CHAPTER 18

FREQUENCY ANALYSIS OF EXTREME EVENTS

Jery R. Stedinger
School of Civil and Environmental Engineering
Cornell University
Ithaca, New York

Richard M. Vogel
Department of Civil Engineering
Tufts University
Medford, Massachusetts

Efi Foufoula-Georgiou
Department of Civil and Mineral Engineering
University of Minnesota
Minneapolis, Minnesota

18.1 INTRODUCTION TO FREQUENCY ANALYSIS

Extreme rainfall events and the resulting floods can take thousands of lives and cause billions of dollars in damage. Flood plain management and designs for flood control works, reservoirs, bridges, and other investigations need to reflect the likelihood or probability of such events. Hydrologic studies also need to address the impact of unusually low stream flows and pollutant loadings because of their effects on water quality and water supplies.

The Basic Problem. Frequency analysis is an information problem: if one had a sufficiently long record of flood flows, rainfall, low flows, or pollutant loadings, then a frequency distribution for a site could be precisely determined, so long as change over time due to urbanization or natural processes did not alter the relationships of concern. In most situations, available data are insufficient to precisely define the risk of large floods, rainfall, pollutant loadings, or low flows. This forces hydrologists to use practical knowledge of the processes involved, and efficient and robust statistical techniques, to develop the best estimates of risk that they can.[115] These techniques are generally restricted, with 10 to 100 sample observations to estimate events exceeded with a chance of at least 1 in 100, corresponding to exceedance probabilities of 1 percent or more. In some cases, they are used to estimate the rainfall exceeded with a chance of 1 in 1000, and even the flood flows for spillway design exceeded with a chance of 1 in 10,000.

The hydrologist should be aware that in practice the true probability distributions of the phenomena in question are not known. Even if they were, their functional representation would likely have too many parameters to be of much practical use. The practical issue is how to select a reasonable and simple distribution to describe the phenomenon of interest, to estimate that distribution's parameters, and thus to obtain risk estimates of satisfactory accuracy for the problem at hand.

Common Problems. The hydrologic problems addressed by this chapter primarily deal with the magnitudes of a single variable. Examples include annual minimum 7-day-average low flows, annual maximum flood peaks, or 24-h maximum precipitation depths. These annual maxima and minima for successive years can generally be considered to be independent and identically distributed, making the required frequency analyses straightforward.

In other instances the risk may be attributable to more than one factor. Flood risk at a site may be due to different kinds of events which occur in different seasons, or due to risk from several sources of flooding or *coincident events,* such as both local tributary floods and large regional floods which result in backwater flooding from a reservoir or major river. When the magnitudes of different factors are independent, a *mixture model* can be used to estimate the combined risk (see Sec. 18.6.2). In other instances, it may be necessary or advantageous to consider all events that exceed a specified threshold because it makes a larger data set available, or because of the economic consequences of every event; such *partial duration series* are discussed in Sec. 18.6.1.

18.1.1 Probability Concepts

Let the upper case letter X denote a *random variable,* and the lower case letter x a possible value of X. For a random variable X, its *cumulative distribution function* (cdf), denoted $F_X(x)$, is the probability the random variable X is less than or equal to x:

$$F_X(x) = P(X \leq x) \tag{18.1.1}$$

$F_X(x)$ is the nonexceedance probability for the value x.

Continuous random variables take on values in a continuum. For example, the magnitude of floods and low flows is described by positive real values, so that $X \geq 0$. The *probability density function* (pdf) describes the relative likelihood that a continuous random variable X takes on different values, and is the derivative of the cumulative distribution function:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{18.1.2}$$

Section 18.2 and Table 18.2.1 provide examples of cdf's and pdf's.

18.1.2 Quantiles, Exceedance Probabilities, Odds Ratios, and Return Periods

In hydrology the *percentiles* or *quantiles* of a distribution are often used as design events. The $100p$ percentile or the pth quantile x_p is the value with cumulative probability p:

$$F_X(x_p) = p \tag{18.1.3}$$

The $100p$ percentile x_p is often called the $100(1-p)$ percent exceedance event because it will be exceeded with probability $1-p$.

The *return period* (sometimes called the *recurrence interval*) is often specified rather than the exceedance probability. For example, the annual maximum flood-flow exceeded with a 1 percent probability in any year, or chance of 1 in 100, is called the 100-year flood. In general, x_p is the T-year flood for

$$T = \frac{1}{1-p} \tag{18.1.4}$$

Here are two ways that return period can be understood. First, in a fixed T-year period the expected number of exceedances of the T-year event is exactly 1, if the distribution of floods does not change over that period; thus on average one flood greater than the T-year flood level occurs in a T-year period.

Alternatively, if floods are independent from year to year, the probability that the first exceedance of level x_p occurs in year k is the probability of $(k-1)$ years without an exceedance followed by a year in which the value of X exceeds x_p:

$$P\,(\text{exactly } k \text{ years until } X \geq x_p) = p^{k-1}\,(1-p). \tag{18.1.5}$$

This is a geometric distribution with mean $1/(1-p)$. Thus the *average* time until the level x_p is exceeded equals T years. However, the probability that x_p is not exceeded in a T-year period is $p^T = (1 - 1/T)^T$, which for $1/(1-p) = T \geq 25$ is approximately 36.7 percent, or about a chance of 1 in 3.

Return period is a means of expressing the exceedance probability. Hydrologists often speak of the 20-year flood or the 1000-year rainfall, rather than events exceeded with probabilities of 5 or 0.1 percent in any year, corresponding to chances of 1 in 20, or of 1 in 1000. Return period has been incorrectly understood to mean that one and only one T-year event should occur every T years. Actually, the probability of the T-year flood being exceeded is $1/T$ in every year. The awkwardness of small probabilities and the incorrect implication of return periods can both be avoided by reporting *odds ratios:* thus the 1 percent exceedance event can be described as a value with a 1 in 100 chance of being exceeded each year.

18.1.3 Product Moments and their Sample Estimators

Several summary statistics can describe the character of the probability distribution of a random variable. Moments and quantiles are used to describe the location or central tendency of a random variable, and its spread, as described in Sec. 17.2 and Table 17.2.1. The *mean* of a random variable X is defined as

$$\mu_X = E[X] \tag{18.1.6}$$

The second moment about the mean is the *variance,* denoted Var (X) or σ_X^2 where

$$\sigma_X^2 = \mathrm{Var}\,(X) = E[(X - \mu_X)^2] \tag{18.1.7}$$

The standard deviation σ_X is the square root of the variance and describes the width or scale of a distribution. These are examples of *product moments* because they depend upon powers of X.

A dimensionless measure of the variability in X, appropriate for use with positive random variables $X \geq 0$, is the *coefficient of variation,* defined in Table 18.1.1. Table 18.1.1 also defines the *coefficient of skewness* γ_X, which describes the relative asymmetry of a distribution, and the *coefficient of kurtosis,* which describes the thickness of a distribution's tails.

Sample Estimators. From a set of observations $(X_1, \ldots, X_n)$, the moments of a distribution can be estimated. Estimators of the mean, variance, and coefficient of skewness are

$$\hat{\mu}_X = \overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$\hat{\sigma}_X^2 = S^2 = \left[\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}\right] \tag{18.1.8}$$

$$\hat{\gamma}_X = G = \frac{n \sum_{i=1}^{n} (X_i - \overline{X})^3}{(n-1)(n-2)S^3}$$

TABLE 18.1.1 Definitions of Dimensionless Product-Moment and L-Moment Ratios

Name	Denoted	Definition
Product-moment ratios		
Coefficient of variation	CV_X	σ_x/μ_x
Coefficient of skewness*	γ_X	$\dfrac{E(X-\mu_x)^3}{\sigma_X^3}$
Coefficient of Kurtosis†		$\dfrac{E(X-\mu_x)^4}{\sigma_X^4}$
L-moment ratios		
L-coefficient of variation‡	L-CV, τ_2	λ_2/λ_1
L-coefficient of skewness	L-skewness, τ_3	λ_3/λ_2
L-coefficient of kurtosis	L-kurtosis, τ_4	λ_4/λ_2

* Some texts define $\beta_1 = [\gamma_X]^2$ as a measure of skewness.

† Some texts define the kurtosis as $\{E[(X-\mu_x)^4]/\sigma_X^4 - 3\}$; others use the term *excess kurtosis* for this difference because the normal distribution has a kurtosis of 3.

‡ Hosking[72] uses τ instead of τ_2 to represent the L-CV ratio.

Some studies use different versions of $\hat{\sigma}_X^2$ and $\hat{\gamma}_X$ that result from replacing $(n-1)$ and $(n-2)$ in Eq. (18.1.8) by n. This makes relatively little difference for large n. The factor $(n-1)$ in the expression for $\hat{\sigma}_X^2$ yields an unbiased estimator of the variance σ_X^2. The factor $n/[(n-1)(n-2)]$ in expression for $\hat{\gamma}_X$ yields an unbiased estimator of $E[(X-\mu_X)^3]$, and generally reduces but does not eliminate the bias of $\hat{\gamma}_X$ (Ref. 159). Kirby[84] derives bounds on the sample estimators of the coefficients of variation and skewness; in fact, the absolute value of both S and G cannot exceed $\sqrt{n}$ for the sample product-moment estimators in Eq. (18.1.8).

Use of Logarithmic Transformations. When data vary widely in magnitude, as often happens in water-quality monitoring, the sample product moments of the logarithms of the data are often employed to summarize the characteristics of a data set or to estimate parameters of distributions. A logarithmic transformation is an effective vehicle for normalizing values which vary by orders of magnitude, and also for keeping occasionally large values from dominating the calculation of product-moment estimators. However, the danger with use of logarithmic transformations is that unusually small observations (or *low outliers*) are given greatly increased weight. This is a concern if it is the large events that are of interest, small values are poorly measured, small values reflect rounding, or small values are reported as zero if they fall below some threshold.

18.1.4 L Moments and Probability-Weighted Moments

L moments are another way to summarize the statistical properties of hydrologic data.[72] The first L-moment estimator is again the mean:

$$\lambda_1 = E[X] \tag{18.1.9}$$

Let $X_{(i|n)}$ be the ith-largest observation in a sample of size n ($i = 1$ corresponds to the largest). Then, for any distribution, the second L moment is a description of scale based on the expected difference between two randomly selected observations:

$$\lambda_2 = \tfrac{1}{2}\, E[X_{(1|2)} - X_{(2|2)}] \tag{18.1.10}$$

Similarly, L-moment measures of skewness and kurtosis use

$$\begin{aligned}\lambda_3 &= \tfrac{1}{3}\, E[X_{(1|3)} - 2X_{(2|3)} + X_{(3|3)}]\\ \lambda_4 &= \tfrac{1}{4}\, E[X_{(1|4)} - 3X_{(2|4)} + 3X_{(3|4)} - X_{(4|4)}]\end{aligned} \tag{18.1.11}$$

as shown in Table 18.1.1.

Advantages of L Moments. Sample estimators of L moments are linear combinations (hence the name *L moments*) of the ranked observations, and thus do not involve squaring or cubing the observations as do the product-moment estimators in Eq. (18.1.8). As a result, L-moment estimators of the dimensionless coefficients of variation and skewness are almost unbiased and have very nearly a normal distribution; the product-moment estimators of the coefficients of variation and of skewness in Table 18.1.1 are both highly biased and highly variable in small samples. Both Hosking[72] and Wallis[163] discuss these issues. In many hydrologic applications an occasional event may be several times larger than other values; when product moments are used, such values can mask the information provided by the other observations, while product moments of the logarithms of sample values can overemphasize small values. In a wide range of hydrologic applications, L moments provide simple and reasonably efficient estimators of the characteristics of hydrologic data and of a distribution's parameters.

L-Moment Estimators. Just as the variance, or coefficient of skewness, of a random variable are functions of the moments $E[X]$, $E[X^2]$, and $E[X^3]$, L moments can be written as functions of *probability-weighted moments* (PWMs),[48,72] which can be defined as

$$\beta_r = E\{X\,[F(X)]^r\} \tag{18.1.12}$$

where $F(X)$ is the cdf for X. Probability-weighted moments are the expectation of X times powers of $F(X)$. (Some authors define PWMs in terms of powers of $[1 - F(X)]$.) For $r = 0$, β_0 is the population mean μ_X.

Estimators of L moments are mostly simply written as linear functions of estimators of PWMs. The first PWM estimator b_0 of β_0 is the sample mean $\bar{X}$ in Eq. (18.1.8).

To estimate other PWMs, one employs the ordered observations, or the *order statistics* $X_{(n)} \leq \cdots \leq X_{(1)}$, corresponding to the sorted or ranked observations in a sample $(X_i | i = 1, \cdots, n)$. A simple estimator of β_r for $r \geq 1$ is

$$b_r^* = \frac{1}{n}\sum_{j=1}^{n} X_{(j)}\left[1 - \frac{(j - 0.35)}{n}\right]^r \tag{18.1.13}$$

where $1 - (j - 0.35)/n$ are estimators of $F(X_{(j)})$. b_r^* is suggested for use when estimating quantiles and fitting a distribution at a single site; though it is biased, it generally yields smaller mean square error quantile estimators than the unbiased estimators in Eq. (18.1.14) below.[68,89]

When unbiasedness is important, one can employ unbiased PWM estimators

$$\begin{aligned}
b_0 &= \bar{X} \\
b_1 &= \sum_{j=1}^{n-1} \frac{(n-j)X_{(j)}}{n(n-1)} \\
b_2 &= \sum_{j=1}^{n-2} \frac{(n-j)(n-j-1)X_{(j)}}{n(n-1)(n-2)} \\
b_3 &= \sum_{j=1}^{n-3} \frac{(n-j)(n-j-1)(n-j-2)X_{(j)}}{n(n-1)(n-2)(n-3)}
\end{aligned} \tag{18.1.14}$$

These are examples of the general formula

$$\hat{\beta}_r = b_r = \frac{1}{n}\sum_{j=1}^{n-r} \frac{\binom{n-j}{r} X_{(j)}}{\binom{n-1}{r}} = \frac{1}{(r+1)}\sum_{j=1}^{n-r} \frac{\binom{n-j}{r} X_{(j)}}{\binom{n}{r+1}} \tag{18.1.15}$$

for $r = 1, \ldots, n - 1$ [see Ref. 89, which defines PWMs in terms of powers of $(1 - F)$]; this formula can be derived using the fact that $(r + 1)\beta_r$ is the expected value of the largest observation in a sample of size $(r + 1)$. The unbiased estimators are recommended for calculating L moment diagrams and for use with regionalization procedures where unbiasedness is important.

For any distribution, L moments are easily calculated in terms of PWMs from

$$\begin{aligned}\lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0\end{aligned} \tag{18.1.16}$$

Estimates of the λ_i are obtained by replacing the unknown β_r by sample estimators b_r from Eq. (18.1.14). Table 18.1.1 contains definitions of dimensionless L-moment coefficients of variation τ_2, of skewness τ_3, and of kurtosis τ_4. L-moment ratios are bounded. In particular, for nondegenerate distributions with finite means, $|\tau_r| < 1$ for $r = 3$ and 4, and for positive random variables, $X > 0$, $0 < \tau_2 < 1$. Table 18.1.2 gives expressions for λ_1, λ_2, τ_3, and τ_4 for several distributions. Figure 18.1.1 shows relationships between τ_3 and τ_4. (A library of FORTRAN subroutines for L-moment analyses is available;[74] see Sec. 18.11.)

Table 18.1.3 provides an example of the calculation of L moments and PWMs. The short rainfall record exhibits relatively little variability and almost zero skewness. The sample product-moment CV for the data of 0.25 is about twice the L-CV $\hat{\tau}_2$ equal to 0.14, which is typical because λ_2 is often about half of σ.

L-Moment and PWM Parameter Estimators. Because the first r L moments are linear combinations of the first r PWMs, fitting a distribution so as to reproduce the first r sample L moments is equivalent to using the corresponding sample PWMs. In fact, PWMs were developed first in terms of powers of $(1 - F)$ and used as effective statistics for fitting distributions.[89,90] Later the PWMs were expressed as L moments which are more easily interpreted.[72,128] Section 18.2 provides formulas for the parameters of several distributions in terms of sample L moments, many of which are obtained by inverting expressions in Table 18.1.2. (See also Ref. 72.)

18.1.5 Parameter Estimation

Fitting a distribution to data sets provides a compact and smoothed representation of the frequency distribution revealed by the available data, and leads to a systematic procedure for extrapolation to frequencies beyond the range of the data set. When flood flows, low flows, rainfall, or water-quality variables are well-described by some family of distributions, a task for the hydrologist is to estimate the parameters Θ of that distribution so that required quantiles and expectations can be calculated with the "fitted" model. For example, the normal distribution has two parameters, μ and σ^2. Appropriate choices for distribution functions can be based on examination of the data using probability plots and moment ratios (discussed in Sec. 18.3), the physical origins of the data, previous experience, and administrative guidelines.

Several general approaches are available for estimating the parameters of a distribution. A simple approach is the *method of moments,* which uses the available sample to compute an estimate $\hat{\Theta}$ of Θ so that the theoretical moments of the distribution of X exactly equal the corresponding sample moments described in Sec. 18.1.3. Alternatively, parameters can be estimated using the sample L moments discussed in Sec. 18.1.4, corresponding to the *method of L moments.*

Still another method that has strong statistical motivation is the *method of maximum likelihood.* Maximum likelihood estimators (MLEs) have very good statistical

TABLE 18.1.2 Values of L Moments and Relationships for the Inverse of the cdf for Several Distributions

Distribution and inverse cdf	L moments
Uniform: $x = \alpha + (\beta - \alpha)F$	$\lambda_1 = \frac{\beta+\alpha}{2}$ $\quad \lambda_2 = \frac{\beta-\alpha}{6}$ $\tau_3 = \tau_4 = 0$
Exponential:* $x = \xi - \frac{\ln[1-F]}{\beta}$	$\lambda_1 = \xi + \frac{1}{\beta}$ $\quad \lambda_2 = \frac{1}{2\beta}$ $\tau_3 = \frac{1}{3}$ $\quad \tau_4 = \frac{1}{6}$
Normal† $x = \mu + \sigma\Phi^{-1}[F]$	$\lambda_1 = \mu$ $\quad \lambda_2 = \frac{\sigma}{\sqrt{\pi}}$ $\tau_3 = 0$ $\quad \tau_4 = 0.1226$
Gumbel: $x = \xi - \alpha \ln[-\ln F]$	$\lambda_1 = \xi + 0.5772\,\alpha$ $\quad \lambda_2 = \alpha \ln 2$ $\tau_3 = 0.1699$ $\quad \tau_4 = 0.1504$
GEV: $x = \xi + \frac{\alpha}{\kappa}\{1 - [-\ln F]^\kappa\}$	$\lambda_1 = \xi + \frac{\alpha}{\kappa}\{1 - \Gamma[1+\kappa]\}$ $\quad \lambda_2 = \frac{\alpha}{\kappa}(1 - 2^{-\kappa})\,\Gamma(1+\kappa)$ $\tau_3 = \left\{\frac{2(1-3^{-\kappa})}{(1-2^{-\kappa})} - 3\right\}$ $\tau_4 = \frac{1 - 5(4^{-\kappa}) + 10(3^{-\kappa}) - 6(2^{-\kappa})}{1 - 2^{-\kappa}}$
Generalized Pareto: $x = \xi + \frac{\alpha}{\kappa}\{1 - [1-F]^\kappa\}$	$\lambda_1 = \xi + \frac{\alpha}{1+\kappa}$ $\quad \lambda_2 = \frac{\alpha}{(1+\kappa)(2+\kappa)}$ $\tau_3 = \frac{1-\kappa}{3+\kappa}$ $\quad \tau_4 = \frac{(1-\kappa)(2-\kappa)}{(3+\kappa)(4+\kappa)}$
Lognormal	See Eqs. (18.2.12), (18.2.13)
Gamma	See Eqs. (18.2.30), (18.2.31)

* Alternative parameterization consistent with that for Pareto and GEV distributions is: $x = \xi - \alpha \ln[1-F]$ yielding $\lambda_1 = \xi + \alpha$; $\lambda_2 = \alpha/2$.
† Φ^{-1} denotes the inverse of the standard normal distribution (see Sec. 18.2.1).
Note: F denotes cdf $F_X(x)$.
Source: Adapted from Ref. 72, with corrections.

properties in large samples, and experience has shown that they generally do well with records available in hydrology. However, often MLEs cannot be reduced to simple formulas, so estimates must be calculated using numerical methods.[85] MLEs sometimes perform poorly when the distribution of the observations deviates in significant ways from the distribution being fit.

A different philosophy is embodied in Bayesian inference, which combines prior information and regional hydrologic information with the likelihood function for available data. Advantages of the Bayesian approach are that it allows the explicit modeling of uncertainty in parameters, and provides a theoretically consistent framework for integrating systematic flow records with regional and other hydrologic information.[3,88,127,150]

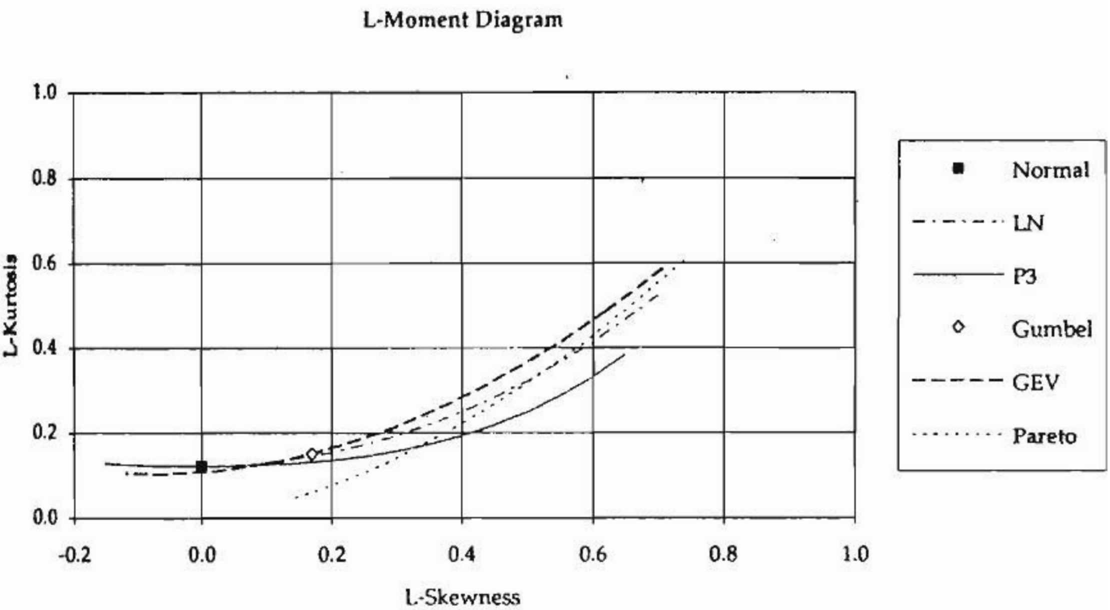

FIGURE 18.1.1 This L-moment diagram illustrates the relationship between the L-kurtosis τ_4 and the L-skewness τ_3 for the normal, lognormal (LN), Pearson type 3 (P3), Gumbel, generalized extreme value (GEV), and Pareto distributions.

TABLE 18.1.3 Example of Calculation of L Moments Using Ranked Annual Maximum 10-min Rainfall Depths for Chicago, 1940–1947

Year	1943	1941	1944	1945	1946	1947	1942	1940
Depth, in*	0.92	0.70	0.66	0.65	0.63	0.60	0.57	0.34
Rank	1	2	3	4	5	6	7	8

Biased PWMs [Eq. (18.1.13)]

$\bar{x} = b_0^* = 0.6338;\ b_1^* = 0.3434;\ b_2^* = 0.2355$

Unbiased PWMs [Eq. (18.1.14)]

$$\bar{x} = b_0 = 0.6338;\ b_1 = \frac{0.92 + (6/7)*(0.70) + (5/7)*(0.66) + \cdots + 0*(0.34)}{8} = 0.3607$$

$$b_2 = \frac{0.92 + (30/42)*(0.70) + (20/42)*(0.66) + \cdots + 0*(0.57) + 0*(0.34)}{8} = 0.2548$$

L moments using unbiased PWMs [Eq. (18.1.16)]

$\hat{\lambda}_1 = b_0 = 0.6338;\ \hat{\lambda}_2 = 2b_1 - b_0 = 0.0877;\ \hat{\lambda}_3 = 6b_2 - 6b_1 + b_0 = -0.0016$

L-CV and L skewness (Table 18.1.1)

$\hat{\tau}_2 = 0.138;\ \hat{\tau}_3 = -0.018$

Product moments for comparison [Eq. (18.1.8)]

$n = 8,\ \bar{x} = 0.6338;\ s = 0.160;\ \mathrm{CV} = 0.252,\ G = -0.088$

* 1 in = 25.4 mm.

Occasionally *nonparametric methods* are employed to estimate frequency relationships. These have the advantage that they do not assume that floods are drawn from a particular family of distributions.[2,7] Modern nonparametric methods have not yet seen much use in practice and have rarely been used officially. However, curve-fitting procedures which employ plotting positions discussed in Sec. 18.3.2 are nonparametric procedures often used in hydrology.

Of concern are the *bias, variability,* and *accuracy* of parameter estimators $\hat{\Theta}[X_1, \ldots, X_n]$, where this notation emphasizes that an estimator $\hat{\Theta}$ is a random variable whose value depends on observed sample values $\{X_1, \ldots X_n\}$. Studies of estimators evaluate an estimator's bias, defined as

$$\text{Bias}\,[\hat{\Theta}] = E[\hat{\Theta}] - \Theta \tag{18.1.17}$$

and sample-to-sample variability, described by Var $[\hat{\Theta}]$. One wants estimators to be nearly unbiased so that on average they have nearly the correct value, and also to have relatively little variability. One measure of accuracy which combines bias and variability is the *mean square error,* defined as

$$\text{MSE}\,[\hat{\Theta}] = E[(\hat{\Theta} - \Theta)^2] = \{\text{Bias}\,[\hat{\Theta}]\}^2 + \text{Var}\,[\hat{\Theta}] \tag{18.1.18}$$

An *unbiased* estimator (Bias $[\hat{\Theta}] = 0$) will have a mean square error equal to its variance. For a given sample size n, estimators with the smallest possible mean square errors are said to be *efficient.*

Bias and mean square error are statistically convenient criteria for evaluating estimators of a distribution's parameters or of quantiles. In particular situations hydrologists can also evaluate the expected probability and under- or overdesign, or use economic loss functions related to operation and design decisions.[112,124]

18.2 PROBABILITY DISTRIBUTIONS FOR EXTREME EVENTS

This section provides descriptions of several families of distributions commonly used in hydrology. These include the normal/lognormal family, the Gumbel/Weibull/generalized extreme value family, and the exponential/Pearson/log-Pearson type 3 family. Table 18.2.1 provides a summary of the pdf or cdf of these probability distributions, and their means and variances. (See also Refs. 54 and 85.) The L moments for several distributions are reported in Table 18.1.2. Many other distributions have also been successfully employed in hydrologic applications, including the five-parameter Wakeby distribution,[69,75] the Boughton distribution,[14] and the TCEV distribution (corresponding to a mixture of two Gumbel distributions[119]).

18.2.1 Normal Family: N, LN, LN3

The normal (N), or Gaussian distribution is certainly the most popular distribution in statistics. It is also the basis of the lognormal (LN) and three-parameter lognormal (LN3) distributions which have seen many applications in hydrology. This section describes the basic properties of the normal distribution first, followed by a discussion of the LN and LN3 distributions. Goodness-of-fit tests are discussed in Sec. 18.3 and standard errors of quantile estimators in Sec. 18.4.2.

The Normal Distribution. The normal distribution is useful in hydrology for describing well-behaved phenomena such as average annual stream flow, or average annual pollutant loadings. The *central limit theorem* demonstrates that if a random variable X is the sum of n independent and identically distributed random variables with finite variance, then with increasing n the distribution of X becomes normal regardless of the distribution of the original random variables.

The pdf for a normal random variable X is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right] \tag{18.2.1}$$

X is unbounded both above and below, with mean μ_X and variance σ_X^2. The normal distribution's skew coefficient is zero because the distribution is symmetric. The product-moment coefficient of kurtosis, $E[(X-\mu_X)^4]/\sigma^4$, equals 3. L moments are given in Table 18.1.2.

The two moments of the normal distribution, μ_X and σ_X^2, are its natural parameters. They are generally estimated by the sample mean and variance in Eq. (18.1.8); these are the maximum likelihood estimates if $(n-1)$ is replaced by n in the denominator of the sample variance. The cdf of the normal distribution is not available in closed form. Selected points z_p for the *standard normal distribution* with zero mean and unit variance are given in Table 18.2.2; because the normal distribution is symmetric, $z_p = -z_{1-p}$.

An approximation, generally adequate for simple tasks and plotting, for the standard normal cdf, denoted $\Phi(z)$, is

$$\Phi(z) = 1 - 0.5 \exp\left[-\frac{(83z+351)z+562}{703/z+165}\right] \tag{18.2.2}$$

for $0 < z \le 5$. An approximation for the inverse of the standard normal cdf, denoted $\Phi^{-1}(p)$ is

$$\Phi^{-1}(p) = z_p = \frac{p^{0.135}-(1-p)^{0.135}}{0.1975} \tag{18.2.3a}$$

or the more accurate expression valid for $10^{-7} < p < 0.5$

$$\Phi^{-1}(p) = z_p = -\sqrt{\frac{y^2[(4y+100)y+205]}{[(2y+56)y+192]y+131}} \tag{18.2.3b}$$

where $y = -\ln(2p)$. [Eqs. (18.2.2) and (18.2.3*b*) are from Ref. 35.]

Lognormal Distribution. Many hydrologic processes are positively skewed and are not normally distributed. However, in many cases for *strictly positive random variables* $X > 0$, their logarithm

$$Y = \ln(X) \tag{18.2.4}$$

is well-described by a normal distribution. This is particularly true if the hydrologic variable results from some multiplicative process, such as dilution. Inverting Eq. (18.2.4) yields

$$X = \exp(Y)$$

TABLE 18.2.1 Commonly Used Frequency Distributions in Hydrology (see also Table 18.1.2)

Distribution	pdf and/or cdf	Range	Moments
Normal	$f_X(x)=\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]$	$-\infty<x<\infty$	μ_X and σ_X^2; $\gamma_X=0$
Lognormal*	$f_X(x)=\frac{1}{x\sqrt{2\pi\sigma_Y^2}}\exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\mu_Y}{\sigma_Y}\right)^2\right]$	$x>0$	$\mu_X=\exp\left(\mu_Y+\frac{\sigma_Y^2}{2}\right)$ $\sigma_X^2=\mu_X^2[\exp(\sigma_Y^2)-1]$ $\gamma_x=3\mathrm{CV}_x+\mathrm{CV}_x^3$
Pearson type 3	$f_X(x)=\lvert\beta\rvert[\beta(x-\xi)]^{\alpha-1}\frac{\exp[-\beta(x-\xi)]}{\Gamma(\alpha)}$	$\alpha>0$	$\mu_x=\xi+\frac{\alpha}{\beta}$; $\sigma_X^2=\frac{\alpha}{\beta^2}$
	$\Gamma(\alpha)$ is the gamma function	for $\beta>0$: $x>\xi$	and $\gamma_x=\frac{2}{\sqrt{\alpha}}$
	(for $\beta>0$ and $\xi=0$: $\gamma_x=2\,\mathrm{CV}_x$)	for $\beta<0$: $x<\xi$	and $\gamma_x=\frac{-2}{\sqrt{\alpha}}$
log-Pearson type 3	$f_X(x)=\lvert\beta\rvert\{\beta[\ln(x)-\xi]\}^{\alpha-1}\frac{\exp\{-\beta[\ln(x)-\xi]\}}{x\Gamma(\alpha)}$	for $\beta<0$, $0<x<\exp(\xi)$; for $\beta>0$, $\exp(\xi)<x<\infty$	See Eq. (18.2.34)
Exponential	$f_X(x)=\beta\exp[-\beta(x-\xi)]$ $F_X(x)=1-\exp\{-\beta(x-\xi)\}$	$x>\xi$ for $\beta>0$	$\mu_X=\xi+\frac{1}{\beta}$; $\sigma_X^2=\frac{1}{\beta^2}$ $\gamma_X=2$
Gumbel	$f_X(x)=\frac{1}{\alpha}\exp\left[-\frac{x-\xi}{\alpha}-\exp\left(-\frac{x-\xi}{\alpha}\right)\right]$ $F_X(x)=\exp\left[-\exp\left(-\frac{x-\xi}{\alpha}\right)\right]$	$-\infty<x<\infty$	$\mu_x=\xi+0.5772\alpha$ $\sigma_X^2=\frac{\pi^2\alpha^2}{6}\approx1.645\alpha^2$; $\gamma_x=1.1396$
GEV	$F_X(x)=\exp\left\{-\left[1-\frac{\kappa(x-\xi)}{\alpha}\right]^{1/\kappa}\right\}$	(σ_x^2 exists for $\kappa>-0.5$) when $\kappa>0$, $x<\left(\xi+\frac{\alpha}{\kappa}\right)$; $\kappa<0$, $x>\left(\xi+\frac{\alpha}{\kappa}\right)$	$\mu_x=\xi+\left(\frac{\alpha}{\kappa}\right)[1-\Gamma(1+\kappa)]$ $\sigma_x^2=\left(\frac{\alpha}{\kappa}\right)^2\{\Gamma(1+2\kappa)-[\Gamma(1+\kappa)]^2\}$
Weibull	$f_X(x)=\left(\frac{k}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^k\right]$ $F_X(x)=1-\exp[-(x/\alpha)^k]$	$x>0$; $\alpha, k>0$	$\mu_x=\alpha\,\Gamma\left(1+\frac{1}{k}\right)$ $\sigma_X^2=\alpha^2\left\{\Gamma\left(1+\frac{2}{k}\right)-\left[\Gamma\left(1+\frac{1}{k}\right)\right]^2\right\}$
Generalized Pareto	$f_X(x)=\left(\frac{1}{\alpha}\right)\left[1-\kappa\frac{(x-\xi)}{\alpha}\right]^{1/\kappa-1}$ $F_X(x)=1-\left[1-\kappa\frac{(x-\xi)}{\alpha}\right]^{1/\kappa}$	for $\kappa<0$, $\xi\le x<\infty$ for $\kappa>0$, $\xi\le x\le\xi+\frac{\alpha}{\kappa}$ (γ_X exists for $\kappa>-0.33$)	$\mu_X=\xi+\frac{\alpha}{(1+\kappa)}$ $\sigma_X^2=\alpha^2/[(1+\kappa)^2(1+2\kappa)]$ $\gamma_X=\frac{2(1-\kappa)(1+2\kappa)^{1/2}}{(1+3\kappa)}$

* Here $Y=\ln(X)$. Text gives formulas for three-parameter lognormal distribution, and for two- and three-parameter lognormal with common base 10 logarithms.

TABLE 18.2.2 Quantiles of the Standard Normal Distribution

p	0.5	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.998	0.999
z_p	0.000	0.253	0.675	0.842	1.282	1.645	1.960	2.326	2.878	3.090

If X has a lognormal distribution, the cdf for X is

$$F_X(x) = P(X \leq x) = P[Y \leq \ln(x)] = P\left[\frac{Y-\mu_Y}{\sigma_Y} \leq \frac{\ln(x)-\mu_Y}{\sigma_Y}\right]$$

$$= \Phi\left[\frac{\ln(x)-\mu_Y}{\sigma_Y}\right] \quad (18.2.5)$$

where Φ is the cdf of the standard normal distribution. The lognormal pdf for X in Table 18.2.1 is illustrated in Fig. 18.2.1. $f_X(x)$ is tangent to the horizontal axis at $x = 0$. As a function of the coefficient of variation CV_X, the skew coefficient is

$$\gamma_X = 3CV_X + CV_X^3$$

As the coefficients of variation and skewness go to zero, the lognormal distribution approaches a normal distribution.

Table 18.2.1 provides formulas for the first three moments of a lognormally distributed variable X in terms of the first two moments of the normally distributed variable Y. The relationships for μ_X and σ_X^2 can be inverted to obtain

$$\sigma_Y = \left[\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)\right]^{1/2} \quad \text{and} \quad \mu_Y = \ln(\mu_X) - \tfrac{1}{2}\sigma_Y^2 \quad (18.2.6)$$

These two equations allow calculation of the method of moments estimators of μ_Y and σ_Y, which are the natural parameters of the lognormal distribution.

Alternatively, the logarithms of the sample (x_i) are a sample of Y's: $[y_i = \ln(x_i)]$.

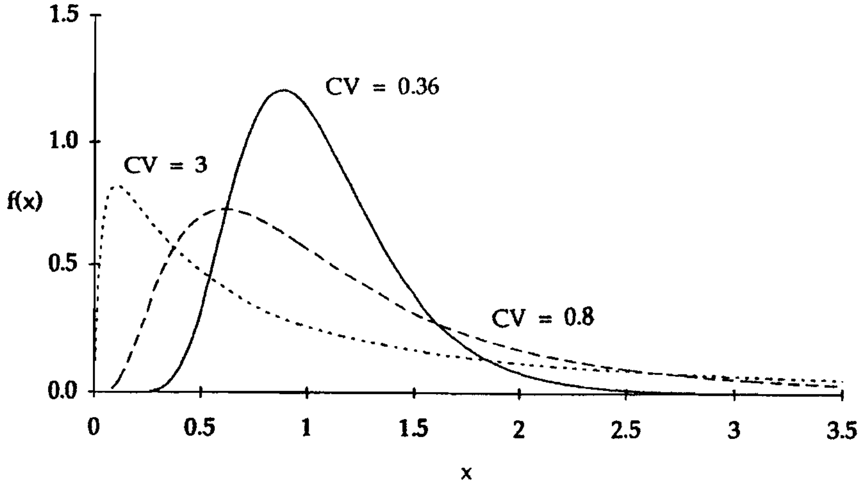

FIGURE 18.2.1 The probability density function of the lognormal distribution with coefficients of variation CV = 0.36, 0.8, and 3, which have coefficients of skewness $\gamma_X = 1.13$, 2.9, and 33 (corresponding to $\mu_Y = 0$ and $\sigma_Y = 0.35, 0.7$, and 1.5 for base e logarithms, or $\mu_W = 0$ and $\sigma_W = 0.15, 0.30$, and 0.65 for common base 10 logarithms.)

The sample mean and variance of the observed (y_i), obtained by using Eq. (18.1.8), are the maximum-likelihood estimators of the lognormal distribution's parameters if $(n-1)$ is replaced by n in the denominator of s_Y^2. The moments of the y_i's are both easier to compute and generally more efficient than the moment estimators in Eq. (18.2.6), provided the sample does not include unusually small values;[126] see discussion of logarithmic transformations in Sec. 18.1.3.

Hydrologists often use common base 10 logarithms instead of natural logarithms. Let W be the common logarithm of X, $\log(X)$. Then Eq. (18.2.5) becomes

$$F_X(x) = P\left[\frac{W-\mu_W}{\sigma_W} \le \frac{\log(x)-\mu_W}{\sigma_W}\right] = \Phi\left[\frac{\log(x)-\mu_W}{\sigma_W}\right]$$

The moments of X in terms of those of W are

$$\mu_X = 10^{\mu_W + \ln(10)\sigma_W^2/2} \quad \text{and} \quad \sigma_X^2 = \mu_X^2\left(10^{\ln(10)\sigma_W^2} - 1\right) \tag{18.2.7}$$

where $\ln(10) = 2.303$. These expressions may be inverted to obtain:

$$\sigma_W = \left[\frac{\log(1+\sigma_X^2/\mu_X^2)}{\ln(10)}\right]^{1/2} \quad \text{and} \quad \mu_W = \log(\mu_X) - \tfrac{1}{2}\ln(10)\,\sigma_W^2 \tag{18.2.8}$$

Three-Parameter Lognormal Distribution. In many cases the logarithms of a random variable X are not quite normally distributed, but subtracting a lower bound parameter ξ before taking logarithms may resolve the problem. Thus

$$Y = \ln(X-\xi) \tag{18.2.9a}$$

is modeled as having a normal distribution, so that

$$X = \xi + \exp(Y) \tag{18.2.9b}$$

For any probability level p, the quantile x_p is given by

$$x_p = \xi + \exp(\mu_Y + \sigma_Y z_p) \tag{18.2.9c}$$

In this case the first two moments of X are

$$\mu_X = \xi + \exp(\mu_Y + \tfrac{1}{2}\sigma_Y^2) \quad \text{and} \quad \sigma_X^2 = [\exp(2\mu_Y + \sigma_Y^2)][\exp(\sigma_Y^2) - 1] \tag{18.2.10a}$$

with skewness coefficient

$$\gamma_X = 3\phi + \phi^3$$

where $\phi = [\exp(\sigma_Y^2) - 1]^{0.5}$. If common base 10 logarithms are employed so that $W = \log(X-\xi)$, the value of ξ and the formula for γ_X are unaffected, but Eq. (18.2.10a) becomes

$$\mu_X = \xi + 10^{\mu_W + \ln(10)\sigma_W^2/2} \quad \text{and} \quad \sigma_X^2 = (\mu_X - \xi)^2\phi^2 \tag{18.2.10b}$$

with $\phi = (10^{\ln(10)\sigma_W^2/2} - 1)^{0.5}$.

Method-of-moment estimators for the three-parameters lognormal distribution are relatively inefficient. A simple and efficient estimator of ξ is the *quantile-lower-bound estimator:*

$$\hat{\xi} = \frac{x_{(1)}x_{(n)} - x_{\text{median}}^2}{x_{(1)} + x_{(n)} - 2x_{\text{median}}} \tag{18.2.11}$$

TABLE 18.2.2 Quantiles of the Standard Normal Distribution

p	0.5	0.6	0.75	0.8	0.9	0.95	0.975	0.99	0.998	0.999
z_p	0.000	0.253	0.675	0.842	1.282	1.645	1.960	2.326	2.878	3.090

If X has a lognormal distribution, the cdf for X is

$$F_X(x) = P(X \le x) = P[Y \le \ln(x)] = P\left[\frac{Y-\mu_Y}{\sigma_Y} \le \frac{\ln(x)-\mu_Y}{\sigma_Y}\right]$$
$$= \Phi\left[\frac{\ln(x)-\mu_Y}{\sigma_Y}\right] \quad (18.2.5)$$

where Φ is the cdf of the standard normal distribution. The lognormal pdf for X in Table 18.2.1 is illustrated in Fig. 18.2.1. $f_X(x)$ is tangent to the horizontal axis at $x = 0$. As a function of the coefficient of variation CV_X, the skew coefficient is

$$\gamma_X = 3\mathrm{CV}_X + \mathrm{CV}_X^3$$

As the coefficients of variation and skewness go to zero, the lognormal distribution approaches a normal distribution.

Table 18.2.1 provides formulas for the first three moments of a lognormally distributed variable X in terms of the first two moments of the normally distributed variable Y. The relationships for μ_X and σ_X^2 can be inverted to obtain

$$\sigma_Y = \left[\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)\right]^{1/2} \quad \text{and} \quad \mu_Y = \ln(\mu_X) - \tfrac{1}{2}\sigma_Y^2 \quad (18.2.6)$$

These two equations allow calculation of the method of moments estimators of μ_Y and σ_Y, which are the natural parameters of the lognormal distribution.

Alternatively, the logarithms of the sample (x_i) are a sample of Y's: $[y_i = \ln(x_i)]$.

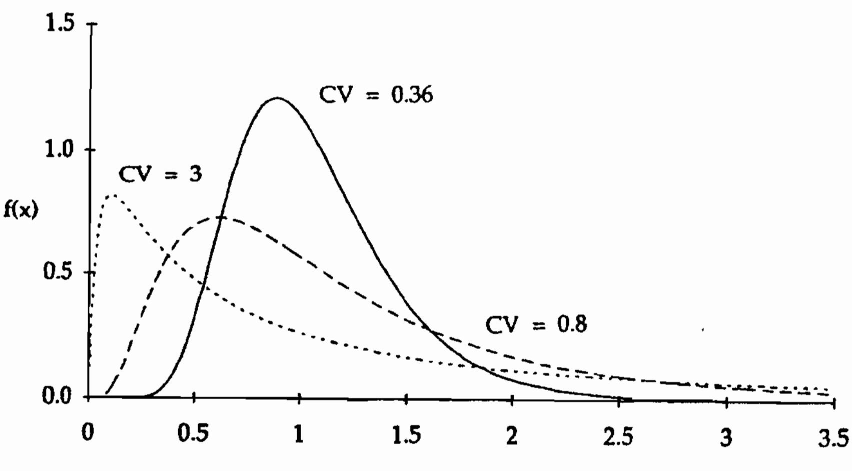

FIGURE 18.2.1 The probability density function of the lognormal distribution with coefficients of variation CV = 0.36, 0.8, and 3, which have coefficients of skewness $\gamma_X = 1.13$, 2.9, and 33 (corresponding to $\mu_Y = 0$ and $\sigma_Y = 0.35, 0.7$, and 1.5 for base e logarithms, or $\mu_W = 0$ and $\sigma_W = 0.15, 0.30$, and 0.65 for common base 10 logarithms.)

he sample mean and variance of the observed (y_i), obtained by using Eq. (18.1.8), re the maximum-likelihood estimators of the lognormal distribution's parameters if $\imath - 1$) is replaced by n in the denominator of s_Y^2. The moments of the y_i's are both asier to compute and generally more efficient than the moment estimators in Eq. 18.2.6), provided the sample does not include unusually small values;[126] see discuson of logarithmic transformations in Sec. 18.1.3.

Hydrologists often use common base 10 logarithms instead of natural logarithms. et W be the common logarithm of X, log (X). Then Eq. (18.2.5) becomes

$$F_X(x) = P\left[\frac{W-\mu_W}{\sigma_W} \le \frac{\log(x)-\mu_W}{\sigma_W}\right] = \Phi\left[\frac{\log(x)-\mu_W}{\sigma_W}\right]$$

he moments of X in terms of those of W are

$$\mu_X = 10^{\mu_W + \ln(10)\sigma_W^2/2} \qquad \text{and} \qquad \sigma_X^2 = \mu_X^2\,(10^{\ln(10)\sigma_W^2} - 1) \tag{18.2.7}$$

here ln(10) = 2.303. These expressions may be inverted to obtain:

$$\sigma_W = \left[\frac{\log(1+\sigma_X^2/\mu_X^2)}{\ln(10)}\right]^{1/2} \qquad \text{and} \qquad \mu_W = \log(\mu_X) - \tfrac{1}{2}\ln(10)\,\sigma_W^2 \tag{18.2.8}$$

hree-Parameter Lognormal Distribution. In many cases the logarithms of a ranom variable X are not quite normally distributed, but subtracting a lower bound ırameter ξ before taking logarithms may resolve the problem. Thus

$$Y = \ln(X - \xi) \tag{18.2.9a}$$

modeled as having a normal distribution, so that

$$X = \xi + \exp(Y) \tag{18.2.9b}$$

ır any probability level p, the quantile x_p is given by

$$x_p = \xi + \exp(\mu_Y + \sigma_Y z_p) \tag{18.2.9c}$$

this case the first two moments of X are

$$= \xi + \exp(\mu_Y + \tfrac{1}{2}\sigma_Y^2) \qquad \text{and} \qquad \sigma_X^2 = [\exp(2\mu_Y + \sigma_Y^2)]\,[\exp(\sigma_Y^2) - 1] \tag{18.2.10a}$$

th skewness coefficient

$$\gamma_X = 3\phi + \phi^3$$

ıere $\phi = [\exp(\sigma_Y^2) - 1]^{0.5}$. If common base 10 logarithms are employed so that $= \log(X - \xi)$, the value of ξ and the formula for γ_X are unaffected, but Eq.].2.10a) becomes

$$\mu_X = \xi + 10^{\mu_W + \ln(10)\,\sigma_W^2/2} \qquad \text{and} \qquad \sigma_X^2 = (\mu_X - \xi)^2\phi^2 \tag{18.2.10b}$$

h $\phi = (10^{\ln(10)\sigma_W^2/2} - 1)^{0.5}$.

Method-of-moment estimators for the three-parameters lognormal distribution relatively inefficient. A simple and efficient estimator of ξ is the *quantile-lowerınd estimator:*

$$\hat{\xi} = \frac{x_{(1)}x_{(n)} - x_{\text{median}}^2}{x_{(1)} + x_{(n)} - 2x_{\text{median}}} \tag{18.2.11}$$

when $x_{(1)} + x_{(n)} - 2x_{\text{median}} > 0$, where $x_{(1)}$ and $x_{(n)}$ are, respectively, the largest and smallest observed values; x_{median} is the sample medium equal to $x_{(k+1)}$ for odd sample sizes $n = 2k + 1$, and $\frac{1}{2}(x_{(k)} + x_{(k+1)})$ for even $n = 2k$. [When $x_{(1)} + x_{(n)} - 2x_{\text{median}} < 0$, the formula provides an upper bound so that $\ln(\xi - x)$ would be normally distributed.] Given ξ, one can estimate μ_Y and σ_Y^2 by using the sample mean and variance of $y_i = \ln(x_i - \hat{\xi})$, or $w_i = \log(x_i - \hat{\xi})$. The quantile-lower-bound estimator's performance with the resultant sample estimators of μ_Y and σ_Y^2 is better than method-of-moments estimators and competitive with maximum likelihood estimators.[67,126]

For the two-parameter and three-parameter lognormal distribution, the second L moment is

$$\lambda_2 = \exp\left(\mu_Y + \frac{\sigma_Y^2}{2}\right) \operatorname{erf}\left(\frac{\sigma_Y}{2}\right) = 2\exp\left(\mu_Y + \frac{\sigma_Y^2}{2}\right)\left[\Phi\left(\frac{\sigma_Y}{\sqrt{2}}\right) - \frac{1}{2}\right] \quad (18.2.12)$$

The following polynomial approximates, within 0.0005 for $|\tau_3| < 0.9$, the relationship between the third and fourth L-moment ratios, and is thus useful for comparing sample values of those ratios with the theoretical values for two- or three-parameter lognormal distributions:[73]

$$\tau_4 = 0.12282 + 0.77518\,\tau_3^2 + 0.12279\,\tau_3^4 - 0.13638\,\tau_3^6 + 0.11368\,\tau_3^8 \quad (18.2.13)$$

18.2.2 GEV Family: Gumbel, GEV, Weibull

Many random variables in hydrology correspond to the maximum of several similar processes, such as the maximum rainfall or flood discharge in a year, or the lowest stream flow. The physical origin of such random variables suggests that their distribution is likely to be one of several *extreme value (EV) distributions* described by Gumbel.[51] The cdf of the largest of n independent variates with common cdf $F(x)$ is simply $F(x)^n$. (See Sec. 18.6.2.) For large n and many choices for $F(x)$, $F(x)^n$ converges to one of three extreme value distributions, called types I, II, and III. Unfortunately, for many hydrologic variables this convergence is too slow for this argument alone to justify adoption of an extreme value distribution as a model of annual maxima and minima.

This section first considers the EV type I distribution, called the *Gumbel distribution*. The *generalized extreme value distribution* (GEV) is then introduced. It spans the three types of extreme value distributions for maxima popularized by Gumbel.[68,80] Finally, the *Weibull distribution* is developed, which is the extreme value type III distribution for minima bounded below by zero. Goodness-of-fit tests are discussed in Sec. 18.3 and standard errors of quantile estimators in Sec. 18.4.4.

The Gumbel Distribution. Let $M_1, \ldots, M_n$ be a set of daily rainfall, stream flow, or pollutant concentrations, and let the random variable $X = \max(M_i)$ be the maximum for the year. If the M_i are independent and identically distributed random variables unbounded above, with an "exponential-like" upper tail (examples include the normal, Pearson type 3, and lognormal distributions), then for large n the variate X has an extreme value type I distribution, or Gumbel distribution.[3,51] For example, the annual-maximum 24-h rainfall depths are often described by a Gumbel distribution, as are annual maximum stream flows.

The Gumbel distribution has the cdf, mean, and variance given in Table 18.2.1, and corresponding L moments are given in Table 18.1.3. The cdf is easily inverted to obtain

$$x_p = \xi - \alpha \ln[-\ln(p)] \quad (18.2.14)$$

The estimator of α obtained by using the second sample L moment is

$$\hat{\alpha} = \frac{\hat{\lambda}_2}{\ln(2)} = 1.443\,\hat{\lambda}_2 \tag{18.2.15}$$

If the sample variance s^2 from Eq. (18.1.8) were employed instead, one obtains

$$\hat{\alpha} = \frac{s\sqrt{6}}{\pi} = 0.7797\,s \tag{18.2.16}$$

The corresponding estimator of ξ in either case is

$$\hat{\xi} = \bar{x} - 0.5772\,\hat{\alpha} \tag{18.2.17}$$

L-moment estimators for the Gumbel distribution are generally as good or better than method-of-moment estimators when the observations are actually drawn from a Gumbel distribution, though maximum likelihood estimators are the best in that case.[89] However, L-moment estimators have been shown to be robust, providing more accurate quantile estimators than product-moment and maximum-likelihood estimators when observations are drawn from a range of reasonable distributions for flood flows.[90]

The Gumbel distribution's density function is very similar to that of the lognormal distribution with $\gamma = 1.13$ in Fig. 18.2.1. Changing ξ and α moves the center of the Gumbel pdf, and changes its width, but does not change the shape of the distribution. The Gumbel distribution has a fixed coefficient of skewness $\gamma = 1.1396$. For large x, the Gumbel distribution is asymptotically equivalent to the exponential distribution with cdf $\{1 - \exp[-(x-\xi)/\alpha]\}$.

The Generalized Extreme Value Distribution. This is a general mathematical form which incorporates Gumbel's type I, II, and III extreme value distributions for maxima.[68,80] The GEV distribution's cdf can be written

$$F(x) = \exp\left\{-\left[1 - \frac{\kappa(x-\xi)}{\alpha}\right]^{1/\kappa}\right\} \qquad \text{for } \kappa \neq 0 \tag{18.2.18}$$

The Gumbel distribution is obtained when $\kappa = 0$. For $|\kappa| < 0.3$, the general shape of the GEV distribution is similar to the Gumbel distribution, though the right-hand tail is thicker for $\kappa < 0$ and thinner for $\kappa > 0$.

Here ξ is a location parameter, α is a scale parameter, and κ is the important shape parameter. For $\kappa > 0$ the distribution has a finite upper bound at $\xi + \alpha/\kappa$ and corresponds to the EV type III distribution for maxima that are bounded above; for $\kappa < 0$, the distribution has a thicker right-hand tail and corresponds to the EV type II distribution for maxima from thick-tailed distributions like the generalized Pareto distribution in Table 18.2.1 with $\kappa < 0$.

The moments of the GEV distribution can be expressed in terms of the gamma function, $\Gamma(\bullet)$. For $\kappa > -1/3$, the mean and variance are given in Table 18.2.1, whereas

$$\gamma_X = \text{Sign}(\kappa)\,\frac{-\Gamma(1+3\kappa) + 3\Gamma(1+\kappa)\,\Gamma(1+2\kappa) - 2\Gamma^3(1+\kappa)}{[\Gamma(1+2\kappa) - \Gamma^2(1+\kappa)]^{3/2}} \tag{18.2.19}$$

where Sign (κ) is plus or minus 1 depending on the sign of κ, and $\Gamma(\,)$ is the gamma function for which an approximation is supplied in Eq. (18.2.21). For $\kappa > -1$ the order r PWM β_r of a GEV distribution is

$$\beta_r = (r+1)^{-1}\left\{\xi + \frac{\alpha}{\kappa}\left[1 - \frac{\Gamma(1+\kappa)}{(r+1)^{\kappa}}\right]\right\} \tag{18.2.20}$$

L moments for the GEV distribution are given in Table 18.1.2.

For $0 \le \delta \le 1$, a good approximation of the *gamma function*, useful with Eqs. (18.2.19) and (18.2.20) is

$$\Gamma(1+\delta) = 1 + \sum_{i=1}^{5} a_i \delta^i + \epsilon \tag{18.2.21}$$

where $a_1 = -0.574\,8646$
$a_2 = 0.951\,2363$
$a_3 = -0.699\,8588$
$a_4 = 0.424\,5549$
$a_5 = -0.101\,0678$

with $|\epsilon| \le 5 \times 10^{-5}$ [Eq. (6.1.35) in Ref. 1]. For larger arguments one can use the relationship $\Gamma(1+w) = w\Gamma(w)$ repeatedly until $0 < w < 1$; for integer w, $\Gamma(1+w) = w!$ is the factorial function.

The parameters of the GEV distribution in terms of L moments are[68]

$$\kappa = 7.8590c + 2.9554c^2 \tag{18.2.22a}$$

$$\alpha = \frac{\kappa\lambda_2}{\Gamma(1+\kappa)\,(1-2^{-\kappa})} \tag{18.2.22b}$$

$$\xi = \lambda_1 + \frac{\alpha\,[\Gamma(1+\kappa)-1]}{\kappa[\cancel{\Gamma(1+\kappa)-1}]} \tag{18.2.22c}$$

where

$$c = \frac{2\lambda_2}{\lambda_3 + 3\lambda_2} - \frac{\ln(2)}{\ln(3)} = \frac{2\beta_1 - \beta_0}{3\beta_2 - \beta_0} - \frac{\ln(2)}{\ln(3)}$$

The quantiles of the GEV distribution can be calculated from

$$x_p = \xi + \frac{\alpha}{\kappa}\{1 - [-\ln(p)]^{\kappa}\} \tag{18.2.23}$$

where p is the cumulative probability of interest. Typically $|\kappa| \le 0.20$.

When data are drawn from a Gumbel distribution ($\kappa = 0$), using the biased estimator b_r^* in Eq. (18.1.13) to calculate the L-moment estimators in Eq. (18.2.22), the resultant estimator of κ has a mean of 0 and variance[68]

$$\mathrm{Var}(\hat{\kappa}) = \frac{0.5633}{n} \tag{18.2.24}$$

Comparison of the statistic $Z = \hat{\kappa}\sqrt{n/0.5633}$ with standard normal quantiles allows construction of a powerful test of whether $\kappa = 0$ or not when fitting a GEV distribution.[68,72] Chowdhury et al.[22] provide formulas for the sampling variance of the sample L-moment skewness and kurtosis $\hat{\tau}_3$ and $\hat{\tau}_4$ as a function of κ for the GEV distribution so that one can test if a particular data set is consistent with a GEV distribution with a regional value of κ.

Weibull Distribution. If W_i are the minimum stream flows in different days of the year, then the annual minimum is the smallest of the W_i, each of which is bounded below by zero. In this case the random variable $X = \min(W_i)$ may be well-described by the EV type III distribution for minima, or the Weibull distribution. Table 18.2.1 includes the Weibull cdf, mean, and variance. The skewness coefficient is the negative of that in Eq. (18.2.19) with $\kappa = 1/k$. The second L moment is

$$\lambda_2 = \alpha(1 - 2^{-1/k})\,\Gamma\left(1 + \frac{1}{k}\right) \tag{18.2.25}$$

Equation (18.2.21) provides an approximation for $\Gamma(1 + \delta)$.

For $k < 1$ the Weibull pdf goes to infinity as x approaches zero, and decays slowly for large x. For $k = 1$ the Weibull distribution reduces to the exponential distribution in Fig. 18.2.2 corresponding to $\gamma = 2$ and $\alpha_{P3} = 1$ in that figure. For $k > 1$, the Weibull density function is like a Pearson type 3 distribution's density function in Fig. 18.2.2 for small x and $\alpha_{P3} = k$, but decays to zero faster for large x. Parameter estimation methods are discussed in Refs. 57 and 85.

There are important relationships between the Weibull, Gumbel, and GEV distributions. If X has a Weibull distribution, then $Y = -\ln[X]$ has a Gumbel distribution. This allows parameter estimation procedures [Eqs. (18.2.15) to (18.2.17)] and goodness-of-fit tests available for the Gumbel distribution to be used for the Weibull; thus if $+\ln(X)$ has mean $\lambda_{1,(\ln X)}$ and L-moment $\lambda_{2,(\ln X)}$, X has Weibull parameters

$$k = \frac{\ln(2)}{\lambda_{2,(\ln X)}} \qquad \text{and} \qquad \alpha = \exp\left(\lambda_{1,(\ln X)} + \frac{0.5772}{k}\right) \tag{18.2.26}$$

Section 18.1.3 discusses use of logarithmic transformations.

If Y has a EV type III distribution (GEV distribution with $\kappa > 0$) for maxima bounded above, then $(\xi + \alpha/\kappa) - Y$ has a Weibull distribution with $k = 1/\kappa$; thus for $k > 0$, the third and fourth L-moment ratios for the Weibull distribution equal $-\tau_3$ and τ_4 for the GEV distribution in Table 18.1.2. A three-parameter Weibull distribution can be fit by the method of L moments by using Eq. (18.2.22) applied to $-X$.

18.2.3 Pearson Type 3 Family: Pearson Type 3 and Log-Pearson Type 3

Another family of distributions used in hydrology is that based on the Pearson type 3 (P3) distribution.[13] It is one of several families of distributions the statistician Pearson proposed as convenient models of random variables. Goodness-of-fit tests are discussed in Sec. 18.3, and standard errors of quantile estimators, in Sec. 18.4.3.

The pdf of the P3 distribution is given in Table 18.2.1. For $\beta > 0$ and lower bound $\xi = 0$, the P3 distribution reduces to the *gamma distribution* for which $\gamma_X = 2\mathrm{CV}_X$. In some instances, the P3 distribution is used with $\beta < 0$, yielding a negatively skewed distribution with an upper bound of ξ.

Figure 18.2.2 illustrates the shape of the P3 pdf for various values of the skew coefficient γ. For a fixed mean and variance, in the limit as the shape parameter α goes to infinity and the skew coefficient γ goes to zero, the Pearson type 3 distribution converges to the normal distribution. For $\alpha < 1$ and skew coefficient $\gamma > 2$, the P3 pdf goes to infinity at the lower bound. For $\alpha = 1$ and $\gamma = 2$, the two-parameter *exponential distribution* is obtained; see Table 18.2.1.

The moments of the P3 distribution are given in Table 18.2.1. The moment equations can be inverted to obtain

$$\alpha = 4/\gamma_X^2$$

$$\beta = \frac{2}{\sigma_X \gamma_X} \tag{18.2.27}$$

$$\xi = \mu_X - \frac{\alpha}{\beta} = \mu_X - 2\frac{\sigma_X}{\gamma_X}$$

which allows computation of method-of-moment estimators. The method of maximum likelihood is seldom used with this distribution; it does not generate estimates of α less than 1, corresponding to skew coefficients in excess of 2.

A closed-form expression for the cdf of the P3 distribution is not available. Tables or approximations must be used. Many tables provide *frequency factors* $K_p(\gamma)$ which are the pth quantile of a standard P3 variate with skew coefficient γ, mean zero, and variance 1.[20,79] For any mean and standard deviation, the pth P3 quantile can be written

$$x_p = \mu + \sigma K_p(\gamma) \tag{18.2.28}$$

With this parameterization, it is not necessary to estimate the underlying values of α and β when the method of moments is used because the quantiles of the fitted distribution are written as a function of the mean, standard deviation, and the frequency factor. Tables of frequency factors are provided in Ref. 79. The frequency factors for $0.01 \le p \le 0.99$ and $|\gamma| < 2$ are well-approximated by the Wilson-Hilferty

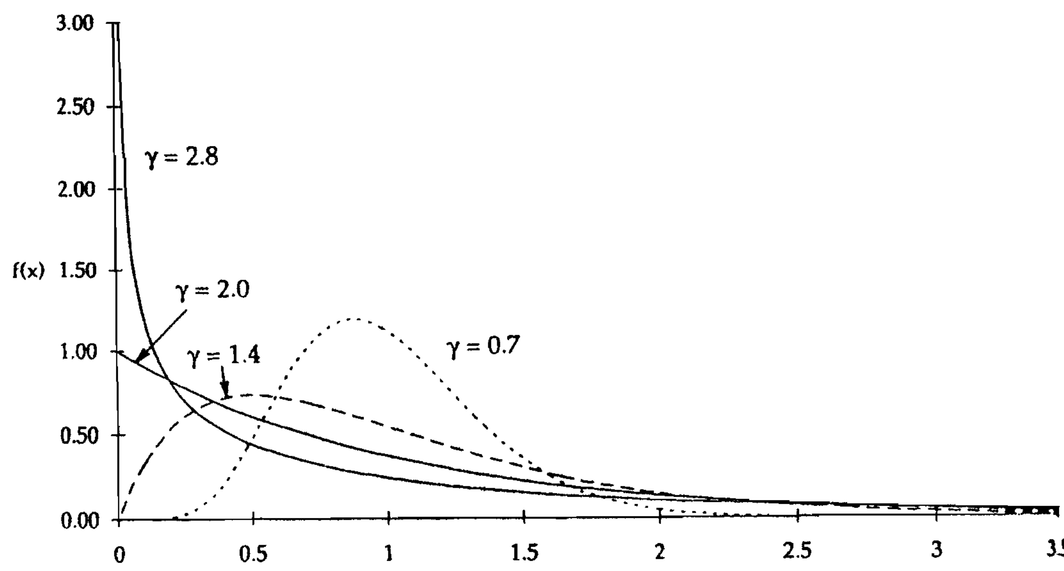

FIGURE 18.2.2 The probability density function for the Pearson type 3 (P3) distribution with lower bound $\xi = 0$, mean $\mu = 1$, and coefficients of skewness $\gamma = 0.7$, 1.4, 2.0, and 2.8 (corresponding to a gamma distribution and shape parameters $\alpha = 8$, 2, 1, and 0.5, respectively).

transformation

$$K_p(\gamma) = \frac{2}{\gamma}\left(1 + \frac{\gamma z_p}{6} - \frac{\gamma^2}{36}\right)^3 - \frac{2}{\gamma} \tag{18.2.29}$$

where z_p is the pth quantile of the zero-mean unit-variance standard normal distribution in Eq. (18.2.3). (Reference 83 provides a better approximation; Ref. 21 evaluates several approximations.)

For the P3 distribution, the first two L moments are

$$\lambda_1 = \xi + \frac{\alpha}{\beta} \qquad \text{and} \qquad \lambda_2 = \frac{\Gamma(\alpha + 0.5)}{\sqrt{\pi}\,\beta\,\Gamma(\alpha)} \tag{18.2.30}$$

An approximation which describes the relationship between the third and fourth L-moment ratios, accurate to within 0.0005 for $|\tau_3| < 0.9$, is[73]

$$\tau_4 = 0.1224 + 0.30115\,\tau_3^2 + 0.95812\,\tau_3^4 - 0.57488\,\tau_3^6 + 0.19383\,\tau_3^8 \tag{18.2.31}$$

Log-Pearson Type 3 Distribution. The log-Pearson type 3 distribution (LP3) describes a random variable whose logarithms are P3-distributed. Thus

$$Q = \exp[X] \tag{18.2.32}$$

where X has a P3 distribution with shape, scale, and location parameters α, β, and ξ. Thus the distribution of the logarithms X of the data is described by Fig. 18.2.2, Eqs. (18.2.27) to (18.2.29), and the corresponding relationships in Table 18.2.1.

The product moments of Q are computed for $\beta > r$ or $\beta < 0$ by using

$$E[Q^r] = e^{r\xi}\left(\frac{\beta}{\beta - r}\right)^\alpha \tag{18.2.33}$$

yielding

$$\mu_Q = e^{\xi}\left(\frac{\beta}{\beta - 1}\right)^\alpha \qquad \sigma_Q{}^2 = e^{2\xi}\left[\left(\frac{\beta}{\beta - 2}\right)^\alpha - \left(\frac{\beta}{\beta - 1}\right)^{2\alpha}\right] \tag{18.2.34}$$

and

$$\gamma_Q = \frac{E[Q^3] - 3\,\mu_Q\,E[Q^2] + 2\,\mu_Q^3}{\sigma_Q^3}$$

The parameter ξ is a lower bound on the logarithms of the random variable if β is positive, and is an upper bound if β is negative. The shape of the real-space flood distribution is a complex function of α and β.[11–13] If one considers W equal to the common logarithm of Q, log (Q), then all the parameters play the same roles, but the new β' and ξ' are smaller by a factor of $1/\ln(10) = 0.4343$.

This distribution was recommended for the description of floods in the United States by the U.S. Water Resources Council in Bulletin 17[79] and in Australia by their Institute of Engineers;[110] Sec. 18.7.2 describes the Bulletin 17 method. It fits a P3 distribution by a modified method of moments to the logarithms of observed flood series using Eq. (18.2.28). Section 18.1.3 discusses pros and cons of logarithmic transformations. Estimation procedures for the LP3 distribution are reviewed in Ref. 5.

18.2.4 Generalized Pareto Distribution

The generalized Pareto distribution (GPD) is a simple distribution useful for describing events which exceed a specified lower bound, such as all floods above a threshold or daily flows above zero. Moments of the GPD are described in Tables 18.1.2 and 18.2.1. A special case is the 2-parameter exponential distribution (for $\kappa = 0$).

For a given lower bound ξ, the shape κ and scale α parameters can be estimated easily with L-moments from

$$\kappa = \frac{\lambda_1 - \xi}{\lambda_2} - 2 \qquad \text{and} \qquad \alpha = (\lambda_1 - \xi)(1 + \kappa) \tag{18.2.35}$$

or the mean and variance formula in Table 18.2.1. In general for $\kappa < 0$, L-moment estimators are preferable. Hosking and Wallis[70] review alternative estimation procedures and their precision. Section 18.6.3 develops a relationship between the Pareto and GEV distributions. If ξ must be estimated, the smaller observation is a good estimator.

18.3 PROBABILITY PLOTS AND GOODNESS-OF-FIT TESTS

18.3.1 Principles and Issues in Selecting a Distribution

Probability plots are extremely useful for visually revealing the character of a data set. Plots are an effective way to see what the data look like and to determine if fitted distributions appear consistent with the data. Analytical goodness-to-fit criteria are useful for gaining an appreciation for whether the lack of fit is likely to be due to sample-to-sample variability, or whether a particular departure of the data from a model is statistically significant. In most cases several distributions will provide statistically acceptable fits to the available data so that goodness-of-fit tests are unable to identify the "true" or "best" distribution to use. Such tests are valuable when they can demonstrate that some distributions appear inconsistent with the data.

Several fundamental issues arise when selecting a distribution.[82] One should distinguish between the following questions:

1. What is the true distribution from which the observations are drawn?
2. What distribution should be used to obtain reasonably accurate and robust estimates of design quantiles and hydrologic risk?
3. Is a proposed distribution consistent with the available data for a site?

Question 1 is often asked. Unfortunately, the true distribution is probably too complex to be of practical use. Still, L-moment skewness-kurtosis and CV-skewness diagrams discussed in Secs. 18.1.4 and 18.3.3 are good for investigating what simple families of distributions are consistent with available data sets for a region. Standard goodness-of-fit statistics, such as probability plot correlation tests in Sec. 18.3.2, have also been used to see how well a member of each family of distributions can fit a sample. Unfortunately, such goodness-of-fit statistics are unlikely to identify the actual family from which the samples are drawn—rather, the most flexible families generally fit the data best. Regional L-moment diagrams focus on the character of sample statistics which describe the "parent" distribution for available samples, rather than goodness-of-fit. Goodness-of-fit tests address Question 3.

Question 2 is important in hydrologic applications and has been the subject of many studies (Ref. 29; examples include Refs. 32, 87, 90, 112, 124). At one time the distribution that best fitted each data set was used for frequency analysis at that site, but this approach has now been largely abandoned. Such a procedure is too sensitive to sampling variations in the data. Operational procedures adopted by different national flood-frequency studies for use in their respective countries should be based on a combination of regionalization of some parameters and split-sample/Monte Carlo evaluations of different estimation procedures to find distribution-estimation procedure combinations which yield reliable flood quantile and risk estimates. Such estimators are called *robust* because they perform reasonably well for a wide range of cases. In the United States, the log-Pearson type 3 distribution with weighted skew coefficient was adopted; an index-flood GEV procedure was selected for the British Isles (see Secs. 18.7.2 and 18.7.3). This principle also applies to frequency analyses of other phenomena.

18.3.2 Plotting Positions and Probability Plots

The graphical evaluation of the adequacy of a fitted distribution is generally performed by plotting the observations so that they would fall approximately on a straight line if a postulated distribution were the true distribution from which the observations were drawn. This can be done with the use of special commercially available probability papers for some distributions, or with the more general technique presented here, on which such special papers are based.[30] Section 17.2.2 also discusses the graphical display of data.

Let $\{X_i\}$ denote the observed values and $X_{(i)}$, the ith largest value in a sample, so that $X_{(n)} \leq X_{(n-1)} \leq \ldots \leq X_{(1)}$. The random variable U_i defined as

$$U_i = 1 - F_X[X_{(i)}] \tag{18.3.1}$$

corresponds to the *exceedance probability* associated with the ith largest observation. If the original observations were independent, in repeated sampling the U_i have a *beta distribution* with mean

$$E[U_i] = \frac{i}{n+1} \tag{18.3.2}$$

and variance

$$\mathrm{Var}\,(U_i) = \frac{i(n-i+1)}{(n+1)^2(n+2)} \tag{18.3.3}$$

Knowing the distribution of the exceedance probabilities U_i, one can develop estimators q_i of their values which can be used to plot each $X_{(i)}$ against a probability scale.

Let $G(x)$ be a proposed cdf for the events. A visual comparison of the data and a fitted distribution is provided by a plot of the ith largest observed event $X_{(i)}$ versus an estimate of what its true value should be. If $G(x)$ is the distribution of X, the value of $X_{(i)} = G^{-1}(1 - U_i)$ should be nearly $G^{-1}(1 - q_i)$, where the *probability-plotting position* q_i is our estimate of U_i. Thus the points $[G^{-1}(1 - q_i), X_{(i)}]$ when plotted would, apart from sampling fluctuation, lie on a straight line through the origin. Such a plot would look like Fig. 18.3.1, which actually displays $[\Phi^{-1}(1 - q_i), \log X_{(i)}]$. The exceed-

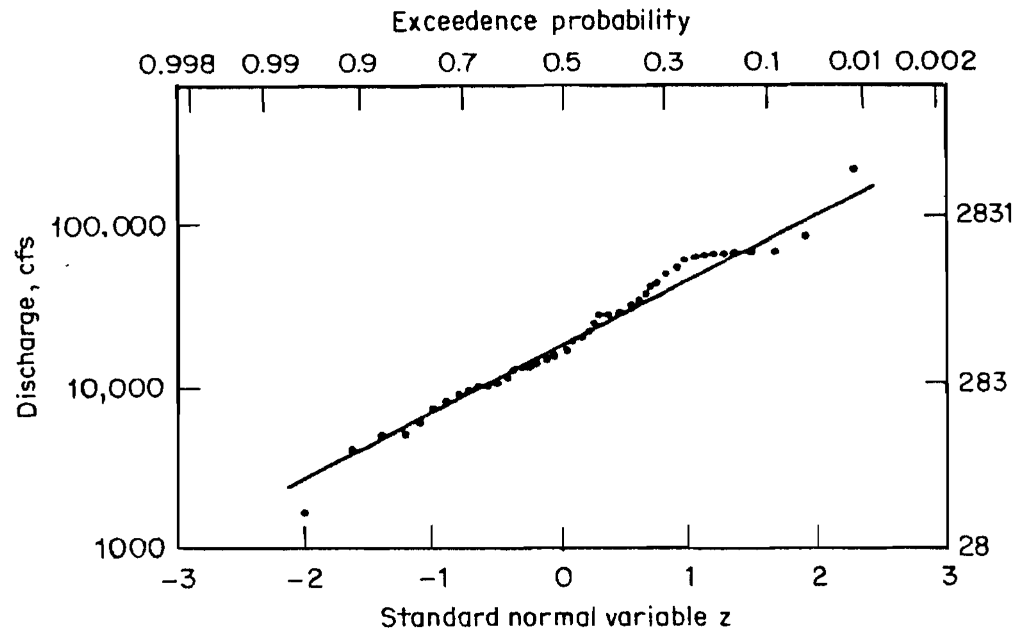

FIGURE 18.3.1 A probability plot using a normal scale of 44 annual maxima for the Guadalupe River near Victoria, Texas. *(Reproduced with permission from Ref. 20, p. 398.)*

ance probability of the ith-largest event is often estimated using the *Weibull plotting position:*

$$q_i = \frac{i}{n+1} \tag{18.3.4}$$

corresponding to the mean of U_i.

Choice of plotting position. Hazen[59] originally developed probability paper and imagined the probability scale divided into n equal intervals with midpoints $q_i = (i - 0.5)/n$, $i = 1, \ldots, n$; these served as his plotting positions. Gumbel[51] rejected this formula in part because it assigned a return period of $2n$ years to the largest observation (see also Harter[58]); Gumbel promoted Eq. (18.3.4).

Cunnane[26] argued that plotting positions q_i should be assigned so that on average $X_{(i)}$ would equal $G^{-1}(1 - q_i)$; that is, q_i would capture the mean of $X_{(i)}$ so that

$$E[X_{(i)}] \approx G^{-1}(1 - q_i) \tag{18.3.5}$$

Such plotting positions would be almost quantile-unbiased. The Weibull plotting positions $i/(n + 1)$ equal the average exceedance probability of the ranked observations $X_{(i)}$, and hence are probability-unbiased plotting positions. The two criteria are different because of the nonlinear relationship between $X_{(i)}$ and $U_{(i)}$.

Different plotting positions attempt to achieve almost quantile-unbiasedness for different distributions; many can be written

$$q_i = \frac{i - a}{n + 1 - 2a} \tag{18.3.6}$$

which is symmetric so that $q_i = 1 - q_{n+1-i}$. Cunanne recommended $a = 0.40$ for obtaining nearly quantile-unbiased plotting positions for a range of distributions.

Other alternatives are Blom's plotting position ($a = \frac{3}{8}$), which gives nearly unbiased quantiles for the normal distribution, and the Gringorten position ($a = 0.44$) which yields optimized plotting positions for the largest observations from a Gumbel distribution.[49] These are summarized in Table 18.3.1, which also reports the return period, $T_1 = 1/q_1$, assigned to the largest observation. Section 18.6.3 develops plotting positions for records that contain censored values.

The differences between the Hazen formula, Cunanne's recommendation, and the Weibull formula is modest for i of 3 or more. However, differences can be appreciable for $i = 1$, corresponding to the largest observation (and $i = n$ for the smallest observation). Remember that the actual exceedance probability associated with the largest observation is a random variable with mean $1/(n + 1)$ and a standard deviation of nearly $1/(n + 1)$; see Eqs. (18.3.2) and (18.3.3). Thus all plotting positions give crude estimates of the unknown exceedance probabilities associated with the largest (and smallest) events.

A good method for illustrating this uncertainty is to consider quantiles of the beta distribution of the actual exceedance probability associated with the largest observation $X_{(1)}$. The actual exceedance probability for the largest observation $X_{(1)}$ in a sample is between $0.29/n$ and $1.38/(n + 2)$ nearly 50 percent of the time; and between $0.052/n$ and $3/(n + 2)$ nearly 90 percent of the time. Such bounds allow one to assess the consistency of the largest (or, by symmetry, the smallest) observation with a fitted distribution better than does a single plotting position.

Probability Paper. It is now possible to see how probability papers can be constructed for many distributions. A probability plot is a graph of the ranked observations $x_{(i)}$ versus an approximation of their expected value $G^{-1}(1 - q_i)$. For the nor-

TABLE 18.3.1 Alternative Plotting Positions and their Motivation*

Name	Formula	a	T_1	Motivation
Weibull	$\dfrac{i}{n+1}$	0	$n + 1$	Unbiased exceedance probabilities for all distributions
Median†	$\dfrac{i-0.3175}{n+0.365}$	0.3175	$1.47n + 0.5$	Median exceedance probabilities for all distributions
APL	$\dfrac{i-0.35}{n}$	~0.35	$1.54n$	Used with PWMs [Eq. (18.1.13)]
Blom	$\dfrac{i-3/8}{n+1/4}$	0.375	$1.60n + 0.4$	Unbiased normal quantiles
Cunnane	$\dfrac{i-0.40}{n+0.2}$	0.40	$1.67n + 0.3$	Approximately quantile-unbiased
Gringorten	$\dfrac{i-0.44}{n+0.12}$	0.44	$1.79n + 0.2$	Optimized for Gumbel distribution
Hazen	$\dfrac{i-0.5}{n}$	0.50	$2n$	A traditional choice

* Here a is the plotting-position parameter in Eq. (18.3.6) and T_1 is the return period each plotting position assigns to the largest observation in a sample of size n.

† For $i = 1$ and n, the exact value is $q_1 = 1 - q_n = 1 - 0.5^{1/n}$.

mal distribution

$$G^{-1}(1-q_i)=\mu+\sigma\,\Phi^{-1}(1-q_i) \tag{18.3.7}$$

Thus, except for intercept and slope, a plot of the observations $x_{(i)}$ versus $G^{-1}[1-q_i]$ is visually identical to a plot of $x_{(i)}$ versus $\Phi^{-1}(1-q_i)$. The values of q_i are often printed along the abscissa or horizontal axis. Lognormal paper is obtained by using a log scale to plot the ordered logarithms log $(x_{(i)})$ versus a normal-probability scale which is equivalent to plotting log $(x_{(i)})$ verus $\Phi^{-1}(1-q_i)$. Figure 18.3.1 illustrates use of lognormal paper with Blom's plotting positions.

For the Gumbel distribution,

$$G^{-1}(1-q_i)=\xi-\alpha\ln[-\ln(1-q_i)] \tag{18.3.8}$$

Thus a plot of $x_{(i)}$ versus $G^{-1}(1-q_i)$ is identical to a plot of $x_{(i)}$ versus the *reduced Gumbel variate*

$$y_i=-\ln[-\ln(1-q_i)] \tag{18.3.9}$$

It is easy to construct probability paper for the Gumbel distribution by plotting $x_{(i)}$ as a function of y_i; the horizontal axis can show the actual values of y or, equivalently, the associated q_i, as in Fig. 18.3.1 for the lognormal distribution.

Special probability papers are not available for the Pearson type 3 or log Pearson type 3 distributions because the frequency factors depend on the skew coefficient. However, for a given value for the coefficient of skewness γ one can plot the observation $x_{(i)}$ for a P3 distribution, or log $(x_{(i)})$ for the LP3 distribution, versus the frequency factors $K_{p_i}(\gamma)$ defined in Eq. (18.2.29) with $p_i=1-q_i$. This should yield a straight line except for sampling error if the correct skew coefficient is employed. Alternatively for the P3 or LP3 distributions, normal or lognormal probability paper is often used to compare the $x_{(i)}$ and a fitted P3 distribution, which plots as a curved

TABLE 18.3.2 Generation of Probability Plots for Different Distributions

Normal probability paper. Plot $x_{(i)}$ versus z_{p_i} given in Eq. (18.2.3), where $p_i=1-q_i$. Blom's formula ($a=3/8$) provides quantile-unbiased plotting positions.

Lognormal probability paper. Plot ordered logarithms log $[x_{(i)}]$ versus z_p. Blom's formula ($a=3/8$) provides quantile-unbiased plotting positions.

Exponential probability paper. Plot ordered observations $x_{(i)}$ versus $\xi-\ln(q_i)/\beta$ or just $-\ln(q_i)$. Gringorten's plotting positions ($a=0.44$) work well.

Gumbel and Weibull probability paper. For Gumbel distribution plot ordered observations $x_{(i)}$ versus $\xi-\alpha\ln[-\ln(1-q_i)]$ or just $y_i=-\ln[-\ln(1-q_i)]$. Gringorten's plotting positions ($a=0.44$) were developed for this distribution. For Weibull distribution plot ln $[x_{(i)}]$ versus $\ln(\alpha)+\ln[-\ln(q_i)]/k$ or just $\ln[-\ln(q_i)]$. (See Ref. 154.)

GEV distribution. Plot ordered observations $x_{(i)}$ versus $\xi+(\alpha/\kappa)\{1-[-\ln(1-q_i)]^{\kappa}$, or just $(1/\kappa)\{1-[-\ln(1-q_i)]^{\kappa}\}$. Alternatively employ Gumbel probability paper on which GEV will be curved. Cunnane's plotting positions ($a=0.4$) are reasonable.[52]

Pearson type 3 probability paper. Plot ordered observations $x_{(i)}$ versus $K_{p_i}(\gamma)$, where $p_i=1-q_i$. Blom's formula ($a=3/8$) is quantile-unbiased for normal distribution and makes sense for small γ. Or employ normal probability paper. (See Ref. 158.)

Log Pearson type 3 probability paper. Plot ordered logarithms log $[x_{(i)}]$ versus $K_{p_i}(\gamma)$ where $p_i=1-q_i$. Blom's formula ($a=3/8$) makes sense for small γ. Or employ lognormal probability paper. (See Ref. 158.)

Uniform probability paper. Plot $x_{(i)}$ versus $1-q_i$, where q_i are the Weibull plotting positions ($a=0$). (See Ref. 154.)

mal distribution

$$G^{-1}(1 - q_i) = \mu + \sigma\,\Phi^{-1}(1 - q_i) \qquad (18.3.7)$$

Thus, except for intercept and slope, a plot of the observations $x_{(i)}$ versus $G^{-1}[1 - q_i$ is visually identical to a plot of $x_{(i)}$ versus $\Phi^{-1}(1 - q_i)$. The values of q_i are often printed along the abscissa or horizontal axis. Lognormal paper is obtained by using a log scale to plot the ordered logarithms log $(x_{(i)})$ versus a normal-probability scale which is equivalent to plotting log $(x_{(i)})$ verus $\Phi^{-1}(1 - q_i)$. Figure 18.3.1 illustrates use of lognormal paper with Blom's plotting positions.

For the Gumbel distribution,

$$G^{-1}(1 - q_i) = \xi - \alpha \ln[-\ln(1 - q_i)] \qquad (18.3.8)$$

Thus a plot of $x_{(i)}$ versus $G^{-1}(1 - q_i)$ is identical to a plot of $x_{(i)}$ versus the *reduced Gumbel variate*

$$y_i = -\ln[-\ln(1 - q_i)] \qquad (18.3.9)$$

It is easy to construct probability paper for the Gumbel distribution by plotting $x_{(i)}$ as a function of y_i; the horizontal axis can show the actual values of y or, equivalently, the associated q_i, as in Fig. 18.3.1 for the lognormal distribution.

Special probability papers are not available for the Pearson type 3 or log Pearson type 3 distributions because the frequency factors depend on the skew coefficient. However, for a given value for the coefficient of skewness γ one can plot the observation $x_{(i)}$ for a P3 distribution, or log $(x_{(i)})$ for the LP3 distribution, versus the frequency factors $K_{p_i}(\gamma)$ defined in Eq. (18.2.29) with $p_i = 1 - q_i$. This should yield a straight line except for sampling error if the correct skew coefficient is employed. Alternatively for the P3 or LP3 distributions, normal or lognormal probability paper is often used to compare the $x_{(i)}$ and a fitted P3 distribution, which plots as a curved

TABLE 18.3.2 Generation of Probability Plots for Different Distributions

Normal probability paper. Plot $x_{(i)}$ versus z_{p_i} given in Eq. (18.2.3), where $p_i = 1 - q_i$. Blom's formula ($a = 3/8$) provides quantile-unbiased plotting positions.

Lognormal probability paper. Plot ordered logarithms log $[x_{(i)}]$ versus z_p. Blom's formula ($a = 3/8$) provides quantile-unbiased plotting positions.

Exponential probability paper. Plot ordered observations $x_{(i)}$ versus $\xi - \ln(q_i)/\beta$ or just $-\ln(q_i)$. Gringorten's plotting positions ($a = 0.44$) work well.

Gumbel and Weibull probability paper. For Gumbel distribution plot ordered observations $x_{(i)}$ versus $\xi - \alpha \ln[-\ln(1 - q_i)]$ or just $y_i = -\ln[-\ln(1 - q_i)]$. Gringorten's plotting positions ($a = 0.44$) were developed for this distribution. For Weibull distribution plot ln $[x_{(i)}]$ versus $\ln(\alpha) + \ln[-\ln(q_i)]/k$ or just $\ln[-\ln(q_i)]$. (See Ref. 154.)

GEV distribution. Plot ordered observations $x_{(i)}$ versus $\xi + (\alpha/\kappa)\{1 - [-\ln(1 - q_i)]^\kappa$, or just $(1/\kappa)\{1 - [-\ln(1 - q_i)]^\kappa\}$. Alternatively employ Gumbel probability paper on which GEV will be curved. Cunnane's plotting positions ($a = 0.4$) are reasonable.[52]

Pearson type 3 probability paper. Plot ordered observations $x_{(i)}$ versus $K_{p_i}(\gamma)$, where $p_i = 1 - q_i$. Blom's formula ($a = 3/8$) is quantile-unbiased for normal distribution and makes sense for small γ. Or employ normal probability paper. (See Ref. 158.)

Log Pearson type 3 probability paper. Plot ordered logarithms log $[x_{(i)}]$ versus $K_{p_i}(\gamma)$ where $p_i = 1 - q_i$. Blom's formula ($a = 3/8$) makes sense for small γ. Or employ lognormal probability paper. (See Ref. 158.)

Uniform probability paper. Plot $x_{(i)}$ versus $1 - q_i$, where q_i are the Weibull plotting positions ($a = 0$). (See Ref. 154.)

line. Table 18.3.2 summarizes how probability plots may be constructed for these and other distributions.

18.3.3 Goodness-of-Fit Tests and L-Moment Diagrams

Rigorous statistical tests are available and are useful for assessing whether or not a given set of observations might have been drawn from a particular family of distributions, as discussed in Sec. 18.3.1. For example, the *Kolmogorov-Smirnov test* provides bounds within which *every observation* on a probability plot should lie if the sample is actually drawn from the assumed distribution; it is useful for evaluating visually the adequacy of a fitted distribution. Stephens[136] gives critical Kolmogorov-Smirnov values for the normal and exponential distributions (reproduced in Ref. 95, p. 112); Chowdhury et al.[22] provide tables for the GEV distribution.

The *probability plot correlation test* discussed below is a more powerful test of whether a sample has been drawn from a postulated distribution; a test with greater *power* has a greater probability of correctly determining that a sample is not from the postulated distribution. *L-moment tests* are also relatively powerful and can be used to determine if a proposed Gumbel, GEV, or normal distribution is consistent with the data. L-moment diagrams are useful as a guide in selecting an appropriate family of distributions for describing a set of variables, such as flood distributions in a region.

Probability Plot Correlation Coefficient Test. A simple but powerful goodness-of-fit test is the probability plot correlation test developed by Filliben.[41] The test uses the correlation r between the ordered observations $x_{(i)}$ and the corresponding fitted quantiles $w_i = G^{-1}(1 - q_i)$, determined by plotting positions q_i for each $x_{(i)}$. Values of r near 1.0 suggest that the observations could have been drawn from the fitted distribution. Essentially, r measures the linearity of the probability plot, providing a quantitative assessment of fit. If $\bar{x}$ denotes the average value of the observations and $\bar{w}$ denotes the average value of the fitted quantiles, then

$$r = \frac{\Sigma\,(x_{(i)} - \bar{x})\,(w_i - \bar{w})}{[\Sigma\,(x_{(i)} - \bar{x})^2\,\Sigma\,(w_i - \bar{w})^2]^{0.5}} \tag{18.3.10}$$

Table 18.3.3 gives critical values of r for the normal distribution, or the logarithms of lognormal variates, based on a plotting position with $a = 3/8$. Values for the Gumbel distribution are reproduced in Table 18.3.4 for use with $a = 0.44$; the table also applies to logarithms of Weibull variates (see Table 18.3.2 and Sec. 18.2.2). Other tables are available for the uniform,[154] the GEV,[22] the Pearson type 3,[158] and exponential and other distributions.[30]

L-Moment Diagrams and Ratio Tests. Figure 18.1.1 provides an example of an L-moment diagram.[72,163] Sample L moments are less biased than traditional product-moment estimators, and thus are better suited for use in constructing moment diagrams. (See Sec. 18.1.4.) Plotting sample statistics on such diagrams allows a choice between alternative families of distributions (Ref. 29). L-moment diagrams include plots of τ_3 versus τ_2 for choosing among two-parameter distributions, or of τ_4 versus τ_3 for choosing among three-parameter distributions. Chowdhury et al.[22] derive the sampling variance of $\hat{\tau}_2$, $\hat{\tau}_3$, and $\hat{\tau}_4$ as a function of κ for the GEV distribution to provide a powerful test of whether a particular data set is consistent with a GEV distribution with a regionally estimated value of κ, or a regional κ and CV. Equation (18.2.24) provides a very powerful test for the Gumbel versus a general

TABLE 18.3.3 Lower Critical Values of the Probability Plot Correlation Test Statistic for the Normal Distribution Using $p_i = (i - 3/8)/(n + 1/4)$

	Significance level		
n	0.10	0.05	0.01
10	0.9347	0.9180	0.8804
15	0.9506	0.9383	0.9110
20	0.9600	0.9503	0.9290
30	0.9707	0.9639	0.9490
40	0.9767	0.9715	0.9597
50	0.9807	0.9764	0.9664
60	0.9835	0.9799	0.9710
75	0.9865	0.9835	0.9757
100	0.9893	0.9870	0.9812
300	0.99602	0.99525	0.99354
1000	0.99854	0.99824	0.99755

Source: Refs. 101, 152, 153. Used with permission.

TABLE 18.3.4 Lower Critical Values of the Probability Plot Correlation Test Statistic for the Gumbel and Two-Parameter Weibull Distributions Using $p_i = (i - 0.44)/(n + 0.12)$

	Significance level		
n	0.10	0.05	0.01
10	0.9260	0.9084	0.8630
20	0.9517	0.9390	0.9060
30	0.9622	0.9526	0.9191
40	0.9689	0.9594	0.9286
50	0.9729	0.9646	0.9389
60	0.9760	0.9685	0.9467
70	0.9787	0.9720	0.9506
80	0.9804	0.9747	0.9525
100	0.9831	0.9779	0.9596
300	0.9925	0.9902	0.9819
1000	0.99708	0.99622	0.99334

Source: Refs. 152, 153. See also Table 18.3.2.

GEV distribution using the sample L-moment estimator of κ. Similarly, if observations have a normal distribution, then $\hat{\tau}_3$ has mean zero and Var $[\hat{\tau}_3] = (0.1866 + 0.8/n)/n$, allowing construction of a powerful test of normality against skewed alternatives[72] using $Z = \hat{\tau}_3/\sqrt{(0.1866/n + 0.8/n^2)}$.

18.4 STANDARD ERRORS AND CONFIDENCE INTERVALS FOR QUANTILES

A simple measure of the precision of a quantile estimator is its variance Var (x_p), which equals the square of the *standard error,* SE, so that $SE^2 = \text{Var}(\hat{x}_p)$. Confidence intervals are another description of precision. Confidence intervals for a quantile are often calculated using the quantile's standard error. When properly constructed, 90 or 99 percent confidence intervals will, in repeated sampling, contain the parameter or quantile of interest 90 or 99 percent of the time. Thus they are an interval which will contain a parameter of interest most of the time.

18.4.1 Confidence Intervals for Quantiles

The classic confidence interval formula is for the mean μ_X of a normally distributed random variable X. If sample observations X_i are independent and normally distributed with the same mean and variance, then a $100(1 - \alpha)\%$ confidence interval for μ_X is

$$\bar{x} - \frac{s_X}{\sqrt{n}} t_{1-\alpha/2,n-1} \le \mu_X \le \bar{x} + \frac{s_X}{\sqrt{n}} t_{1-\alpha/2,n-1} \tag{18.4.1}$$

where $t_{1-\alpha/2,n-1}$ is the upper $100(\alpha/2)\%$ percentile of Student's t distribution with $n - 1$ degrees of freedom. Here $s_X/\sqrt{n}$ is the estimated *standard error* of the sample mean; that is, it is the square root of the variance of the estimator $\bar{X}$ of μ_X. In large samples ($n > 40$), the t distribution is well-approximated by a normal distribution, so that $z_{1-\alpha/2}$ from Table 18.2.2 can replace $t_{1-\alpha/2,n-1}$ in Eq. (18.4.1).

In hydrology, attention often focuses on quantiles of various distributions, such as the 10-year 7-day low flow, or the rainfall depth exceeded with a 1 percent probability. Confidence intervals can be constructed for quantile estimators. Asymptotically (with increasingly large n), most quantile estimators $\hat{x}_p$ are normally distributed. If $\hat{x}_p$ has variance Var $(\hat{x}_p)$ and is essentially normally distributed, then an approximate $100(1 - \alpha)\%$ confidence interval based on Eq. (18.4.1) is

$$\hat{x}_p - z_{1-\alpha/2} \sqrt{\text{Var}(\hat{x}_p)} \quad \text{to} \quad \hat{x}_p + z_{1-\alpha/2} \sqrt{\text{Var}(\hat{x}_p)} \tag{18.4.2}$$

Equation (18.4.2) allows calculation of approximate confidence intervals for quantiles (or parameters) of distributions for which good estimates of their standard errors, $\sqrt{\text{Var}(\hat{x}_p)}$, are available.[85]

18.4.2 Results for Normal/Lognormal Quantiles

For a normally distributed random variable, the traditional estimator of x_p is

$$\hat{x}_p = \bar{x} + z_p s_X \tag{18.4.3}$$

Asymptotically, the variance of this estimator is

$$\mathrm{Var}\,(\hat{x}_p) = \frac{s_x^2}{n}\left(1 + \frac{1}{2}\,z_p^2\right) \tag{18.4.4}$$

Thus, an approximate $100(1-\alpha)\%$ confidence interval for x_p is

$$(\bar{x} + z_p s_x) \pm z_{1-\alpha/2}\sqrt{\frac{s_x^2}{n}\left(1 + \frac{1}{2}\,z_p^2\right)} \tag{18.4.5}$$

These results can also be used to obtain confidence intervals for quantiles of the two-parameter lognormal distribution. If X is lognormally distributed, then $Y = \ln(X)$ is normally distributed and

$$x_p = \exp\,(\mu_y + z_p\,\sigma_y) \tag{18.4.6}$$

The maximum likelihood estimator of x_p is essentially $\exp(\bar{y} + z_p s_y)$; for this estimator,[85,129]

$$\mathrm{Var}\,(\hat{x}_p) \approx x_p^2\left[\frac{\sigma_y^2}{n}\left(1 + \frac{1}{2}z_p^2\right)\right] \tag{18.4.7}$$

A simple but approximate $100(1-\alpha)\%$ confidence interval for the lognormal quantile x_p is

$$\exp\left[(\bar{y} + z_p s_y) \pm z_{1-\alpha/2}\sqrt{\frac{s_y^2}{n}\left(1 + \frac{1}{2}\,z_p^2\right)}\right] \tag{18.4.8}$$

Confidence intervals obtained by substituting Eq. (18.4.7) into Eq. (18.4.2) are not as good as Eq. (18.4.8).[129]

For the normal (and lognormal) distribution, it is also possible to calculate exact confidence intervals using the noncentral t distribution. Let $\zeta_{\alpha,p}$ and $\zeta_{1-\alpha,p}$ denote the 100α and $100(1-\alpha)$ percentiles of the noncentral t distribution. Then an exact $100(1-2\alpha)\%$ confidence interval for $x_p = \mu + z_p\sigma$ when X has a normal distribution is

$$\bar{x} + \zeta_{\alpha,p} s_X < x_p < \bar{x} + \zeta_{1-\alpha,p}\, s_X \tag{18.4.9}$$

Stedinger[129] and App. 9 in Ref. 79 provide tables of percentage points of the ζ distribution. An approximation for $\zeta_{\alpha,p}$ is

$$\zeta_{\alpha,p} \approx \frac{z_p + z_\alpha\sqrt{\dfrac{1}{n} + \dfrac{z_p^2}{2(n-1)} - \dfrac{z_\alpha^2}{2n(n-1)}}}{1 - \dfrac{z_\alpha^2}{2(n-1)}} \tag{18.4.10}$$

which is reasonably accurate for $n \geq 15$ and $\alpha \geq 0.05$ (Ref. 21). $\zeta_{1-\alpha,p}$ is obtained using Eq. (18.4.10) by replacing z_α by $z_{1-\alpha}$ which equals $-z_\alpha$.

Less work has been done on formulas for the variances of lognormal quantiles when three parameters are estimated. Formulas for maximum likelihood, moment, and moment/quantile-lower-bound estimators are evaluated in Ref. 67.

18.4.3 Results for Pearson/Log-Pearson Type 3 Quantiles

Confidence intervals for normal quantiles can be extended to obtain approximate confidence intervals for Pearson Type 3 (P3) quantiles y_p for known skew coefficient γ by using a scaling factor η, obtained from a first-order asymptotic approximation of the P3/normal quantile variance ratio:

$$\eta = \left[\frac{\mathrm{Var}\,(\hat{y}_p)}{\mathrm{Var}\,(\hat{x}_p)}\right]^{1/2} \cong \sqrt{\frac{1 + \gamma K_p + \tfrac{1}{2}\,(1 + \tfrac{3}{4}\,\gamma^2)\,K_p^2}{1 + \tfrac{1}{2}\,z_p^2}} \tag{18.4.11}$$

where K_p is the standard P3 quantile (frequency factor) in Eqs. (18.2.28) and (18.2.29) with cumulative probability p for skew coefficient γ;[129] z_p is the frequency factor for the standard normal distribution in Eq. (18.2.3) employed to compute $\hat{x}_p$ in Eq. (18.4.3). An approximate $100(1 - 2\alpha)\%$ confidence interval for the pth P3 quantile is

$$\hat{y}_p + \eta(\zeta_{1-\alpha,p} - z_p)\, s_y < y_p < \hat{y}_p + \eta(\zeta_{\alpha,p} - z_p) s_y \tag{18.4.12}$$

where $\hat{y}_p = \bar{y} + K_p s_y$.

Chowdhury and Stedinger[21] show that a generalization of Eq. (18.4.12) should be used when the skew coefficient γ is estimated by the at-site sample skew coefficient G_s, a generalized regional estimate G_g, or a weighted estimate of G_s and G_g. For example, if a generalized regional estimate G_g of the coefficient of skewness is employed, and G_g has variance $\mathrm{Var}\,(G_g)$ about the true skew coefficient, then the scaling factor in Eq. (18.4.12) should be calculated as[21]

$$\eta = \sqrt{\frac{1 + \gamma K_p + \tfrac{1}{2}\,(1 + \tfrac{3}{4}\,\gamma^2)\,K_p^2 + n\,\mathrm{Var}\,(G_g)\,(\partial K_p/\partial\gamma)^2}{1 + \tfrac{1}{2}\,z_p^2}} \tag{18.4.13}$$

where, from Eq. (18.2.29),

$$\frac{\partial K_p}{\partial \gamma} \approx \frac{1}{6}\,(z_p^2 - 1)\left[1 - 3\left(\frac{\gamma}{6}\right)^2\right] + (z_p^3 - 6z_p)\,\frac{\gamma}{54} + \frac{2}{3}\,z_p\left(\frac{\gamma}{6}\right)^3 \tag{18.4.14}$$

for $|\gamma| \le 2$ and $0.01 \le p \le 0.99$.

18.4.4 Results for Gumbel and GEV Quantiles

For the Gumbel distribution with two parameters estimated by the method of moments, the variance of the pth quantile is asymptotically[105]

$$\mathrm{Var}\,(\hat{x}_p) = \frac{\alpha^2(1.11 + 0.52y + 0.61y^2)}{n} \tag{18.4.15}$$

for a sample of size n where $y = -\ln\,[-\ln\,(p)]$ is the Gumbel reduced variate.

For unbiased L-moment estimators[38,109]

$$\mathrm{Var}\,(\hat{x}_p) = \frac{\alpha^2[(1.1128 - 0.9066/n) - (0.4574 - 1.1722/n)y + (0.8046 - 0.1855/n)y^2]}{n-1} \tag{18.4.16}$$

Equation (18.4.16) also provides a reasonable estimate of Var (x_p) for use with biased PWMs. These values can be used in Eq. (18.4.2) to obtain approximate confidence intervals. Reference 109 provides formulas for Var $(\hat{x}_p)$ when maximum likelihood estimators are employed.

GEV Index Flood Procedures. The Gumbel and GEV distributions are often used as normalized regional distributions or regional growth curves, as discussed in Sec. 18.5.1. In that case the variance of $\hat{x}_p$ is given by Eq. (18.5.3).

GEV with Fixed κ. The GEV distribution can be used when the location and scale parameters are estimated by using L moments via Eqs. (18.2.22*b*) and (18.2.22*c*) with a fixed regional value of the shape parameter κ, corresponding to a two-parameter index flood procedure (Sec. 18.5.1). For fixed κ the asymptotic variance of the pth quantile estimator with unbiased L-moment estimators is

$$\mathrm{Var}\,(\hat{x}_p) = \frac{\alpha^2(c_1 + c_2 y + c_3 y^2)}{n} \tag{18.4.17}$$

where $y = 1 - [\ln(p)]^\kappa$ when $\kappa \neq 0$ and c_1, c_2, c_3 are coefficients which depend on κ. The asymptotic values of c_1, c_2, c_3 for $-0.33 < \kappa < 0.3$ are well-approximated by[96]

$$c_1 = 1.1128 - 0.2384\kappa + 0.0908\kappa^2 + 0.1084\kappa^3$$

where, for $\kappa > 0$,

$$c_2 = 0.4580 - 3.0561\kappa + 1.1104\kappa^2 - 0.4071\kappa^3$$

$$c_3 = 0.8046 - 2.8890\kappa + 8.7874\kappa^2 - 10.375\kappa^3$$

and, for $\kappa < 0$,

$$c_2 = 0.4580 - 7.5124\kappa + 5.0832\kappa^2 - 11.623\kappa^3 + 2.250\ln(1 + 2\kappa)$$

$$c_3 = 0.8046 - 2.6215\kappa + 6.8989\kappa^2 + 0.003\kappa^3 - 0.1\ln(1 + 3\kappa)$$

For $\kappa = 0$, use Eq. (18.4.16).

Estimation of Three GEV Parameters. All three parameters of the GEV distribution can be estimated with L moments by using Eq. (18.2.22).[68] Asymptotic formulas for the variance of three-parameter GEV quantile estimators are relatively inaccurate in small samples;[96] an estimate of the variance of the pth quantile estimator with

TABLE 18.4.1 Coefficients for an Eq. (18.4.18) That Approximates Variance of Three-Parameter GEV Quantile Estimators

	Cumulative probability level p						
Coefficient	0.80	0.90	0.95	0.98	0.99	0.998	0.999
a_0	−1.813	−2.667	−3.222	−3.756	−4.147	−5.336	−5.943
a_1	3.017	4.491	5.732	7.185	8.216	10.711	11.815
a_2	−1.401	−2.207	−2.367	−2.314	−0.2033	−1.193	−0.630
a_3	0.854	1.802	2.512	4.075	4.780	5.300	6.262

Source: Ref. 96.

unbiased L-moment estimators for $-0.33 < \kappa < 0.3$ is

$$\mathrm{Var}\,(\hat{x}_p) = \frac{\exp\,[a_0(p) + a_1(p)\exp(-\kappa) + a_2(p)\kappa^2 + a_3(p)\kappa^3]}{n} \tag{18.4.18}$$

with coefficients $a_i(p)$ for selected probabilities p in Table 18.4.1 based on the actual sampling variance of unbiased L-moment quantile estimators in samples of size $n = 40$; the variances provided by Eq. (18.4.18) are relatively accurate for sample sizes $20 \le n \le 70$ and $\kappa > -0.20$.

18.5 REGIONALIZATION

Frequency analysis is a problem in hydrology because sufficient information is seldom available at a site to adequately determine the frequency of rare events. At some sites no information is available. When one has 30 years of data to estimate the event exceeded with a chance of 1 in 100 (the 1 percent exceedance event), extrapolation is required. Given that sufficient data will seldom be available at the site of interest, it makes sense to use climatic and hydrologic data from nearby and similar locations.

The National Research Council (Ref. 104, p. 6) proposed three principles for hydrometeorological modeling: "(1) 'substitute space for time'; (2) introduction of more 'structure' into models; and (3) focus on extremes or tails as opposed to, or even to the exclusion of, central characteristics." One substitutes space for time by using hydrologic information at different locations to compensate for short records at a single site. This is easier to do for rainfall which in regions without appreciable relief should have fairly uniform characteristics over large areas. It is more difficult for floods and particularly low flows because of the effects of catchment topography and geology. A successful example of regionalization is the index flood method discussed below. Many other regionalization procedures are available.[28] See also Secs. 18.7.2 and 18.7.3.

Section 18.5.2 discusses regression procedures for deriving regional relationships relating hydrologic statistics to physiographic basin characteristics. These are particularly useful at *ungauged sites.* When floods at a short-record site are highly correlated with floods at a site with a longer record, the *record augmentation* procedures described in Sec. 18.5.3 can be employed. These are both ways of making use of regional hydrologic information.

18.5.1 Index Flood

The index flood procedure is a simple regionalization technique with a long history in hydrology and flood frequency analysis.[31] It uses data sets from several sites in an effort to construct more reliable flood-quantile estimators. A similar regionalization approach in precipitation frequency analysis is the station-year method, which combines rainfall data from several sites without adjustment to obtain a large composite record to support frequency analyses.[15] One can also smooth the precipitation quantiles derived from analysis of the records from different stations.[63]

The concept underlying the index flood method is that the distributions of floods at different sites in a region are the same except for a scale or index-flood parameter which reflects the size, rainfall, and runoff characteristics of each watershed. Generally the mean is employed as the index flood. The problem of estimating the pth quantile x_p is then reduced to estimation of the mean for a site μ_X, and the ratio x_p/μ_X of the pth quantile to the mean. The mean can often be estimated adequately with the record available at a site, even if that record is short. The indicated ratio is estimated by using regional information. The British *Flood Studies Report*[105] calls these normalized regional flood distributions *growth curves*. The index flood method was also found to be an accurate and reproducible method for use at ungauged sites.[107]

At one time the British attempted to normalize the floods available at each site so that a large composite sample could be constructed to estimate their growth curves;[105] this approach was shown to be relatively inefficient.[69] Regional PWM index flood frequency estimation procedures that employ PWM and L moments, and often the GEV or Wakeby distributions, have been studied.[71,81,91,112,162] These results demonstrate that L-moment/GEV index flood procedures should in practice with appropriately defined regions be reasonably robust and more accurate than procedures that attempt to estimate two or more parameters with the short records often available at many sites. Outlined below is the L-moment/GEV version of the algorithm initially proposed by Landwehr, Matalas, and Wallis (personal communication, 1978), and popularized by Wallis and others.[69,160,162]

Let there be K sites in a region with records $[x_t(k)]$, $t = 1, \ldots, n_k$, and $k = 1, \ldots, K$. The L-moment/GEV index-flood procedure is

1. At each site k compute the three L-moment estimators $\hat{\lambda}_1(k)$, $\hat{\lambda}_2(k)$, $\hat{\lambda}_3(k)$ using the unbiased PWM estimators b_r.
2. To obtain a normalized frequency distribution for the region, compute the regional average of the normalized L moments of order $r = 2$ and 3:

$$\hat{\lambda}_r^R = \frac{\sum_{k=1}^{K} w_k [\hat{\lambda}_r(k)/\hat{\lambda}_1(k)]}{\sum_{k=1}^{K} w_k} \qquad \text{for } r = 2, 3 \tag{18.5.1}$$

   For $r = 1$, $\hat{\lambda}_1^R = 1$. Here w_k are weights; a simple choice is $w_k = n_k$, where n_k is the sample size for site k. However, weighting by the sample sizes when some sites have much longer records may give them undue influence. A better choice which limits the weight assigned to sites with longer records is

$$w_k = \frac{n_k n_R}{n_k + n_R}$$

   where n_k are the sample sizes and $n_R \approx 25$; the optimal value of the weighting parameter n_R depends on the heterogeneity of a region.[138,141]
3. Using the average normalized L moments $\hat{\lambda}_1^R$, $\hat{\lambda}_2^R$, and $\hat{\lambda}_3^R$ in Eqs. (18.2.22) and (18.2.23), determine the parameters and quantiles $\hat{x}_p^R$ of the normalized regional GEV distribution.
4. The estimator of the $100p$ percentile of the flood distribution at any site k is

$$\hat{x}_p(k) = \hat{\lambda}_1^k \, \hat{x}_p^R \tag{18.5.2}$$

   where λ_1^k is the at-site sample mean for site k:

$$\hat{\lambda}_1^k = \frac{1}{n_k} \sum_{t=1}^{n_k} x_t(k)$$

Of value is an estimate of the precision of flood quantiles obtained with Eq. (18.5.2). Across the region of interest, let the variance of the differences $\tilde{x}_p^S - \hat{x}_p^R$ between the actual normalized quantile $\tilde{x}_p^S$ for a random site and the average regional estimator $\hat{x}_p^R$ be denoted ϑ^2; ϑ^2 describes the heterogeneity of a region. Then the variance of the error associated with the flood quantile estimator $\hat{x}_p$, equal to $\hat{\lambda}_1\, \hat{x}_p^R$ for at-site sample mean $\hat{\lambda}_1$, can be written

$$\mathrm{Var}\,(\hat{x}_p) = E[\hat{\lambda}_1\, \hat{x}_p^R - \lambda_1\, \tilde{x}_p^S]^2 = \mathrm{Var}\,(\hat{\lambda}_1)\, E[(\hat{x}_p^R)^2] + (\lambda_1)^2\, \vartheta^2 \qquad (18.5.3)$$

The expected error in $\hat{x}_p$ is a combination of sampling error in site k's sample mean

$$\mathrm{Var}\,(\hat{\lambda}_1) = \mathrm{Var}\,[\bar{x}(k)] = \frac{\sigma_x^2}{n_k}$$

and the precision ϑ^2 of the regional flood quantile $\hat{x}_p^R$ as an estimator of the normalized quantile $\tilde{x}_p^S$ for a site in the region. In practice ϑ^2 is generally difficult to estimate. The generalized least squares regional regression methodology in Sec. 18.5.2 addresses the relevant issues and can provide a useful estimator.

A key to the success of the index flood approach is identification of reasonably similar sets of basins to keep the error in the regional quantiles ϑ^2 small.[94] Basins can be grouped geographically, as well as by physiographic characteristics including drainage area and elevation. Regions need not be geographically contiguous. Each site can potentially be assigned its own unique region consisting of sites with which it is particularly similar,[17] or regional regression equations can be derived to compute normalized regional quantiles as a function of a site's physiographic characteristics, or other statistics.[120]

For regions which exhibit a large ϑ^2, or when the record length for a site is on the order of 40 or more, then a two-parameter index flood procedure that uses the regional value of κ with at-site estimates of the GEV distribution's ξ and α parameters becomes attractive.[94] Chowdhury et al.[22] provide goodness-of-fit tests to assess whether a particular dimensionless regional GEV distribution, or a specified regional κ, is consistent with the data available at a questionable site.

18.5.2 Regional Regression

Regression can be used to derive equations to predict the values of various hydrologic statistics (including means, standard deviations, quantiles, and normalized regional flood quantiles) as a function of physiographic characteristics and other parameters. Such relationships are needed when little or no flow data are available at or near a site. Figure 18.5.1 illustrates the estimated prediction errors for regression models of low-flow, mean annual flows, and flood flows in the Potomac River Basin in the United States. Regional regression models have long been used to predict flood quantiles at ungauged sites, and in a nationwide test this method did as well or better than more complex rainfall-runoff modeling procedures.[107]

Consider the traditional log-linear model for a statistic y_i which is to be estimated by using watershed characteristics such as drainage area and slope:

$$y_i = \alpha + \beta_1 \log\,(\text{area}) + \beta_2 \log\,(\text{slope}) + \ldots + \epsilon \qquad (18.5.4)$$

A challenge in analyzing this model and estimating its parameters with available records is that one only obtains sample estimates, denoted $\hat{y}_i$, of the hydrologic statistics y_i. Thus the observed error ϵ is a combination of: (1) the time-sampling error

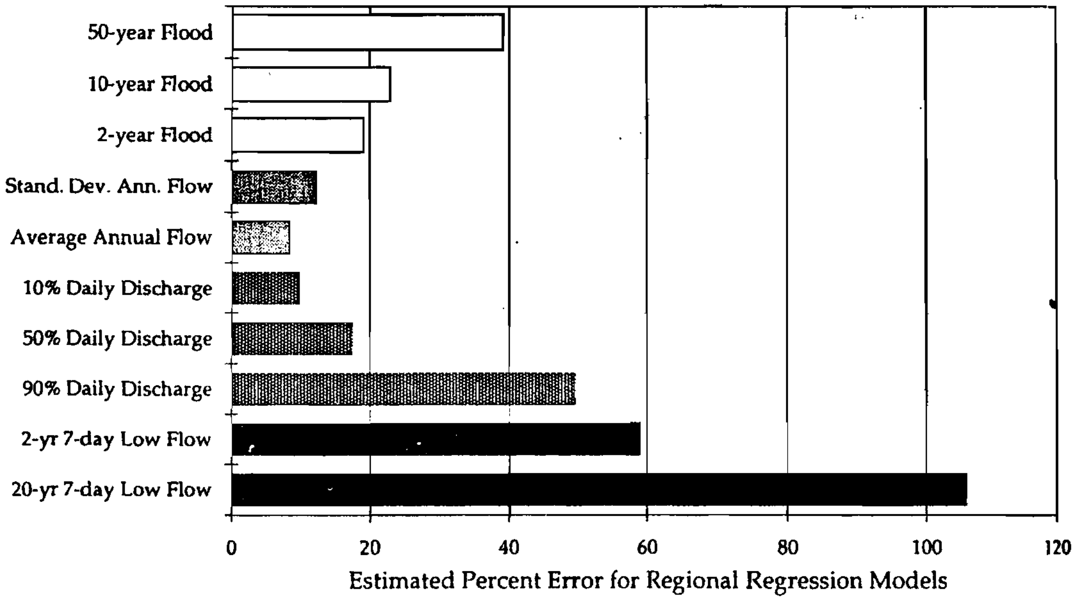

FIGURE 18.5.1 Percentage error for regional regression estimators of different statistics in the Potomac River Basin in the United States. *(From Ref. 142.)*

in sample estimators of y_i (these errors at different sites can be cross-correlated if the records are concurrent) and (2) underlying model error (lack of fit) due to failure of the model to exactly predict the true value of the y_i's at every site. Often these problems have been ignored and standard ordinary least squares (OLS) regression has been employed.[142] Stedinger and Tasker[130-132] develop a specialized generalized least squares (GLS) regression methodology to address these issues. Advantages of the GLS procedure include more efficient parameter estimates when some sites have short records, an unbiased model-error estimator, and a better description of the relationship between hydrologic data and information for hydrologic network analysis and design;[130,141] see also Sec. 17.4.8. Examples are provided by Tasker and Driver,[140] Vogel and Kroll,[156] and Potter and Faulkner.[111]

18.5.3 Record Augmentation and Extension

One can fill in missing observations in a short record by using a longer nearby record with which observations in the short record are highly correlated. Such cross-correlation can be used to fill in a few scattered missing observations, to extend the shorter record, or to improve estimates of the mean and variance of the events at the short-record site. For this third purpose it is not necessary to actually construct the extended series; one needs only the improved estimates of the moments. This idea is called *record augmentation* (Ref. 97, Ref. 105, App. 7 in Ref. 79).

Let x and y denote the flow record at the long- and short-record sites, respectively; let subscript 1 denote sample means and variances calculated for the period of concurrent record and subscript 2 denote the sample mean and variance for the long-record x site calculated using only the observations for which there is no corresponding y. The Matalas-Jacobs augmented-record estimator of the mean is

$$\hat{\mu}_Y = \bar{y}_1 + \frac{n_2}{n_1 + n_2} b(\bar{x}_2 - \bar{x}_1) \qquad n_1 \geq 4 \tag{18.5.5}$$

where n_1 is the number of concurrent observations and n_2 is the number of additional observations available at the x site. Their estimator of the variance is essentially

$$\hat{\sigma}_Y^2 = s_{y1}^2 + \frac{n_2}{n_1 + n_2} b^2 (s_{x2}^2 - s_{x1}^2) \qquad n_1 \geq 6 \tag{18.5.6}$$

except for several negligible adjustments; here

$$b = \hat{\rho}_{xy} \frac{s_{y1}}{s_{x1}} \tag{18.5.7}$$

is the standard linear regression estimator of change in y from a change in x. Equation (18.5.5) is relatively effective at improving estimates of the mean when the cross-correlation is 0.70 or greater; Eq. (18.5.6) transfers less information about the variance, which generally requires a cross-correlation of at least 0.85 to be worthwhile.[151] If the observations are serially correlated, considerably less information is transferred.[137,157]

In some cases one actually wants to create a longer series that will be used in simulation or archived as described in Sec. 17.4.10. In such cases it would be preferable if the extended series y_t had the same variance as the original series and was not smoothed by the process of regressing one record on another. This idea of *record extension* is developed in Refs. 64, 65, and 151 and, for the multivariate case with cross-correlation, in Ref. 50.

18.6 PARTIAL DURATION SERIES, MIXTURES, AND CENSORED DATA

This section discusses situations where data describing hydrologic events are not a simple series of annual values. Partial duration series and mixture models discussed in Secs. 18.6.1 and 18.6.2 describe hydrologic events by more than an average or a single annual maximum or minimum. These are examples of the idea of stochastic structure discussed in the introduction to Sec. 18.5. Section 18.6.3 discusses methods for dealing with censored data sets that occur when some observations fall below a recording threshold.

18.6.1 Partial Duration Series

Two general approaches are available for modeling flood, rainfall, and many other hydrologic series. Using an *annual maximum series,* one considers the largest event in each year; using a *partial duration series* (PDS) or peaks-over-threshold (POT) approach, the analysis includes all peaks above a truncation or threshold level. An objection to using annual maximum series is that it employs only the largest event in each year, regardless of whether the second largest event in a year exceeds the largest events of other years. Moreover, the largest annual flood flow in a dry year in some arid or semiarid regions may be zero, or so small that calling them floods is misleading.

Partial duration series analyses avoid such problems by considering all independent peaks which exceed a specified threshold. Fortunately one can estimate annual exceedance probabilities from the analysis of PDS with Eq. (18.6.4), below, or empirical relationships.[63] Arguments in favor of PDS are that relatively long and reliable PDS records are often available, and if the arrival rate for peaks over the threshold is large enough (1.65 events/year for the Poisson arrival with exponential exceedance model), PDS analyses should yield more accurate estimates of extreme quantiles than the corresponding annual-maximum frequency analyses.[105,118,145] Still, a drawback of PDS analyses is that one must have criteria to identify only independent peaks (and not multiple peaks corresponding to the same event); thus PDS analysis can be more complicated than analyses using annual maxima.

Partial duration models are applicable to modeling flood or rainfall events that exceed some threshold depth, or the occurrence of runoff carrying non-point-pollution loads. Partial duration models, perhaps with parameters that vary by season, are often used to estimate expected damages from hydrologic events when more than one damage-causing event can occur in a season or within a year.[108]

Two issues arise in modeling PDS. First, one must model the arrival rate of events larger than the threshold level; second, one must model the magnitudes of those events. For example, a Poisson distribution is often used to model the arrival of events, and an exponential distribution to describe the magnitudes of peaks which exceed the threshold.[16] For large-return-period events, the actual probabilistic model for arrivals is not important, provided different models yield the same average number of arrivals per year.[27,105]

There are several general relationships between the probability distribution for annual maximum and the frequency of events in a partial duration series. For a PDS, let λ be the arrival rate, equal to the average number of events per year larger than a threshold x_0; let $G(x)$ be the probability that events when they occur are less than x, and thus falls in the range (x_0, x). Then the arrival rate for any level x, with $x \geq x_0$, is

$$\lambda^* = \lambda[1 - G(x)] \tag{18.6.1}$$

The cdf $F_a(x)$ for the corresponding annual maximum series is the probability that the annual maximum for a year will not exceed x. For independent events, the probability of *no exceedances of* x over a 1-year period is given by the Poisson distribution, so that

$$F_a(x) = \exp(-\lambda^*) = \exp\{-\lambda[1 - G(x)]\} \tag{18.6.2}$$

[This relationship can be derived by dividing a year into m intervals, each with arrival rate λ^*/m. Then for small λ^*/m, the probability of no arrivals in a year is essentially $(1 - \lambda^*/m)^m$. Equation (18.6.2) is obtained in the limit as $m \to \infty$.]

Equation 18.6.2 reveals the relationship between the cdf for the annual maximums, and the arrival rate of and distribution for partial duration peaks. If the annual exceedance probability $1 - F_a(x)$ is denoted $1/T_a$, for an annual return period T_a (denoted as T elsewhere in the chapter) and the corresponding exceedance probability $[1 - G(x)]$ for level x in the partial duration series is denoted q_e, then Eq. (18.6.2) can be written

$$\frac{1}{T_a} = 1 - \exp(-\lambda\, q_e) = 1 - \exp\left(-\frac{1}{T_p}\right) \tag{18.6.3a}$$

where $T_p = 1/\lambda q_e$ is the average return period for level x in the PDS. Equation (18.6.3a) can be solved for T_p to obtain

$$T_p = -\frac{1}{\ln(1 - 1/T_a)} \tag{18.6.3b}$$

T_p is less than T_a because more than one event can occur per year in a PDS.

Equation (18.6.3*a*) transforms the *average arrival rate* λq_e for events larger than x into the *annual exceedance probability* $1/T_a$ in the annual maximum series. For levels x with $T_a > 10$, corresponding to infrequent events, the annual exceedance probability $1/T_a$ essentially equals the average arrival rate $\lambda q_e = \lambda[1 - G(x)]$ for the PDS, so that $T_a = T_p$ (Ref. 93). [See also Eq. (18.10.1).]

Consider a useful application of Eq. (18.6.2). Suppose a generalized Pareto distribution (Sec. 18.2.4) describes the distribution $G(x)$ of the magnitude of events larger than a threshold x_0:

$$G(x) = 1 - \left[1 - \kappa\left(\frac{x - x_0}{\alpha}\right)\right]^{1/\kappa} \qquad \text{for } \kappa \neq 0 \qquad (18.6.4)$$

For positive κ this cdf has upper bound $x_{\max} = x_0 + \alpha/\kappa$; for $\kappa < 0$, an unbounded and thick-tailed distribution results; $\kappa = 0$ yields a two-parameter exponential distribution. Substitution of Eq. (18.6.4) for $G(\cdot)$ into Eq. (18.6.2) yields a GEV distribution for the annual maximum series greater than x_0 if $\kappa \neq 0$:[34,70,125]

$$F_a(x) = \exp\left[-\left(1 - \kappa\frac{x - \xi}{\alpha^*}\right)^{1/\kappa}\right] \qquad \kappa \neq 0 \qquad (18.6.5a)$$

and a Gumbel distribution for $\kappa = 0$:[16]

$$F_a(x) = \exp\left[-\left(\frac{x - \xi}{\alpha}\right)\right] \qquad (18.6.5b)$$

when $x \geq x_0$; the transformed parameters ξ and α^* are defined by

$$\begin{aligned} \xi &= x_0 + \frac{\alpha(1 - \lambda^{-\kappa})}{\kappa} \qquad \alpha^* = \alpha\lambda^{-\kappa} \qquad && \kappa \neq 0 \\ \xi &= x_0 + \alpha \ln(\lambda) && \kappa = 0 \end{aligned} \qquad (18.6.6)$$

This general Poisson-Pareto model is a flexible and physically reasonable model of many phenomena. It has the advantage that regional estimates of the GEV distribution's shape parameter κ from annual maximum and PDS analyses can be used interchangeably.

In practice the arrival rate λ can simply be estimated by the average number of exceedances of x_0 per year. For either the exponential or generalized Pareto distributions in Table 18.1.2 for $G(x)$, the lower bound (denoted ξ in Table 18.1.2) equals x_0. The other parameters in Eqs. (18.6.5) and (18.6.6) can be estimated by substituting sample estimators into the inverse of the relationships in Table 18.1.2:

$$\begin{aligned} &\text{For } \kappa \neq 0: \qquad \kappa = \frac{\mu - x_0}{\lambda_2} - 2; \ \alpha = (\mu - x_0)(1 + \kappa) \\ &\text{For fixed } \kappa = 0: \qquad \frac{1}{\beta} = \alpha = \mu - x_0 \end{aligned} \qquad (18.6.7)$$

where $\mu = \lambda_1$ is the mean of X, λ_2 is the second L moment, and β is the exponential distribution's scale parameter in Tables 18.1.2 and 18.2.1.

18.6.2 Mixtures

A common problem in hydrology is that annual maximum series are composed of events that may arise from distinctly different processes. Precipitation series may correspond to different storm types in different seasons (such as summer thunderstorms, winter frontal storms, and remnants of tropical hurricanes). Floods arising from different types of precipitation events, or from snow melt, may have distinctly different probability distributions.[168]

The annual maximum series M can be viewed as the maximum of the maximum summer event S and the maximum winter event W:

$$M = \max(S, W) \tag{18.6.8}$$

Here S and W may be defined by a rigidly specified calendar period, a loosely defined climatic period, or the physical characteristics of the phenomena.

Let the cdf of the S and W variables be $F_S(s)$ and $F_W(w)$. Then, if the magnitudes of the summer and winter events are *statistically independent*, meaning that knowing one has no effect on the probability distribution of the other, the cdf for M is

$$F_M(m) = P[M = \max(S, W) \leq m] = F_S(m)\, F_W(m) \tag{18.6.9}$$

because M will be less than m only if *both* S and W are less than m. If two or more independent series of events contribute to an annual maximum, the distribution of the maximum is the product of their cdfs.

An important question is when it is advisable to model several different component precipitation or flood series separately, and when it is better to model the composite annual maximum series directly. If several series are modeled, then more parameters must be estimated, but more data are available than if the annual maximum series (or the partial duration series) for each type of event is employed. Fortunately, the distributions of large events caused by different mechanisms can be relatively similar.[62] Modeling the component series separately is most attractive when the annual maximum series is composed of components with distinctly different distributions which are individually easy to model because classical two-parameter Gumbel or lognormal distributions describe them well, and such a simple model provides a poor description of the composite annual maximum series.

18.6.3 Analysis of Censored Data

In some water-quality investigations, a substantial portion of reported values of many contaminants is below limits of detection. Likewise, low-flow and sometimes flood-flow observations are rounded to or reported as zero. Such data sets are called *censored samples* because it is as if the values of observations in a *complete sample* that fell above or below some level were removed, or censored. Several approaches are available for analysis of censored data sets, including probability plots and probability-plot regression, weighted-moment estimators, maximum likelihood estimators, and conditional probability models.[55,57,61] See also Section 17.5.

Probability-plot methods for use with censored data are discussed below. They are relatively simple and efficient when the majority of values are observed, and unobserved values are known to be below (above) some detection limit or perception threshold which serves as an upper (lower) bound. In such cases, probability-plot regression estimators of moments and quantiles are as accurate as maximum likelihood estimators, and almost as good as estimators computed with complete samples.[33,60] Partial PWMs are the expectation of $xF(x)^r$ for x values above a threshold; they are conceptually similar to probability-plot regression estimators and provide a useful alternative for fitting some distributions.[164]

Weighted moment estimators are used in flood frequency analyses with data sets that include both a complete gauged record and a historical flood record consisting of all events above a perception threshold.[79,133,165] (See Sec. 18.7.4.) Weighted moment estimators weight values above and below the threshold levels so as to obtain moment estimators consistent with a complete sample. These methods are reasonable when a substantial fraction of the observations remain after censoring (at least 10 percent), and a value is either observed accurately or falls below a threshold and thus is censored.

Maximum likelihood estimators are quite flexible, and are more efficient than plotting and weighted moment estimators when the frequency with which a threshold was exceeded represents most of the sample information.[23,133] They allow the recorded values to be represented by exact values, ranges, and various thresholds that either were or were not exceeded at various times; this can be particularly important with historical flood data sets because the magnitudes of many historical floods are not recorded precisely, and it may be known that a threshold was never crossed or was crossed at most once or twice in a long period.[23] (See Sec. 18.7.4.) In these cases maximum likelihood estimators are perhaps the only approach that can make effective use of the available information.[33]

Conditional probability models are appropriate for simple cases wherein the censoring occurs because small observations are recorded as zero, as often happens with low-flow and some flood records. An extra parameter describes the probability p_0 that an observation is zero. A continuous distribution $G(x)$ is derived for the strictly positive nonzero values of X; the parameters of the cdf G can be estimated by any procedure appropriate for complete samples. The unconditional cdf $F(x)$ for any value $x > 0$ is then

$$F(x) = p_0 + (1 - p_0)\, G(x) \tag{18.6.10}$$

Equations (18.7.6) to (18.7.8) provide an example of such a model.

Plotting Positions for Censored Data. Section 18.3.2 discusses plotting positions useful for graphical fitting methods, as well as visual displays of data. Suppose that among n samples a detection limit or perception threshold is exceeded by water-quality observations or flood flows r times. The natural estimator of the exceedance probability q_e of the perception threshold is r/n. If the r values which exceeded the threshold are indexed by $i = 1, \ldots, r$, reasonable plotting positions approximating the exceedance probabilities within the interval $(0, q_e)$ are

$$q_i = q_e\left(\frac{i-a}{r+1-2a}\right) = \frac{r}{n}\left(\frac{i-a}{r+1-2a}\right) \tag{18.6.11}$$

where a is a value from Table 18.3.1. For $r \gg (1 - 2a)$, q_i is indistinguishable from $(i - a)/(n + 1 - 2a)$ for a single threshold. Reasonable choices for a generally make little difference to the resulting plotting positions.[60]

The idea of an exceedance probability for the threshold is important when detection limits change over time, generating multiple thresholds. In such cases, an exceedance probability should be estimated for each threshold so that a consistent set of plotting positions can be computed for observations above, below, or between thresholds.[60,66] For example, consider a historical flood record with an h-year historical period in addition to a complete s-year gauged flood record. Assume that during the total $n = (s + h)$ years of record, a total of r floods exceeded a perception threshold (censoring level) for historical floods. These r floods can be plotted by using Eq. (18.6.11).

Let e be the number of gauged-record floods that exceeded the threshold and hence are counted among the r exceedances of that threshold. Plotting positions within $(q_e, 1)$ for the remaining $(s - e)$ below-threshold gauged-record floods are

$$q_j = q_e + (1 - q_e)\left(\frac{j - a}{s - e + 1 - 2a}\right) \tag{18.6.12}$$

for $j = 1$ through $s - e$, where again a is a value from Table 18.3.1. This approach directly generalizes to several thresholds.[60,66] For records with an r of only 1 or 2, Ref. 166 proposes fitting a parametric model to the gauged record to estimate q_e; these are cases when nonparametric estimators of q_e and q_i in Eq. (18.6.11) are inaccurate,[64] and MLEs are particularly attractive for parameter estimation.[23,133]

Probability-Plot Regression. Probability-plot regression has been shown to be a robust procedure for fitting a distribution and estimating various statistics with censored water-quality data.[60] When water-quality data is well-described by a lognormal distribution, available values $\log [X_{(1)}] \geq \ . \ . \ . \ \geq \log [X_{(r)}]$ can be regressed upon $\Phi^{-1}[1 - q_i]$ for $i = 1, \ldots, r$, where the r largest observation in a sample of size n are available; and q_i are their plotting positions. If regression yields constant m and slope s, a good estimator of the pth quantile is

$$\hat{x}_p = 10^{m + s z_p} \tag{18.6.13}$$

for cumulative probability $p > (1 - r/n)$. To estimate sample means and other statistics, one can fill in the missing observations as

$$X_{(i)} = 10^{y(i)} \qquad \text{for } i = r + 1, \ldots, n \tag{18.6.14}$$

where $y(i) = m + s\ \Phi^{-1}(1 - q_i)$ and an approximation for Φ^{-1} is given in Eq. (18.2.3). Once a complete sample is constructed, standard estimators of the sample mean and variance can be calculated, as can medians and ranges. By filling in the missing small observations, and then using complete-sample estimators of statistics of interest, the procedure is made relatively insensitive to the assumption that the observations actually have a lognormal distribution.[60]

18.7 FREQUENCY ANALYSIS OF FLOODS

Lognormal, Pearson type 3, and generalized extreme value distributions are reasonable choices for describing flood flows using the fitting methods described in Sec. 18.2. However, as suggested in Sec. 18.3.3, it is advisable to use regional experience to select a distribution for a region and to reduce the number of parameters estimated for an individual site. This section describes sources of flood flow data and particular procedures adopted for flood flow frequency analysis in the United States and the United Kingdom, and discusses the use of historical flood flow information.

18.7.1 Selection of Data and Sources

A convenient way to find information on United States water data is through the U.S. National Water Data Exchange (NAWDEX) assistance centers. [For information contact NAWDEX, U.S. Geological Survey (USGS), 421 National Center, Reston, Va. 22092; tel. 703-648-6848.] Records are also published in annual U.S. Geological Survey water data reports. Computerized records are stored in the National Water Data Storage and Retrieval System (WATSTORE). Many of these records (climate data, daily and annual maximum stream flow, water-quality parameters) have been put on compact disc read-only memories (CD-ROMs) sold by EarthInfo Inc. (5541 Central Ave., Boulder, Colo. 80301; tel. 303-938-1788; fax 303-938-8183) so that the data can be accessed directly with personal computers. The WATSTORE peak-flow records contain annual maximum instantaneous flood-peak discharge and stages, and dates of occurrence as well as associated partial duration series for many sites. USGS offices also publish sets of regression relationships (often termed *state equations*) for predicting flood and low-flow quantiles at ungauged sites in the United States.

18.7.2 Bulletin 17B Frequency Analysis

Recommended procedures for flood-frequency analyses by U.S. federal agencies are described in Bulletin 17B.[79] Bulletin no. 15, "A Uniform Technique for Determining Flood Flow Frequencies," released in December 1967, recommended the log-Pearson type 3 distribution for use by U.S. federal agencies. The original Bulletin 17, released in March 1976, extended Bulletin 15 and recommended the log-Pearson type 3 distribution with a regional estimator of the log-space skew coefficient. Bulletin 17A followed in 1977. Bulletin 17B was issued in September 1981 with corrections in March 1982. Thomas[143] describes the development of these procedures.

The Bulletin 17 procedures were essentially finalized in the mid-1970s, so they did not benefit from subsequent advances in multisite regionalization techniques. Studies in the 1980s demonstrated that use of reasonable index flood procedures should provide substantially better flood quantile estimates, with perhaps half the standard error.[91,112,162] Bulletin 17 procedures are much less dependent on regional multisite analyses than are index flood estimators, and Bulletin 17 is firmly established in the United States, Australia, and other countries. However, App. 8 of the bulletin does describe a procedure for weighting the bulletin's at-site estimator and a regional regression estimator of the logarithms of a flood quantile by the available record length and the effective record length, respectively. The resulting weighted estimator reflects a different approach to combining regional and at-site information than that employed by index flood procedures.

Bulletin 17B recommends special procedures for zero flows, low outliers, historic peaks, regional information, confidence intervals, and expected probabilities for estimated quantiles. This section describes only major features of Bulletin 17B. The full Bulletin 17B procedure is described in that publication and is implemented in the HECWRC computer program discussed in Sec. 18.11.

The bulletin describes procedures for computing flood flow frequency curves using annual flood series with at least 10 years of data. The recommended technique fits a Pearson type 3 distribution to the common base 10 logarithms of the peak discharges. The flood flow Q associated with cumulative probability p is then

$$\log (Q_p) = \overline{X} + K_p S \tag{18.7.1}$$

where $\overline{X}$ and S are the sample mean and standard deviation of the base 10 logarithms, and K_p is a frequency factor which depends on the skew coefficient and selected exceedance probability; see Eq. (18.2.28) and discussion of the log-Pearson type 3 distribution in Sec. 18.2.3. The mean, standard deviation, and skew coefficient of station data should be computed by Eq. (18.1.8), where X_i are the base 10 logarithms of the annual peak flows. Section 18.1.3 discusses advantages and disadvantages of logarithmic transformations.

The following sections discuss three major features of the bulletin: generalized skew coefficients, outliers, and the conditional probability adjustment. Expected probability adjustments are also discussed. Confidence intervals for Pearson distributions with known and generalized skew coefficient estimators are discussed in Sec. 18.4.3. Use of historical information is discussed in Sec. 18.7.4, mixed populations in Sec. 18.6.2, and record augmentation in Sec. 18.5.3.

Generalized Skew Coefficient. Because of the variability of at-site sample skew coefficients in small samples, the bulletin recommends weighting the station skew coefficient with a generalized coefficient of skewness, which is a regional estimate of the log-space skewness. In the absence of detailed studies, the generalized skew coefficient G_g for sites in the United States can be read from Plate I in the bulletin. Assuming that the generalized skew coefficient is unbiased and independent of the station skew coefficient, the mean square error (MSE) of the weighted estimate is minimized by weighting the station and generalized skew coefficients inversely proportional to their individual mean square errors:

$$G_w = \frac{G_s/\mathrm{MSE}(G_s) + G_g/\mathrm{MSE}(G_g)}{1/\mathrm{MSE}(G_s) + 1/\mathrm{MSE}(G_g)} \tag{18.7.2}$$

Here G_w is the weighted skew coefficient, G_s is the station skew coefficient, and G_g is the generalized regional estimate of the skew coefficient; MSE( ) is the mean square error of the indicated variable. When generalized regional skew coefficients are read from its Plate I, Bulletin 17 recommends using $\mathrm{MSE}(G_g) = 0.302$.

From Monte Carlo experiments,[159] the bulletin recommends that $\mathrm{MSE}(G_s)$ be estimated using the bulletin's Table 1, or an expression equivalent to

$$\mathrm{MSE}(G_s) = \frac{10^{a+b}}{n^b} \tag{18.7.3}$$

where

$$\begin{aligned} a &= -0.33 + 0.08|G_s| && \text{if} \quad |G_s| \le 0.90 \\ &= -0.52 + 0.30|G_s| && \text{if} \quad |G_s| > 0.90 \\ b &= 0.94 - 0.26|G_s| && \text{if} \quad |G_s| \le 1.50 \\ &= 0.55 && \text{if} \quad |G_s| > 1.50 \end{aligned}$$

$\mathrm{MSE}(G_s)$ is essentially $5/n$ for small G_s and $10 \le n \le 50$. G_g should be used in place of G_s in Eq. (18.7.3) when estimating $\mathrm{MSE}(G_s)$ to avoid correlation between G_s and the estimate of $\mathrm{MSE}(G_s)$ (Ref. 138). McCuen[98] and Tasker and Stedinger[138] discuss the development of skew-coefficient maps, and regression estimators of G_g and $\mathrm{MSE}(G_g)$.

Outliers. Bulletin 17B defines outliers to be "Data points which depart significantly from the trend of the remaining data." In experimental statistics an outlier is often a rogue observation which may result from unusual conditions or observational or recording error; such observations are often discarded. In this application low outliers are generally valid observations, but because Bulletin 17 uses the logarithms of the observed flood peaks to fit a two-parameter distribution with a generalized skew coefficient, one or more unusual low-flow values can distort the entire fitted frequency distribution. Thus detection of such values is important and fitted distributions should be compared graphically with the data to check for problems.

The thresholds used to define high and low outliers in log space are

$$\bar{X} \pm K_n S \tag{18.7.4}$$

where $\bar{X}$ and S are the mean and standard deviations of the logarithms of the flood peaks, excluding outliers previously detected, and K_n is a critical value for sample size n. For normal data the largest observation will exceed $\bar{X} + K_n S$ with a probability of only 10 percent; thus Eq. (18.7.4) is a one-sided outlier test with a 10 percent significance level. Values of K_n are tabulated in Bulletin 17B; for $5 \leq n \leq 150$, K_n can be computed by using the common base 10 logarithm of the sample size

$$K_n = -0.9043 + 3.345 \sqrt{\log(n)} - 0.4046 \log(n) \tag{18.7.5}$$

Flood peaks identified as low outliers are deleted from the record and a conditional probability adjustment is recommended. High outliers are retained unless historical information is identified showing that such floods are the largest in an extended period.

Conditional Probability Adjustment. A conditional probability procedure is recommended for frequency analysis at sites whose record of annual peaks is truncated by the omission of peaks below a minimum recording threshold, years with zero flow, or low outliers. The bulletin does not recommend this procedure when more than 25 percent of the record is below the truncation level. Section 18.6.3 discusses other methods.

Let $G(x)$ be the Pearson type 3 (P3) distribution fit to the r logarithms of the annual maximum floods *that exceeded the truncation level* and are included in the record, after deletions of zero, low outliers, and other events. If the original record spanned n years ($n > r$), then an estimator of the probability the truncation level is exceeded is

$$q_e = \frac{r}{n} \tag{18.7.6}$$

Flood flows exceeded with a probability $q \leq q_e$ in any year are obtained by solving

$$q = q_e[1 - G(x)] \tag{18.7.7}$$

to obtain

$$G(x) = 1 - \frac{q}{q_e} = 1 - q\left(\frac{n}{r}\right) \tag{18.7.8}$$

Bulletin 17 uses Eq. (18.7.8) to calculate the logarithms of flood flows which will be exceeded with probabilities of $q = 0.50$, 0.10, and 0.01. These three values are used to define a new Pearson type 3 distribution for the logarithms of the flood flows which reflects the unconditional frequency of above threshold values. The new Pearson type 3 distribution is defined by its mean M_a, variance S_a^2, and skew coefficient

G_a, which are calculated as

$$G_a = -2.50 + 3.12 \frac{\log (Q_{0.99}/Q_{0.90})}{\log (Q_{0.90}/Q_{0.50})}$$

$$S_a = \frac{\log (Q_{0.99}/Q_{0.50})}{K_{0.99} - K_{0.50}} \tag{18.7.9}$$

$$M_a = \log (Q_{0.50}) - K_{0.50}\, S_a$$

for log-space skew coefficients between -2.0 and $+2.5$. The Pearson type 3 distribution obtained with the moments in Eq. (18.7.9) should not be used to describe the frequency of flood flows below the median $Q_{0.50}$. Fitted quantiles near the threshold are likely to be particularly poor if the P3 distribution $G(x)$ fit to the above threshold values has a lower bound less than the truncation level for zeros and low outliers, which is thus a lower bound for x.

Expected Probability. A fundamental issue is what a hydrologist should provide when requested to estimate the flood flow exceeded with probability $q = 1/T$ using short flood flow records. It is agreed that one wants the flood quantile x_{1-q} which will be exceeded with probability q. An unresolved question is what should be the statistical characteristics of estimators $\hat{x}_{1-q}$. Most estimators in Sec. 18.2 yield $\hat{x}_{1-q}$ that are almost unbiased estimators of x_{1-q}:

$$E[\hat{x}_{1-q}] \approx x_{1-q} \tag{18.7.10}$$

and which have a relatively small variance or mean square error. However, an equally valid argument suggests that one wants $\hat{x}_{1-q}$ to be a value which in the future will be exceeded with probability q, so that

$$\mathrm{P}(X > \hat{x}_{1-q}) \approx q \tag{18.7.11}$$

when both X and $\hat{x}_{1-q}$ are viewed as random variables. If one had a very long record, these two criteria would lead to almost the same design value x_{1-q}. With short records they lead to different estimates because of the effect of the uncertainty in the estimated parameters.[8,113,127]

For normal samples, App. 11 in Bulletin 17B[79] (see also Ref. 20) provides formulas for the probabilities that the almost-unbiased estimator $\hat{x}_p = \bar{x} + z_p s$ of the $100p$ percentile will be exceeded. For $p = 0.99$ the formula is

$$\text{Average exceedance probability for } \hat{x}_{0.99} = 0.01\left(1 + \frac{26}{n^{1.16}}\right) \tag{18.7.12}$$

For samples of size 16, estimates of the 99 percentile will be exceeded with a probability of 2 percent on average. Bulletin 17B notes that for lognormal or log-Pearson distributions, the equations in its App. 11 can be used to make an expected probability adjustment.

Unfortunately, while the expected probability correction can eliminate the bias in the expected exceedance probability of a computed T-year event, the corrections would generally increase the bias in estimated damages calculated for dwellings and economic activities located at fixed locations in a basin.[4,127] This paradox arises because the estimated T-year flood is a (random) level computed by the hydrologist based on the fitted frequency distribution, whereas the expected damages are calculated for human and economic activities at fixed flood levels. Expected probability issues are related to Bayesian inference.[127]

18.7.3 British Frequency Analysis Procedures

The *Flood Studies Report*[105] contains hydrological, meteorological, and flood routing studies for the British Isles. The report concluded that the GEV distribution provided the best description of British and Irish annual maximum flood distributions and was recommended for general use in those countries. The three parameter P3 and LP3 distribution also described the data well (Ref. 105, pp. 241, 242).

A key recommendation was use of an index flood procedure. The graphically derived normalized regional flood distributions were summarized by dimensionless GEV distributions called *growth curves*. Reevaluation of the original method showed that the L-moment index flood procedure is to be preferred.[69] (See Sec. 18.5.1.) The report distinguishes between sites with less than 10 years of record, those with 10 to 25 years of record, and those with more than 25 years of record (Ref. 105, pp. 14 and 243):

Sites with $n < 10$ Years. The report recommends a regional growth curve with $\overline{Q}$ obtained from catchment characteristics (see Sec. 18.5.2), or at-site data extended if possible by correlation with data at other sites (see Sec. 18.5.3).

Sites with $10 \leq n \leq 25$ Years. Use either the index flood procedure with at-site data to estimate $\overline{Q}$ or, if the return period $T < 2n$, the Gumbel distribution.

Sites with $n > 25$ Years. Use either an index flood estimator with at-site data to estimate $\overline{Q}$ or, if the return period $T < 2n$, GEV distribution (see Sec. 18.2.2).

For $T > 500$. Use $\overline{Q}$ with a special country-wide growth curve.

18.7.4 Historical Flood Information

Available at-site systematic gauged records are the traditional and most obvious source of information on the frequency of floods, but they are of limited length. Another source of potentially valuable at-site information is historical and paleoflood records. Historical information includes written and other records of large floods left by human observers: newspaper accounts, letters, and flood markers. The term paleoflood information describes the many botanical and geophysical sources of information on large floods which are not limited to the locations of past human observations or recording devices.[6,25,134] Botanical data can consist of the systematic interpretation of tipped trees, scars, and abnormal tree rings along a water course providing a history of the frequency with which one or more thresholds were exceeded.[77,78] Recent advances in physical paleoflood reconstruction have focused on the use of slack-water deposits and scour lines, as indicators of paleoflood stages, and the absence of large flows that would have left such evidence; such physical evidence of flood stage along a water course has been used with radiocarbon and other dating techniques to achieve a relatively accurate and complete catalog of paleofloods in favorable settings with stable channels.[86]

Character of Information. Different processes can generate historical and physical paleoflood records. A flood leaving a high-water mark, or known to be the largest flood of record from written accounts, is the largest flood to have occurred in some period of time which generally extends back beyond the date at which that flood occurred.[66] In other cases, several floods may be recorded (or none at all), because they exceed some perception level defined by the location of dwellings and economic activities, and thus sufficiently disrupted people's lives for their occurrence to be noted, or for the resultant botanical or physical damage to document the event. In statistical terms, historical information represents a *censored sample* because only the largest floods are recorded, either because they exceeded a threshold of perception for the occupants of the basin, or because they were sufficiently large to leave physical evidence which was preserved. To correctly interpret such data, hydrologists should understand the mechanisms or reasons that historical, botanical, or geophysical records document that floods of different magnitudes either did, or did not, occur. The historical record should represent a complete catalog of all events that exceeded various thresholds so that it can serve as the basis for frequency analyses.

Estimation Procedures. A general discussion of estimation techniques with censored data is provided in Sec. 18.6.3, including plotting positions and curve fitting based on a graphical representation of systematic and historical flood peaks. Bulletin 17B[79] recommends a historically weighted moments procedure. A similar partial PWM method has been developed.[165] Curve fitting and weighted moments require that historical flood peaks above the perception level be assigned specific values. Even when the magnitudes of the few observed historical floods are available, historically weighted moments are not as efficient as maximum likelihood estimators.[92,133] The value of historical information using maximum likelihood estimation techniques is well-documented.[23,81,133] Maximum likelihood estimation is quite flexible and allows the historical record to be represented by thresholds that were not exceeded and by flood events whose magnitude is known only to have exceeded a threshold, to lie within some range, or which can be described by a precise value.[135]

18.8 FREQUENCY ANALYSIS OF STORM RAINFALL

The frequency of rainfall of various intensities and durations is used in the hydrologic design of structures that control storm runoff and floods, such as storm sewers, highway culverts, and dams. Precipitation frequency analysis typically provides rainfall accumulation values at a point for a specified exceedance probability and various durations. Basin-average rainfall values are usually developed from point rainfall by using a correction factor for basin areas greater than 10 mi^2 (25.9 km^2), as shown in Fig. 3.9.2.[100,103]

18.8.1 Selection of Data and Sources

United States precipitation data are published in *Climatological Data* and *Hourly Precipitation Data* by the National Oceanic and Atmospheric Administration (NOAA) from their National Climatic Data Center (NCDC); precipitation records and publications can be obtained directly from the center (NCDC, Federal Building, Asheville, NC 28801; tel. 704-259-0682; fax 704-259-0876). Climatic data have been put on CD-ROMs sold by EarthInfo Inc. (5541 Central Ave., Boulder, Colo. 80301; tel. 303-938-1788; fax 303-938-8183). NRC[104] discusses the availability and interpretation of United States rainfall data. Other national and regional agencies publish their own precipitation records.

The user of precipitation data should be aware of possible errors in data collection caused by wind effects, changes in the station environment, and observers. Users l check data for outliers and consistency. Interpretation of data is needed to nt for liquid precipitation versus snow equivalent and observation time differ- Stations submitting data to NCDC are expected to operate standard equip- and follow standard procedures with observations taken at standard times.[169] infall frequency analysis is usually based on annual maximum series or partial on series at one site (at-site analysis) or several sites (regional analysis). Since ll data are usually published for fixed time intervals, e.g., clock hours, they yield the true maximum amounts for the indicated durations. For example, ınnual maximum 24-h rainfalls for the United States are on the average 13 nt greater than annual maximum daily values corresponding to a fixed 24-h l.[63] Adjustment factors are usually employed with the results of a frequency sis of annual maximum series. Such factors depend on the number of observa- l reporting times within the duration of interest. (See Ref. 172, p. 5-36).

ıother source of data which has been used to derive estimates of the probable mum precipitation, and to a lesser extent for rainfall frequency analysis, is the Army Corps of Engineers catalog of extreme storms. The data collection and ssing were a joint effort of the U.S. Army Corps of Engineers and the U.S. her Bureau. Currently, a total of 563 storms, most of which occurred between and 1940, have been completed and published in Ref. 146; see also Refs. 104 123. There are problems associated with the use of this catalog for frequency sis. It may be incomplete because the criteria used for including a storm in the og are not well-defined and have changed. Also, the accuracy in the estimation e storm depths varies.

.2 Frequency Analysis Studies

Rainfall Frequency Atlas,[63] known as TP-40, provides an extended rainfall ıency study for the United States from approximately 4000 stations. The Gum- istribution (Sec. 18.2.2; see also Ref. 172) was used to produce the point precipi- n frequency maps of durations ranging from 30 min to 24 h and exceedance abilities from 10 to 1 percent. The report also contains diagrams for making ipitation estimates for other durations and exceedance probabilities. The U.S. ther Bureau, in a publication called TP-49,[149] published rainfall maps for dura- s of 2 to 10 days. Isohyetal maps (which partially supersede TP-40) for durations to 60 min are found in Ref. 46, known as HYDRO-35, and for 6 to 24 h for the ern United States in *NOAA Atlas 2.*[100] Examples of frequency maps can be found hap. 3.

or a site for which rainfall data are available, a frequency analysis can be per- ned. Common distributions for rainfall frequency analysis are the Gumbel, log- rson type 3, and GEV distributions with $\kappa < 0$, which is the standard distribution l in the British Isles.[105]

Maps presented in TP-40 and subsequent publications have been produced by rpolation and smoothing of at-site frequency analysis results. Regional frequency lysis, which uses data from many sites, can reduce uncertainties in estimators of eme quantiles (Refs. 15 and 161; see Sec. 18.5.1). Regional analysis requires ction of reasonably homogeneous regions. Schaefer[120] found that rainfall data in shington State have CV and γ which systematically vary with mean areal precipi- on. He used mean areal precipitation as an explanatory variable to develop a ıonal analysis methodology for a heterogeneous region, thereby eliminating ındary problems that would be introduced if subregions were defined.

Models of daily precipitation series (as opposed to annual maxima) are constructed for purposes of simulating some hydrologic systems. As Chap. 3 discusses, models of daily series need to describe the persistence of wet-day and dry-day sequences. The mixed exponent distribution, and the Weibull distribution with $k = 0.7$ to 0.8, have been found to be good models of daily precipitation depths on rainy days, though an exponential distribution has often been used.[122,173]

18.8.3 Intensity-Duration-Frequency Curves

Rainfall intensity-duration-frequency (IDF) curves allow calculation of the average design rainfall intensity for a given exceedance probability over a range of durations. IDF curves are available for several U.S. cities; two are shown in Fig. 18.8.1.[148] When an IDF curve is not available, or a longer data base is available than in TP-25 or TP-40, a hydrologist may need to perform the frequency analyses necessary to construct an IDF curve (see p. 456 in Ref. 20).

IDF curves can be described mathematically to facilitate calculations. For example, one can use

$$i = \frac{c}{t^e + f} \tag{18.8.1}$$

where i is the design rainfall intensity (inches per hour), t is the duration (minutes), c is a coefficient which depends on the exceedance probability, and e and f are coefficients which vary with location.[170] For a given return period, the three constants can be estimated to reproduce i for three different t's spanning a range of interest. For example, for a 1 in 10 year event, values for Los Angeles are $c = 20.3$, $e = 0.63$, and $f = 2.06$, while for St. Louis $c = 104.7$, $e = 0.89$, and $f = .9.44$.

More recently, generalized intensity-duration-frequency relationships for the United States have been constructed by Chen[19] using three depths: the 10-year 1-h rainfall (R_1^{10}), the 10-year 24-h rainfall (R_{24}^{10}), and the 100-year 1-h rainfall (R_1^{100}) from TP-40. These depths describe the geographic pattern of rainfall in terms of the depth-duration ratio (R_1^T/R_{24}^T) for any return period T, and the depth-frequency ratio (R_t^{100}/R_t^{10}) for any duration t. Chen's general rainfall IDF relation for the rainfall depth R_t^T in inches for any duration t (in minutes) and any return period T (in years) is

$$R_t^T = \frac{a_1 R_1^{10}\,[x - 1)\log(T_p/10) + 1]\left(\dfrac{t}{60}\right)}{(t + b_1)^{c_1}} \tag{18.8.2}$$

where $x = (R_1^{100}/R_1^{10})$, T_p is the return period for the partial duration series (equal to the reciprocal of the average number of exceedances per year), and a_1, b_1, and c_1 are coefficients obtained from Fig. 18.8.2 as functions of (R_1^{10}/R_{24}^{10}) with the assumption that this ratio does not vary significantly with T. Chen uses $T_p = -1/\ln(1 - 1/T)$ to relate T_p to the return period T for the annual maximum series (see Sec. 18.6.1); for $T > 10$ there is little difference between the two return periods. The coefficients obtained from Fig. 18.8.2 are intended for use with TP-40 rainfall quantiles.

For many design problems, the time distribution of precipitation (hyetograph) is needed. In the design of a drainage system the time of occurrence of the maximum rainfall intensity in relation to the beginning of the storm may be important. Design hyetographs can be developed from IDF curves following available procedures (see Chap. 3).

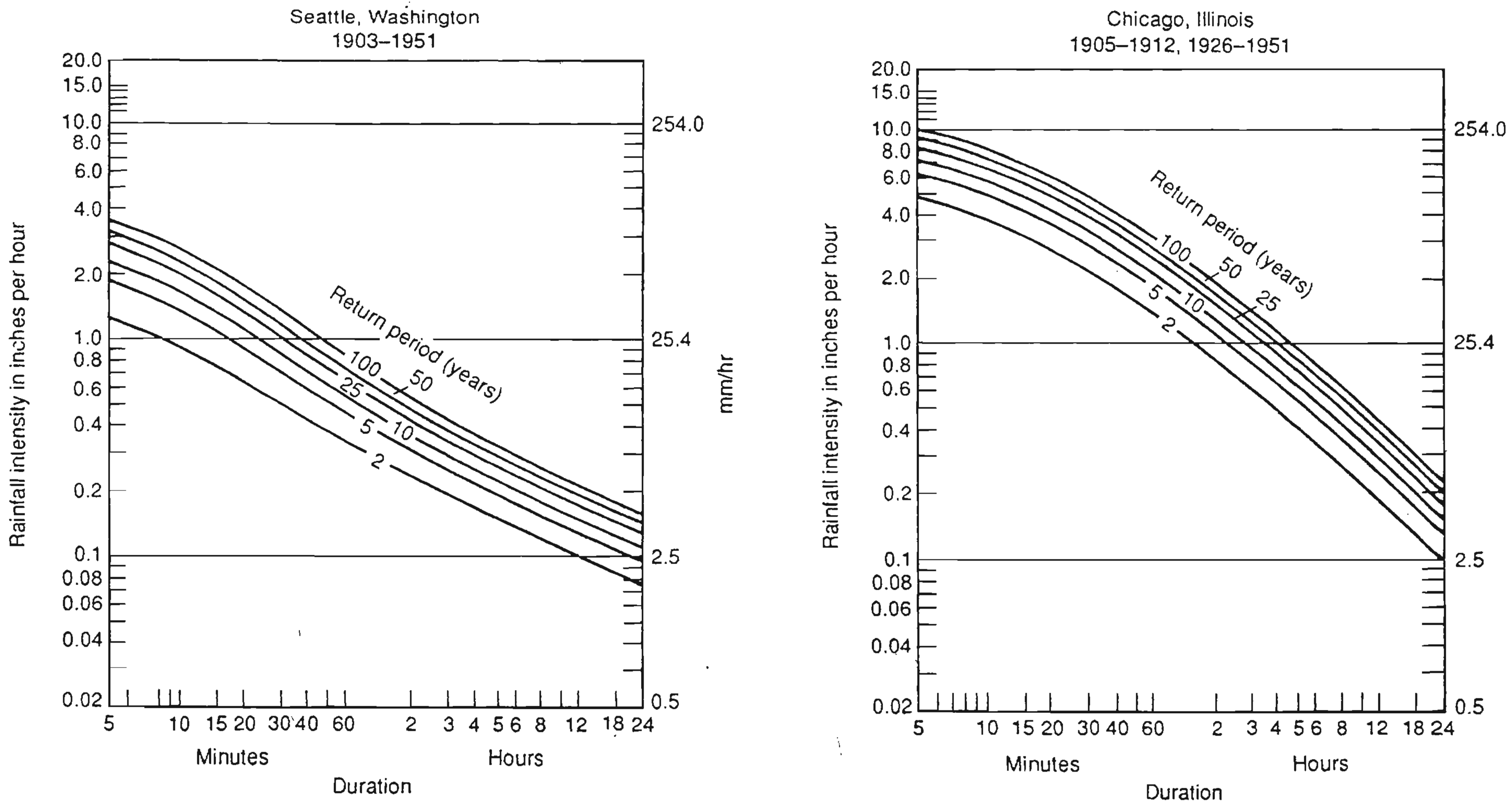

FIGURE 18.8.1 Typical intensity-duration-frequency curves. *(From Ref. 148.)*

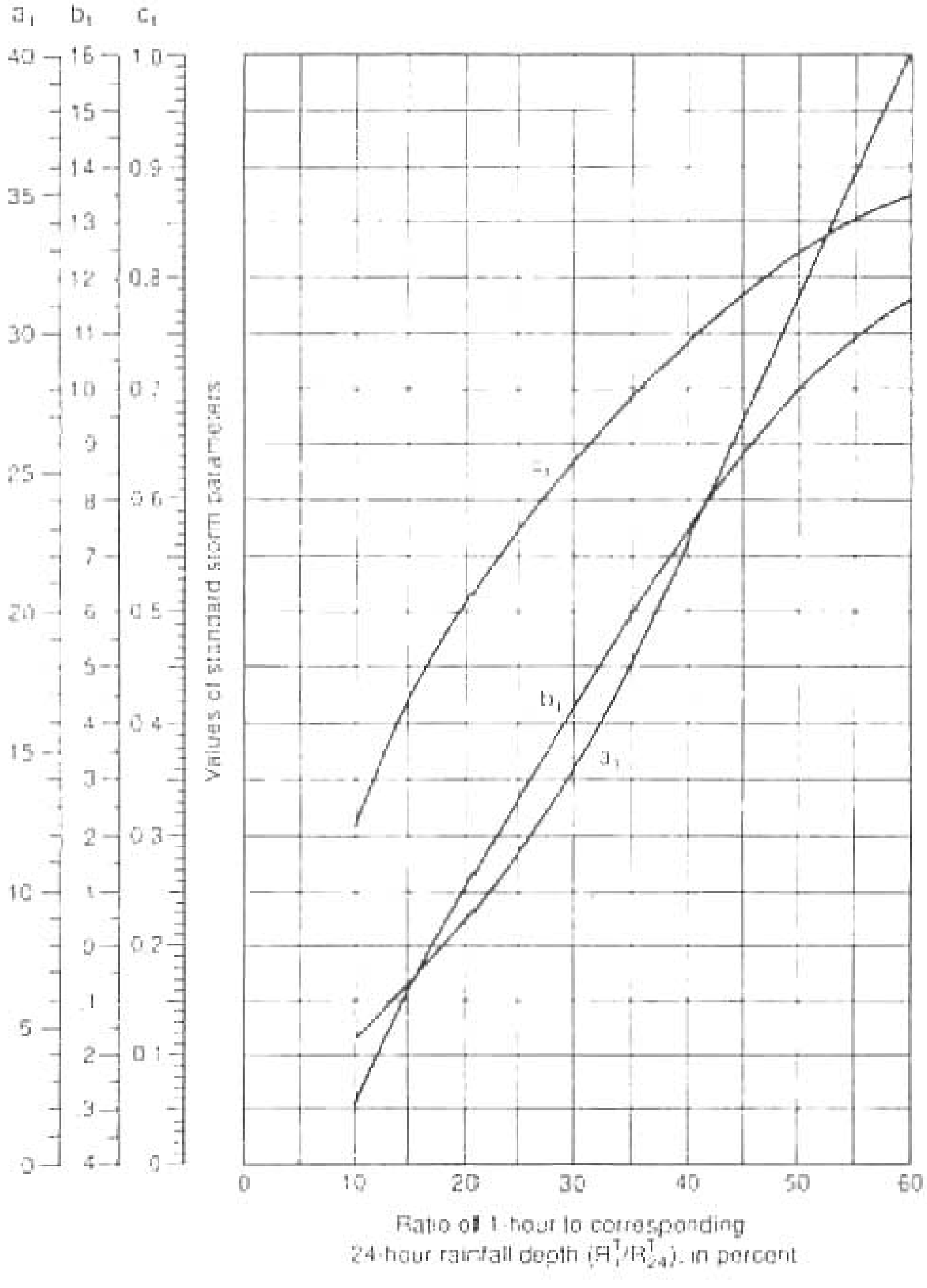

FIGURE 18.8.2 Relationship between standard storm parameters and ratio of 1-h to corresponding 24-h rainfall depth. *(Reproduced from Ref. 19 with permission.)*

18.8.4 Frequency Analysis of Basin Average Extreme Storm Depths

The stochastic storm transposition (SST) methodology has been developed for very low frequency rainfall (exceedance probabilities less than 1 in a 1000). SST provides estimates of the annual exceedance probability of the average catchment depth, which is the storm depth deposited over the catchment of interest. The estimate is based on regionalized storm characteristics and estimation of the joint probability distribution of storm characteristics and storm occurrences within a region.[42,44,106,171]

The SST method first selects a large climatologically homogeneous area (called the *storm transposition area*). Extreme storms of record within that area are employed to estimate the joint probability distribution of selected storm characteristics (such as maximum storm-center depth, storm shape parameters, storm orientation, storm depth spatial variability, etc.). Then, the probability distribution of the position of the storm centers within the transposition area is determined. This distribution is generally not uniform because the likelihood of storms of a given magnitude will vary across a region.[45] Integration over the distribution of storm characteristics and storm locations allows calculation of annual exceedance probabilities for various catchment depths. An advantage of the SST method is that it explicitly considers the morphology of the storms including the spatial distribution of storm depth and its relation to the size and shape of the catchment of interest.

18.9 FREQUENCY ANALYSIS OF LOW FLOWS

Low-flow quantiles are used in water-quality management applications including waste-load allocations and discharge permits, and in siting treatment plants and sanitary landfills. Low-flow statistics are also used in water-supply planning to determine allowable water transfers and withdrawals. Other applications of low-flow frequency analysis include determination of minimum downstream release requirements from hydropower, water-supply, cooling plants, and other facilities.

18.9.1 Selection of Data and Sources

Annual-Event-Based Low-Flow Statistics. Sources of streamflow and low-flow data are discussed in Sec. 18.7.1. The most widely used low-flow index in the United States is the one in 10-year 7-day-average low flow, denoted $Q_{7,0.10}$ (Ref. 117). In general, $Q_{d,p}$ is the annual minimum d-day consecutive average discharge not exceeded with probability p.

Prior to performing low-flow frequency analyses, an effort should be made to "deregulate" low flow series to obtain "natural" streamflows. This includes accounting for the impact of large withdrawals and diversions including water- and wastewater-treatment facilities, as well as urbanization, lake regulation, and other factors. Since low flows result primarily from groundwater inflow to the stream channel, substantial year-to-year carryover in groundwater storage can cause sequences of annual minimum low flows to be correlated from one year to the next (see Fig. 9 in Ref. 116 or Fig. 2 in Ref. 157). Low-flow series should be subjected to trend analysis so that identified trends can be reflected in frequency analyses.

The Flow-Duration Curve. Flow-duration curves are an alternative to analysis of annual minimum d-day averages (see Sec. 8.6.1). A flow-duration curve is the empirical cumulative distribution function of all the daily (or weekly) streamflow recorded at a site. A flow-duration curve describes the fraction of the time over the entire record that different daily flow levels were exceeded. Flow-duration curves are often used in hydrologic studies for run-of-river hydropower, water supply, irrigation planning and design, and water-quality management.[36,40,43,53,121] The flow-duration curves should not be interpreted on an annual event basis, as is $Q_{d,p}$, because the flow-duration curve provides only the fraction of the time that a stream flow level was exceeded; it does not distinguish between regular seasonal variation in flow levels, and random variations from seasonal averages.

18.9.2 Frequency Analysis Methods for Low Flows and Treatment of Zeros

Estimation of $Q_{d,p}$ from stream flow records is generally done by fitting a probability distribution to the annual minimum d-day-average low-flow series. The literature on low-flow frequency analysis is relatively sparse. The extreme value type III or Weibull distribution (see Sec. 18.2.2) is a theoretically plausible distribution for low flows. Studies in Canada and the eastern United States have recommended the three-parameter Weibull, the log-Pearson type 3, and the two-parameter and three-parameter lognormal distributions based on apparent goodness-of-fit.[24,139,154] Fitting methods for complete samples are described in Sec. 18.2.

Low-flow series often contain years with zero values. In some arid areas, zero flows are recorded more often than nonzero flows. Stream flows recorded as zero imply either that the stream was completely dry or that the actual stream flow was below a recording limit. At most U.S. Geological Survey gauges there is a lower stream flow level (0.05 ft^3/s) below which any measurement is reported as a zero. This implies that low-flow series are censored data sets, discussed in Sec. 18.6.3. Zero values should not simply be ignored, nor do they necessarily reflect accurate measurements of the minimum flow in a channel. Based on the hydraulic configuration of a gauge, and knowledge of the rating curve and recording policies, one can generally determine the lowest discharge which can be reliably estimated and would not be recorded as a zero.

The plotting-position method or the conditional probability model in Sec. 18.6.3 are reasonable procedures for fitting a probability distribution with data sets containing recorded zeros. The graphical plotting position approach without a formal statistical model is often sufficient for low-flow frequency analyses. One can define visually the low-flow frequency curve or estimate the parameters of a parametric distribution using probability-plot regression.

18.9.3 Regional Estimates of Low-Flow Statistics

Regional regression procedures are often employed at ungauged sites to estimate low-flow statistics by using basin characteristics. If no reliable regional regression equations are available, one can also consider the drainage area ratio, regional statistics, or base-flow correlation methods described below.

Regional Regression Procedures. Many investigators have developed regional models for the estimation of low-flow statistics at ungauged sites using physiographic basin parameters. This methodology is discussed in Sec. 18.5.2. Unfortunately, most low-flow regression models have large prediction errors, as shown in Fig. 18.5.1, because they are unable to capture important land-surface and subsurface geological characteristics of a basin. In a few regions, efforts to regionalize low-flow statistics have been improved by including basin parameters which in some manner describe the geohydrologic response of each watershed.[9,18,156] Conceptual watershed models can be used to formulate regional regression models of low-flow statistics.[155]

Drainage Area Ratio Method. Perhaps the simplest regional approach for estimation of low-flow statistics is the drainage area ratio method, which would estimate a low-flow quantile y_p for an ungauged site as

$$y_p = \left(\frac{A_y}{A_x}\right) x_p \qquad (18.9.1)$$

where x_p is the corresponding low-flow quantile for a nearby gauging station and A_x and A_y are the drainage areas for the gauging station and ungauged site, respectively. Seepage runs, consisting of a series of discharge measurements along a river reach during periods of base flow, are useful for determining the applicability of this simple linear drainage-area discharge relation.[116] Some studies employ a scaling factor $(A_y/A_x)^b$ to allow for losses by using an exponent $b < 1$ derived by regional regression. (See Sec. 18.5.2.)

Regional Statistics Methods. One can sometimes use a gauging station record to construct a monthly streamflow record at an ungauged site using

$$y(i, j) = M(y_i) + \frac{S(y_i)\,[x(i, j) - M(x_i)]}{S(x_i)} \tag{18.9.2}$$

where $y(i, j)$ and $x(i, j)$ are monthly stream flows at the ungauged and nearby gauged sites, respectively, in month i and year j; $M(x_i)$ and $S(x_i)$ are the mean and standard deviation of the observed flows at the gauged site in month i; and $M(y_i)$ and $S(y_i)$ are the corresponding mean and standard deviation of the monthly flows at the ungauged site obtained from regional regression equations, discussed in Sec. 18.5.2. Hirsch[64] found that this method transferred the characteristics of low flows from the gauged site to the ungauged site.

Base Flow Correlation Procedures. When *base flow measurements* (instantaneous or average daily values) can be obtained at an otherwise ungauged site, they can be correlated with concurrent stream flows at a nearby gauged site for which a long flow record is available.[114,116] Estimators of low-flow moments at the ungauged site can be developed by using bivariate and multivariate regression, as well as estimates of their standard errors.[144] This is an extension to the record augmentation idea in Sec. 18.5.3. Ideally the nearby gauged site is hydrologically similar in terms of topography, geology, and base flow recession characteristics. For a single gauged record, if regression of concurrent daily flows at the two stations yields a model

$$y = a + bx + \epsilon \qquad \text{with Var}(\epsilon) = s_\epsilon^2 \tag{18.9.3}$$

estimators of the mean and variance of annual minimum d-day-average flows y are

$$\begin{aligned} M(y) &= a + b\,M(x) \\ S^2(y) &= b^2\,S^2(x) + s_\epsilon^2 \end{aligned} \tag{18.9.4}$$

where $M(x)$ and $S^2(x)$ are the estimators of the mean and variance of the annual minimum d-day averages at the gauged x site. Base flow correlation procedures are subject to considerable error when only a few discharge measurements are used to estimate the parameters of the model in Eq. (18.9.3), as well as error introduced from use of a model constructed between base flows for use in relating annual minimum d-day averages. (Thus the model R^2 should be at least 70 percent; see also Refs. 56 and 144.)

18.10 FREQUENCY ANALYSIS OF WATER-QUALITY VARIABLES

In the early 1980s, most U.S. water-quality improvement programs aimed at obvious and visible pollution sources resulting from direct *point discharges* of sewage and wastewaters to surface waters. This did not always lead to major improvements in the quality of receiving waters subject to *non-point-source pollution* loadings, corresponding to storm-water runoff and other discharges that carry sediment and pollutants from various distributed sources. The analyses of point- and nonpoint-source water-quality problems differ. Point sources are often sampled regularly, are less variable, and frequently have severe impacts during periods of low stream flow. Nonpoint sources often occur only during runoff-producing storm events, which is when nonpoint discharges generally have their most severe effect on water quality.

18.10.1 Selection of Data and Water-Quality Monitoring

Water-quality problems can be quantified by a number of variables, including biodegradable oxygen demand (BOD) and concentrations of nitrogenous compounds, chlorophyll, metals, organic pesticides, and suspended or dissolved solids. Water-quality monitoring activities include both design and actual data acquisition, and the data's utilization.[102,167] Monitoring programs require careful attention to the definition of the population to be sampled and the sampling plan.[47] A common issue is the detection of trends or changes that have occurred over time because of development or pollution control efforts, as discussed in Chap. 17.

The statistical analysis of water-quality data is complicated by the facts that quality and quantity measurements are not always made simultaneously, the time interval between water-quality samples can be irregular, the precision with which different constituents can be measured varies, and base flow samples (from which background levels may be derived) are often unavailable in studies of non-point-source pollution. The U.S. National Water Data Exchange provides access to data on ambient water quality; see Sec. 18.7.1. A list of other non-point-source water-quality data bases appears in Ref. 76.

18.10.2 Frequency Analysis Methods and Water-Quality Data

It is often inappropriate to use the conventional approach of selecting a single design flow for managing the quality of receiving waters. The most critical impact of pollutant loadings on receiving water quality does not necessarily occur under low flow conditions; often the shock loads associated with intermittent urban storm-water runoff are more critical.[99] Nevertheless, water-quality standards are usually stated in terms of a maximum allowable d-day average concentration.[147] The most common type of design event for the protection of aquatic life is based on the one in T-year d-day average annual low stream flow.[10]

For problems with regular data collection programs yielding continuous or regularly spaced observations, traditional methods of frequency analysis can be employed to estimate exceedance probabilities for annual maxima, or the probability that monthly observations will exceed various values. Event mean concentrations (EMC) corresponding to highway storm-water runoff, combined sewer overflows, urban runoff, sewage treatment plants, and agricultural runoff are often well approximated by lognormal distributions,[39] which have been a common choice for constituent concentrations.[47] Section 18.1.3 discusses advantages and disadvantages of logarithmic transformations. Procedures in Sec. 18.3 for selecting an appropriate probability distribution may be employed.

Investigations of the concentrations associated with trace substances in receiving waters are faced with a recurring problem: a substantial portion of water sample concentrations are below the limits of detection for analytical laboratories. Measurements below the detection limit are often reported as "less than the detection limit" rather than by numerical values, or as zero.[47] Such data sets are called *censored data* in the field of statistics. Probability-plot regression and maximum likelihood techniques for parameter estimation with censored data sets are discussed in Sec. 18.6.3.[60,61]

For intermittent loading problems the situation is more difficult and corresponds roughly to partial duration series, discussed in Sec. 18.6.1. In the context of urban storm-water problems, the average recurrence interval (in years) of a design event has been estimated as

$$T = \frac{1}{NP(C \geq C_0)} \tag{18.10.1}$$

where N is the average number of rainfall runoff events in a year, C is a constituent concentration, and the probability $P(C \geq C_0)$ that an observation C exceeds a standard C_0 in a runoff event is obtained by fitting a frequency distribution to the concentrations measured in observed runoff events.[37] This corresponds to Eq. (18.6.3) of Sec. 18.6.1 with event arrival rate $\lambda = N$. Models such as SWMM (see Chap. 21) can also be used to estimate T directly.

18.11 COMPUTER PROGRAMS FOR FREQUENCY ANALYSIS

Many frequency computations are relatively simple and are easily performed with standard functions on hand calculators, spreadsheets, or general-purpose statistical packages. However, maximum likelihood estimators and several other procedures can be quite involved. Water management agencies in most countries have computer packages to perform the standard procedures they employ. Four sets of routines for flood frequency analyses are discussed below.

U.S. Army Corps of Engineers Flood Flow Frequency Analysis (HECWRC). The U.S. Army Corps of Engineers has developed a library of 60 FORTRAN routines to support statistical analysis on MS-DOS personal computers. The library includes routines for performing the standard Bulletin 17B analyses, as well as general-purpose functions including general statistics, time series, duration curves, plotting positions, and graphical display. Information can be obtained by contacting Flood Frequency Investigations, Department of the Army, COE 51 Support Center, Hydrologic Engineering Center, 609 Second St., Davis, Calif. 95616-46897. HECWRC and other HEC software, with some improvements in the user interface and user support, are actively marketed by several private vendors. The U.S. Geological Survey also provides a program for Bulletin 17B analyses (Chief Hydrologist, U.S. Geological Survey, National Center, Mail Stop 437, Reston, Va. 22092).

British Flood Studies Software. Micro-FSR is a microcomputer-based implementation of the flood-frequency analysis methods developed by the Institute of Hydrology.[105] It also contains menu-driven probable maximum precipitation, unit hydrograph, and reservoir routing calculations for personal computers running MS-DOS. The package and training information can be obtained from Software Sales, Institute of Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire

OX10 38800, United Kingdom; phone, 0491 38800; telex, 849365 HYDROL G; fax, 0491 32256

Consolidated Frequency Analysis (CFA) Package. The package incorporates FORTRAN routines developed by Environment Canada for flood-frequency analyses in that country with MS-DOS computers. Routines allow fitting of three-parameter lognormal, Pearson, log-Pearson, GEV, and Wakeby distributions using moments, maximum likelihood, and sometimes probability-weighted moment analyses. Capabilities are provided for nonparametric trend and tests of independence, as well as employing maximum likelihood procedures for historical information with a single threshold. Contact Dr. Paul Pilon, Inland Waters Directorate, Water Resources Branch, Ottawa, Ontario K1A 0E7, Canada.

FORTRAN Routines for Use with the Method of L Moments. Hosking[74] describes a set of FORTRAN-77 subroutines useful for analyses employing L moments, including subroutines to fit 10 different distributions. Index-flood procedures and regional diagnostic analyses are included. Contact Dr. J. R. M. Hosking, Mathematical Sciences Dept., IBM Research Division, T. J. Watson Research Center, Yorktown Heights, N.Y. 10598. The routines are available through STATLIB, a system for distribution of statistical software by electronic mail. To obtain the software send the message "send lmoments from general" to the e-mail address: statlib@lib.stat.cmu.edu.

REFERENCES

1. Abramowitz, M., and I. A. Stegun, *Handbook of Mathematical Functions,* Appl. Math. Ser. 55, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., 1964.
2. Adamowski, K., "A Monte Carlo Comparison of Parametric and Nonparametric Estimation of Flood Frequencies," *J. Hydrol.,* vol. 108, pp. 295–308, 1989.
3. Ang, A. H-S., and W. H. Tang, *Probability Concepts in Engineering, Planning and Design,* vol. I, *Basic Principles,* Wiley, New York, 1984.
4. Arnell, N., "Expected Annual Damages and Uncertainties in Flood Frequency Estimation," *J. Water Resour. Plann. Manage.,* vol. 115, no. 1, pp. 94–107, 1989.
5. Arora, K., and V. P. Singh, "A Comparative Evaluation of the Estimators of the Log Pearson Type (LP) 3 Distribution," *J. Hydrol.,* vol. 105, no. 1/2, pp. 19–37, 1989.
6. Baker, V. R., R. C. Kochel, and P. C. Patton, eds., *Flood Geomorphology,* Wiley, New York, 1988.
7. Bardsley, W. E., "Using Historical Data in Nonparametric Flood Estimation," *J. Hydrol.,* vol. 108, no. 1–4, pp. 249–255, 1989.
8. Beard, L. R., "Impact of Hydrologic Uncertainties on Flood Insurance," *J. Hydraul. Div., Am. Soc. Civ. Eng.,* (now *J. Hydraul. Eng.*), vol. 104, no. HY11, pp. 1473–1483, 1978.
9. Bingham, R. H., "Regionalization of Low-Flow Characteristics of Tennessee Streams," U.S. Geological Survey, Water Resources Investigations report 85-4191, 1986.
10. Biswas, H., and B. A. Bell, "A Method for Establishing of Site-Specific Stream Design Flows for Wasteload Allocations," *J. Water Pollut. Control Fed.,* vol. 56, no. 10, pp. 1123–1130, 1984.
11. Bobée, B., "The Log Pearson Type 3 Distribution and Its Applications in Hydrology," *Water Resour. Res.,* vol. 14, no. 2, pp. 365–369, 1975.

12. Bobée, B., and R. Robitaille, "The Use of the Pearson Type 3 Distribution and Log Pearson Type 3 Distribution Revisited," *Water Resour. Res.*, vol. 13, no. 2, pp. 427–443, 1977.

13. Bobée, B., and F. Ashkar, *The Gamma Distribution and Derived Distributions Applied in Hydrology*, Water Resources Press, Littleton, Colo., 1991.

14. Boughton, W. C., "A Frequency Distribution for Annual Floods," *Water Resour. Res.*, vol. 16, no. 2, pp. 347–354, 1980.

15. Buishand, T. A., "Bivariate Extreme-Value Data and the Station-Year Method," *J. Hydrol.*, vol. 69, pp. 77–95, 1984.

16. Buishand, T. A., "Bias and Variance of Quantile Estimates from a Partial Duration Series," *J. Hydrol.*, vol. 120, no. 1/4, pp. 35–49, 1990.

17. Burn, D. H., "Evaluation of Regional Flood Frequency Analysis with a Region of Influence Approach," *Water Resour. Res.*, vol. 26, no. 10, pp. 2257–2265, 1990.

18. Cervione, M. A., R. L. Melvin, and K. A. Cyr, "A Method for Estimating the 7-Day 10-Year Low Flow of Streams in Connecticut," Connecticut Water Resources Bulletin No. 34, Department of Environmental Protection, Hartford, Conn., 1982.

19. Chen, C., "Rainfall Intensity-Duration-Frequency Formulas," *J. Hydraul. Eng.*, vol. 109, no. 12, pp. 1603–1621, 1983.

20. Chow, V. T., D. R. Maidment, and L. W. Mays, *Applied Hydrology*, McGraw-Hill, New York, 1988.

21. Chowdhury, J. U., and J. R. Stedinger, "Confidence Intervals for Design Floods with Estimated Skew Coefficient," *J. Hydraul. Eng.*, vol. 117, no. 7, pp. 811–831, 1991.

22. Chowdhury, J. U., J. R. Stedinger, and L-H. Lu, "Goodness-of-Fit Tests for Regional GEV Flood Distributions," *Water Resour. Res.*, vol. 27, no. 7, pp. 1765–1776, 1991.

23. Cohn, T. A., and J. R. Stedinger, "Use of Historical Information in a Maximum-Likelihood Framework," *J. Hydrol.*, vol. 96, no. 1-4, pp. 215–223, 1987.

24. Condie, R., and G. A. Nix, "Modeling of Low Flow Frequency Distributions and Parameter Estimation," Int. Water Resour. Symp. on Water for Arid Lands, Teheran, Iran, Dec. 8–9, 1975.

25. Costa, J. E., "A History of Paleoflood Hydrology in the United States, 1800–1970," *EOS*, vol. 67, pp. 425–430, 1986.

26. Cunnane, C., "Unbiased Plotting Positions—A Review," *J. Hydrol.*, vol. 37, no. 3/4, pp. 205–222, 1978.

27. Cunnane, C., "A Note on the Poisson Assumption in Partial Duration Series Models," *Water Resour. Res.*, vol. 15, no. 2, pp. 489–494, 1979.

28. Cunnane, C., "Methods and Merits of Regional Flood Frequency Analysis," *J. Hydrol.*, vol. 100, pp. 269–290, 1988.

29. Cunnane, C., "Statistical Distributions for Flood Frequency Analysis, Operational Hydrology Report No. 33, World Meteorological Organization WMO—No. 718, Geneva, Switzerland, 1989.

30. D'Agostino, R. B., and M. A. Stephens, *Goodness-of-Fit Procedures*, Marcel Dekker, New York, 1986.

31. Dalrymple, T., "Flood Frequency Analysis," U.S. Geological Survey, *Water Supply Paper* 1543-A, 1960.

32. Damázio, J. M., Comment on "Quantile Estimation with More or Less Floodlike Distributions," by J. M. Landwehr, N. C. Matalas, and J. R. Wallis, *Water Resour. Res.*, vol. 20, no. 6, pp. 746–750, 1984.

33. David, H. A., *Order Statistics*, 2d ed., Wiley, New York, 1981.

34. Davison, A. C., "Modeling Excesses over High Thresholds, with an Application," in *Statistical Extremes and Applications*, J. Tiago de Oliveira, ed., D. Reidel, Dordrecht, Holland, pp. 461–482, 1984.

35. Derenzo, S. E., "Approximations for Hand Calculators Using Small Integer Coefficients,' *Math. Computation,* vol. 31, no. 137, pp. 214–225, 1977.

36. Dingman, S. L., "Synthesis of Flow-Duration Curves for Unregulated Streams in New Hampshire," *Water Resour. Bull.,* vol. 14, no. 6, pp. 1481–1502, 1978.

37. DiToro, D. M., and E. D. Driscoll, "A Probabilistic Methodology for Analyzing Water Quality Effects of Urban Runoff on Rivers and Streams," U.S. Environmental Protection Agency, Washington, D.C., 1984.

38. Downton, F., "Linear Estimates of Parameters of the Extreme Value Distribution,' *Technometrics,* vol. 8, pp. 3–17, 1966.

39. Driscoll, E. D., "Lognormality of Point and Non-Point Source Pollutant Concentrations," *Urban Runoff Quality,* Proceedings of an Engineering Foundation Conference New England College, Henniker, N.H., pp. 438–458, 1986.

40. Fennessey, N., and R. M. Vogel, "Regional Flow-Duration Curves for Ungauged Sites in Massachusetts," *J. Water Resour. Plann. Manage.,* vol. 116, no. 4, pp. 530–549, 1990

41. Filliben, J. J., "The Probability Plot Correlation Test for Normality," *Technometrics,* vol 17, no. 1, pp. 111–117, 1975.

42. Fontaine, T. A., and K. W. Potter, "Estimating Probabilities of Extreme Rainfalls," *J Hydraulic Eng. ASCE,* vol. 115, no. 11, pp. 1562–1575, 1989.

43. Foster, H. A., "Duration Curves," *Trans. Am. Soc. Civ. Eng.,* vol. 99, pp. 1213–1267 1934.

44. Foufoula-Georgiou, E., "A Probabilistic Storm Transposition Approach for Estimating Exceedance Probabilities of Extreme Precipitation Depths," *Water Resour. Res.,* vol. 25 no. 5, pp. 799–815, 1989.

45. Foufoula-Georgiou, E., and L. L. Wilson," In Search of Regularities in Extreme Rainstorms," *J. Geophys. Res.,* vol. 95, no. D3, pp. 2061–2072, 1990.

46. Frederick, R. H., V. A. Meyers, and E. P. Auciello, "Five and Sixty Minute Precipitation Frequency for Eastern and Central United States," NOAA tech. memorandum Hydro-35 Silver Springs, Md., 1977.

47. Gilbert, R. O., *Statistical Methods for Environmental Pollution Monitoring,* Van Nostrand Reinhold, New York, 1987.

48. Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis, "Probability Weighted Moments: Definitions and Relation to Parameters of Several Distributions Expressible in Inverse Form," *Water Resour. Res.,* vol. 15, no. 6, pp. 1049–1054, 1979.

49. Gringorten, I. I., "A Plotting Rule for Extreme Probability Paper," *J. Geophys. Res.,* vol 68, no. 3, pp. 813–814, 1963.

50. Grygier, J. C., J. R. Stedinger, and H. B. Yin, "A Generalized Move Procedure for Extending Correlated Series," *Water Resour. Res.,* vol. 25, no. 3, pp. 345–350, 1989.

51. Gumbel, E. J., *Statistics of Extremes,* Columbia University Press, New York, 1958.

52. Guo, S. L., "A Discussion on Unbiased Plotting Positions for the General Extreme Value Distribution," *J. Hydrol.,* vol. 121, no. 1-4, pp. 33–44, 1990.

53. Gupta, R. S., *Hydrology and Hydraulic Systems,* Prentice-Hall, Englewood Cliffs, N.J., pp. 363–367, 1989.

54. Haan, C. T., *Statistical Methods in Hydrology,* Iowa State University Press, Ames, Iowa, 1977.

55. Haas, C. N., and P. A. Scheff, "Estimation of Averages in Truncated Samples," *Environ. Sci. Technol.,* vol. 24, no. 6, pp. 912–919, 1990.

56. Hardison, C. H., and M. E. Moss, "Accuracy of Low-Flow Characteristics Estimated by Correlation of Base-flow Measurements," U.S. Geological Survey, *Water Supply Paper* 1542-B, pp. 350–355, 1972.

57. Harlow, D. G., "The Effect of Proof-Testing on the Weibull Distribution," *J. Material Science,* vol. 24, pp. 1467–1473, 1989.

58. Harter, H. L., "Another Look at Plotting Positions," *Commun. Statistical-Theory and Methods*, vol. 13, no. 13, pp. 1613–1633, 1984.

59. Hazen, A., "Discussion on 'Flood flows' by W. E. Fuller," *Trans. Amer. Soc. Civ. Eng.*, vol. 77, pp. 526–632, 1914.

60. Helsel, D. R., and T. A. Cohn, "Estimation of Descriptive Statistics for Multiply Censored Water Quality Data," *Water Resour. Res.*, vol. 24, no. 12, pp. 1997–2004, 1988.

61. Helsel, D. R., "Less Than Obvious: Statistical Treatment of Data Below the Detection Limit," *Environ. Sci. Technol.*, vol. 24, no. 12, pp. 1767–1774, 1990.

62. Hershfield, D. M., and W. T. Wilson, "A Comparison of Extreme Rainfall Depths from Tropical and Nontropical Storms," *J. Geophys. Res.*, vol. 65, pp. 959–982, 1960.

63. Hershfield, D. M., "Rainfall Frequency Atlas of the United States for Durations from 30 Minutes to 24 Hours and Return Periods from 1 to 100 Years," tech. paper 40, U.S. Weather Bureau, Washington, D.C., 1961.

64. Hirsch, R. M., "An Evaluation of Some Record Reconstruction Techniques," *Water Resour. Res.*, vol. 15, no. 6, pp. 1781–1790, 1979.

65. Hirsch, R. M., "A Comparison of Four Record Extension Techniques," *Water Resour. Res.*, vol. 18, no. 4, pp. 1081–1088, 1982.

66. Hirsch, R. M., and J. R. Stedinger, "Plotting Positions for Historical Floods and Their Precision," *Water Resour. Res.*, vol. 23, no. 4, pp. 715–727, 1987.

67. Hoshi, K., J. R. Stedinger, and S. Burges, "Estimation of Log Normal Quantiles: Monte Carlo Results and First-Order Approximations," *J. Hydrol.*, vol. 71, no. 1/2, pp. 1–30, 1984.

68. Hosking, J. R. M., J. R. Wallis, and E. F. Wood, "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability Weighted Moments," *Technometrics*, vol. 27, no. 3, pp. 251–261, 1985.

69. Hosking, J. R. M., J. R. Wallis, and E. F. Wood, "An Appraisal of the Regional Flood Frequency Procedure in the UK Flood Studies Report," *Hydrol. Sci. Jour.*, vol. 30, no. 1, pp. 85–109, 1985.

70. Hosking, J. R. M., and J. R. Wallis, "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, vol. 29, no. 3, pp. 339–349, 1987.

71. Hosking, J. R. M., and J. R. Wallis, "The Effect of Intersite Dependence on Regional Flood Frequency Analysis," *Water Resour. Res.*, vol. 24, no. 4, pp. 588–600, 1988.

72. Hosking, J. R. M., "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," *J. Royal Statistical Soc., B*, vol. 52, no. 2, pp. 105–124, 1990.

73. Hosking, J. R. M., "Approximations for Use in Constructing L-moment Ratio Diagrams," research report RC-16635, IBM Research Division, T. J. Watson Research Center, Yorktown Heights, N.Y., March 12, 1991.

74. Hosking, J. R. M., "Fortran Routines for Use with the Method of L-Moments," research report RC-17097, IBM Research Division, T. J. Watson Research Center, Yorktown Heights, N.Y., August 21, 1991.

75. Houghton, J. C., "Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows," *Water Resour. Res.*, vol. 14, no. 6, pp. 1105–1109, 1978.

76. Huber, W. C., and J. P. Heaney, "Urban Rainfall-Runoff-Quality Data Base," EPA-600/8-77-009 (NTIS PB-270 065), Environmental Protection Agency, Cincinnati, 1977.

77. Hupp, C. R., "Botanical Evidence of Floods and Paleoflood History," *Regional Flood Frequency Analysis*, V. P. Singh, ed., D. Reidel, Dordrecht, pp. 355–369, 1987.

78. Hupp, C. R., "Plant Ecological Aspects of Flood Geomorphology and Paleoflood History," Chap. 20 in *Flood Geomorphology*, V. R. Baker, R. C. Kochel and P. C. Patton, Wiley, New York, 1988.

79. Interagency Advisory Committee on Water Data, *Guidelines for Determining Flood Flow*

Frequency, Bulletin 17B, U.S. Department of the Interior, U.S. Geological Survey, Office of Water Data Coordination, Reston, Va., 1982.

80. Jenkinson, A. F., "Statistics of Extremes," in "Estimation of Maximum Floods," WMO No. 233, TP 126 (tech. note no. 98), pp. 183–228, 1969.
81. Jin, M., and J. R. Stedinger, "Flood Frequency Analysis with Regional and Historical Information," *Water Resour. Res.,* vol. 25, no. 5, pp. 925–936, 1989.
82. Kelman, J., "Cheias E Aproveitamentos Hidrelétricos," *Revista Brasileria de Engenharia,* Rio de Janeiro, 1987.
83. Kirby, W., "Computer Oriented Wilson-Hilferty Transformation That Preserves the First Three Moments and the Lower Bound of the Pearson Type 3 Distribution," *Water Resour. Res.,* vol. 8, no. 5, pp. 1251–1254, 1972.
84. Kirby, W., "Algebraic Boundness of Sample Statistics," *Water Resour. Res.,* vol. 10, no. 2, pp. 220–222, 1974.
85. Kite, G. W., *Frequency and Risk Analysis in Hydrology,* Water Resources Publication, Littleton, Colo., 1988.
86. Kochel, R. C., and V. R. Baker, "Paleoflood Analysis Using Slackwater Deposits," Chap. 21 in *Flood Geomorphology,* V. R. Baker, R. C. Kochel and P. C. Patton, eds., Wiley, New York, 1988.
87. Kuczera, G., "Robust Flood Frequency Models," *Water Resour. Res.,* vol. 18, no. 2, pp. 315–324, 1982.
88. Kuczera, G., "Effects of Sampling Uncertainty and Spatial Correlation on an Empirical Bayes Procedure for Combining Site and Regional Information," *J. Hydrol.,* vol. 65, no. 4, pp. 373–398, 1983.
89. Landwehr, J. M., N. C. Matalas, and J. R. Wallis, "Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles," *Water Resour. Res.,* vol. 15, no. 5, pp. 1055–1064, 1979.
90. Landwehr, J. M., N. C. Matalas, and J. R. Wallis, "Quantile Estimation with More or Less Floodlike Distributions," *Water Resour. Res.,* vol. 16, no. 3, pp. 547–555, 1980.
91. Landwehr, J. M., G. D. Tasker, and R. D. Jarrett, "Discussion of Relative Accuracy of Log Pearson III Procedures, by J. R. Wallis and E. F. Wood," *J. Hydraul. Eng.,* vol. 111, no. 7, pp. 1206–1210, 1987.
92. Lane, W. L., "Paleohydrologic Data and Flood Frequency Estimation," in *Regional Flood Frequency Analysis,* V. P. Singh, ed., D. Reidel, Dordrecht, pp. 287–298, 1987.
93. Langbein, W. B., "Annual Floods and the Partial Duration Flood Series," EOS, *Trans. AGU,* vol. 30, no. 6, pp. 879–881, 1949.
94. Lettenmaier, D. P., J. R. Wallis, and E. F. Wood, "Effect of Regional Heterogeneity on Flood Frequency Estimation," *Water Resour. Res.,* vol. 23, no. 2, pp. 313–324, 1987.
95. Loucks, D. P., J. R. Stedinger, and D. A. Haith, *Water Resource Systems Planning and Analysis,* Prentice-Hall, Englewood Cliffs, N.J., 1981.
96. Lu, L., and J. R. Stedinger, "Variance of 2- and 3-Parameter GEV/PWM Quantile Estimators: Formulas, Confidence Intervals and a Comparison," *J. Hydrol.,* vol. 138, no. 1/2, pp. 247–268, 1992.
97. Matalas, N. C., and B. Jacobs, "A Correlation Procedure for Augmenting Hydrologic Data," *U.S. Geological Survey,* prof. paper 434-E, E1–E7, 1964.
98. McCuen, R. H., "Map Skew???," *J. Water Resour. Plann. Manage. Div., Am. Soc. Civ. Eng.* (now) *J. Water Res. Plann. Manage.,* vol. 105, no. WR2, pp. 265–277 [with Closure, vol. 107, no. WR2, p. 582, 1981], 1979.
99. Medina, M. A., "Level III: Receiving Water Quality Modeling for Urban Stormwater Management," EPA-600/2-79-100, Environmental Protection Agency, Cincinnati, 1979.
100. Miller, J. F., R. H. Frederick, and R. J. Tracey, "Precipitation-Frequency Atlas of the Conterminous Western United States (by states)," *NOAA Atlas 2,* National Weather Service, Silver Spring, Md., 1973.

101. Minitab, *Minitab Reference Manual,* Release 5, Minitab, Inc., State College, Pa., December 1986.

102. Montgomery, R. H., and T. G. Sanders, "Uncertainty in Water Quality Data," in *Statistical Aspects of Water Quality Monitoring,* A. H. El-Sharwi and R. E. Kwiatkowski, eds., Elsevier, New York, pp. 17–29, 1985.

103. Myers, V. A., and R. M. Zehr, "A Methodology for Point-to-Area Rainfall Frequency Ratios," *NOAA,* National Weather Service tech. rep. 24, 1980.

104. National Research Council, Committee on Techniques for Estimating Probabilities of Extreme Floods, *Estimating Probabilities of Extreme Floods, Methods and Recommended Research,* National Academy Press, Washington, D.C., 1988.

105. Natural Environmental Research Council, *Flood Studies Report,* vol. I, Hydrological Studies, London, 1975.

106. Newton, D. W., "Realistic Assessment of Maximum Flood Potentials," *J. Hydraul. Eng.,* ASCE, vol. 109, no. 6, pp. 905–918, 1983.

107. Newton, D. W., and J. C. Herrin, "Assessment of commonly used flood frequency methods," Transportation Research Record Series TRR896, Washington, D.C., 1982.

108. North, M., "Time-dependent Stochastic Model of Floods," *J. Hydraul. Div., Am. Soc. Civ. Eng.* (now *J. Hydraul. Eng.*), vol. 106, no. HY5, pp. 649–665, 1980.

109. Phien, H. N., "A Method of Parameter Estimator for the Extreme Value type I Distribution," *J. Hydrol.,* vol. 90, pp. 251–268, 1987.

110. Pilgrim, D. H., ed., *Australian Rainfall and Runoff, A Guide to Flood Estimation,* vol. I, The Institute of Engineers, Barton ACT, Australia, (rev. ed.), 1987.

111. Potter, K. W., and E. B. Faulkner, "Catchment Response Times as a Predictor of Flood Quantiles," *Water Resour. Bull.,* vol. 23, no. 5, pp. 857–861, 1987.

112. Potter, K. W., and D. P. Lettenmaier, "A Comparison of Regional Flood Frequency Estimation Methods Using a Resampling Method," *Water Resour. Res.,* vol. 26, no. 3, pp. 415–424, 1990.

113. Rasmussen, P. F., and D. Rosbjerg, "Risk Estimation in Partial Duration Series," *Water Resour. Res.,* vol. 25, no. 11, pp. 2319–2330, 1989.

114. Riggs, H. C., "Estimating Probability Distributions of Drought Flows," *Water and Sewage Works,* vol. 112, no. 5, pp. 153–157, 1965.

115. Riggs, H. C., Chap. 3, "Frequency Curves," U.S. Geological Survey Surface Water Techniques Series, Book 2, U.S. Geological Survey, Washington, D.C., 1966.

116. Riggs, H. C., "Low Flow Investigations," *U.S. Geological Survey Techniques of Water Resources Investigations,* Book 4, U.S. Geological Survey, Washington, D.C., 1972.

117. Riggs, H. C., et al., "Characteristics of Low Flows, Report of an ASCE Task Committee," *J. Hydraul. Eng.,* vol. 106, no. 5, pp. 717–731, 1980.

118. Rosbjerg, D., "Estimation in Partial Duration Series with Independent and Dependent Peak Values," *J. Hydrol.,* vol. 76, no. 1, pp. 183–196, 1985.

119. Rossi, F., M. Fiorentino, and P. Versace, "Two Component Extreme Value Distribution for Flood Frequency Analysis," *Water Resour. Res.,* vol. 20, no. 7, pp. 847–856, 1984.

120. Schaefer, M. G., "Regional Analyses of Precipitation Annual Maxima in Washington State," *Water Resour. Res.,* vol. 26, no. 1, pp. 119–131, 1990.

121. Searcy, J. K., "Flow-Duration Curves," U.S. Geological Survey, *Water Supply Paper* 1542-A, 1959.

122. Selker, J. S., and D. A. Haith, "Development and Testing of Single-Parameter Precipitation Distributions," *Water Resour. Res.,* vol. 26, no. 11, pp. 2733–2740, 1990.

123. Shipe, A. P., and J. T. Riedel, "Greatest Known Areal Storm Rainfall Depths for the Contiguous United States," NOAA tech. memo. Hydro-33, National Weather Service, 1976.

124. Slack, J. R., J. R. Wallis, and N. C. Matalas, "On the Value of Information in Flood Frequency Analysis," *Water Resour. Res.,* vol. 11, no. 5, pp. 629–648, 1975.

125. Smith, R. L., "Threshold Methods for Sample Extremes," in *Statistical Extremes and Applications,* J. Tiago de Oliveira, ed., D. Reidel, Dordrecht, Holland, pp. 621–638, 1984.

126. Stedinger, J. R., "Fitting Log Normal Distributions to Hydrologic Data," *Water Resour. Res.,* vol. 16, no. 3, pp. 481–490, 1980.

127. Stedinger, J. R., "Design Events with Specified Flood Risk," *Water Resour. Res.,* vol. 19, no. 2, pp. 511–522, 1983.

128. Stedinger, J. R., "Estimating a Regional Flood Frequency Distribution," *Water Resour. Res.,* vol. 19, no. 2, pp. 503–510, 1983.

129. Stedinger, J. R., "Confidence Intervals for Design Events," *J. Hydraul. Div., Am. Soc. Civ. Eng.* (now *J. Hydraul. Eng.*) vol. 109, no. HY1, pp. 13–27, 1983.

130. Stedinger, J. R., and G. D. Tasker, "Regional Hydrologic Regression, 1. Ordinary, Weighted and Generalized Least Squares Compared," *Water Resour. Res.,* vol. 21, no. 9, pp. 1421–1432, 1985.

131. Stedinger, J. R., and G. D. Tasker, "Correction to 'Regional Hydrologic Analysis, 1. Ordinary, Weighted and Generalized Least Squares Compared,'" *Water Resour. Res.,* vol. 22, no. 5, p. 844, 1986.

132. Stedinger, J. R., and G. D. Tasker, "Regional Hydrologic Analysis, 2. Model Error Estimates, Estimation of Sigma, and Log-Pearson Type 3 Distributions," *Water Resour. Res.,* vol. 22, no. 10, pp. 1487–1499, 1986.

133. Stedinger, J. R., and T. A. Cohn, "Flood Frequency Analysis with Historical and Paleoflood Information," *Water Resour. Res.,* vol. 22, no. 5, pp. 785–793, 1986.

134. Stedinger, J. R., and V. R. Baker, "Surface Water Hydrology: Historical and Paleoflood Information," *Reviews of Geophysics,* vol. 25, no. 2, pp. 119–124, 1987.

135. Stedinger, J., R. Surani, and R. Therivel, "Max Users Guide: A Program for Flood Frequency Analysis Using Systematic-Record, Historical, Botanical, Physical Paleohydrologic and Regional Hydrologic Information Using Maximum Likelihood Techniques," Department of Environmental Engineering, Cornell University, Ithaca, N.Y., 1988.

136. Stephens, M., "E.D.F. Statistics for Goodness of Fit," *J. Amer. Statistical Assoc.,* vol. 69, pp. 730–737, 1974.

137. Tasker, G. D., "Effective Record Length for the T-Year Event," *J. Hydrol.,* vol. 64, pp. 39–47, 1983.

138. Tasker, G. D., and J. R. Stedinger, "Estimating Generalized Skew with Weighted Least Squares Regression," *J. Water Resour. Plann. Manage.,* vol. 112, no. 2, pp. 225–237, 1986.

139. Tasker, G. D., "A Comparison of Methods for Estimating Low Flow Characteristics of Streams," *Water Resour. Bull.,* vol. 23, no. 6, pp. 1077–1083, 1987.

140. Tasker, G. D., and N. E. Driver, "Nationwide Regression Models for Predicting Urban Runoff Water Quality at Unmonitored Sites," *Water Resour. Bull,* vol. 24, no. 5, pp. 1091–1101, 1988.

141. Tasker, G. D., and J. R. Stedinger, "An Operational GLS Model for Hydrologic Regression," *J. Hydrol.,* vol. 111, pp. 361–375, 1989.

142. Thomas, D. M., and M. A. Benson, "Generalization of Streamflow Characteristics from Drainage-Basin Characteristics," U.S. Geological Survey, *Water Supply Paper* 1975, 1970.

143. Thomas, W. O., "A Uniform Technique for Flood Frequency Analysis," *J. Water Resour. Plann. Management,* vol. 111, no. 3, pp. 321–337, 1985.

144. Thomas, W. O., and J. R. Stedinger, "Estimating Low-Flow Characteristics at Gaging Stations and Through the Use of Base-Flow Measurements," pp. 197–205, in *Proceedings of the US-PRC Bilateral Symposium on Droughts and Arid-Region Hydrology,* W. H. Kirby and W. Y. Tan, eds., open-file report 81-244, U.S. Geological Survey, Reston, Va., 1991.

145. Todorovic, P., "Stochastic Models of Floods," *Water Resour. Res.*, vol. 14, no. 2, pp. 345–356, 1978.

146. U.S. Army Corps of Engineers, "Storm Rainfall on the United States," Office of the Chief Engineer, Washington, D.C., ongoing publication since 1945.

147. U.S. Environmental Protection Agency, "Quality Criteria for Water—1986," EPA 440/5-86-001, Office of Water Regulations and Standards, Washington, D.C., 1986.

148. U.S. Weather Bureau, "Rainfall Intensity-Duration Frequency Curves (for Selected Stations in the United States, Alaska, Hawaiian Islands, and Puerto Rico)," tech. paper no. 25, Washington, D.C., 1955.

149. U.S. Weather Bureau, "Two-to-Ten-Day Precipitation for Return Period of 2 to 100 Years in the Contiguous United States," tech. paper no. 49, Washington, D.C., 1964.

150. Vicens, G. J., I. Rodríguez-Iturbe, and J. C. Schaake, Jr., "A Bayesian Framework for the Use of Regional Information in Hydrology," *Water Resour. Res.*, vol. 11, no. 3, pp. 405–414, 1975.

151. Vogel, R. M., and J. R. Stedinger, "Minimum Variance Streamflow Record Augmentation Procedures," *Water Resour. Res.*, vol. 21, no. 5, pp. 715–723, 1985.

152. Vogel, R. M., "The Probability Plot Correlation Coefficient Test for Normal, Lognormal, and Gumbel Distributional Hypotheses," *Water Resour. Res.*, vol. 22, no. 4, pp. 587–590, 1986.

153. Vogel, R. M., "Correction to 'The Probability Plot Correlation Coefficient Test for Normal, Lognormal, and Gumbel Distributional Hypotheses," *Water Resour. Res.*, vol. 23, no. 10, p. 2013, 1987.

154. Vogel, R. M., and C. N. Kroll, "Low-Flow Frequency Analysis Using Probability-Plot Correlation Coefficients," *J. Water Resour. Plann. Manage., ASCE,* vol. 115, no. 3, pp. 338–357, 1989.

155. Vogel, R. M., C. N. Kroll, and K. M. Driscoll, "Regional Geohydrologic-Geomorphic Relationships for the Estimation of Low-Flows," *Proceedings of the International Conference on Channel Flow and Catchment Runoff,* B. C. Yen, ed., University of Virginia, Charlottesville, pp. 267–277, 1989.

156. Vogel, R. M., and C. N. Kroll, "Generalized Low-Flow Frequency Relationships for Ungauged Sites in Massachusetts," *Water Resour. Bull,* vol. 26, no. 2, pp. 241–253, 1990.

157. Vogel, R. M., and C. N. Kroll, "The Value of Streamflow Augmentation Procedures in Low Flow and Flood-Flow Frequency Analysis," *J. Hydrol.*, vol. 125, pp. 259–276, 1991.

158. Vogel, R. M., and D. E. McMartin, "Probability Plot Goodness-of-Fit and Skewness Estimation Procedures for the Pearson Type III Distribution," *Water Resour. Res.*, vol. 27, no. 12, pp. 3149–3158, 1991.

159. Wallis, J. R., N. C. Matalas, and J. R. Slack, "Just a Moment!," *Water Resour. Res.*, vol. 10, no. 2, pp. 211–221, 1974.

160. Wallis, J. R., "Risk and Uncertainties in the Evaluation of Flood Events for the Design of Hydraulic Structures," in *Piene e Siccita,* E. Guggino, G. Rossi, and E. Todini, eds., Fondazione Politecnica del Mediterraneo, Catania, Italy, pp. 3–36, 1980.

161. Wallis, J. R., "Probable and Improbable Rainfall in California," res. rep. RD9350, IBM, Armonk, N.Y., 1982.

162. Wallis, J. R., and E. F. Wood, "Relative Accuracy of Log Pearson III Procedures," *J. Hydraul. Eng.*, vol. 111, no. 7, pp. 1043–1056 [with discussion and closure, *J. Hydraul. Eng.*, vol. 113, no. 7, pp. 1205–1214, 1987], 1985.

163. Wallis, J. R., "Catastrophes, Computing and Containment: Living in Our Restless Habitat," *Speculation in Science and Technology,* vol. 11, no. 4, pp. 295–315, 1988.

164. Wang, Q. J., "Estimation of the GEV Distribution from Censored Samples by Method of Partial Probability Weighted Moments," *J. Hydrol.*, vol. 120, no. 1-4, pp. 103–114, 1990*a*.

165. Wang, Q. J., "Unbiased Estimation of Probability Weighted Moments and Partial Proba-

bility Weighted Moments for Systematic and Historical Flood Information and Their Application to Estimating the GEV Distribution," *J. Hydrol.*, vol. 120, no. 1-4, pp. 115–124, 1990.

166. Wang, Q. J., "Unbiased Plotting Positions for Historical Flood Information," *J. Hydrol.*, vol. 124, no. 3-4, pp. 197–206, 1991.

167. Ward, R. C., J. C. Loftis, and G. B. McBride, *Design of Water Quality Monitoring Systems,* Van Nostrand Reinhold, New York, 1991.

168. Waylen, P., and M-K. Woo, "Prediction of Annual Floods Generated by Mixed Processes," *Water Resour. Res.*, vol. 18, no. 4, pp. 1283–1286, 1982.

169. WB-ESSA, Weather Bureau Observing Handbook No. 2, Substation Observations, Office of Meteorological Operations, Weather Bureau, Environmental Sciences Administration, Silver Spring, Md., 1970.

170. Wenzel, H. G., "Rainfall for Urban Stormwater Design," in *Urban Storm Water Hydrology,* D. F. Kibler, ed., Water Resources Monograph 7, AGU, Washington, D.C., 1982.

171. Wilson, L. L., and E. Foufoula-Georgiou, "Regional Rainfall Frequency Analysis via Stochastic Storm Transposition," *J. Hydraul. Eng.*, vol. 116, no. 7, pp. 859–880, 1990.

172. World Meteorological Organization, *Guide to Hydrological Practices,* vol. II, *Analysis, Forecasting, and Other Applications,* WMO no. 168 (4th ed.), Geneva, 1983.

173. Woolhiser, D. A., and J. Roldan, "Stochastic Daily Precipitation Models, 2. A Comparison of Distributions of Amounts," *Water Resour. Res.*, vol. 18, no. 5, pp. 1461–1468, 1982.