CHAPTER 12

FREQUENCY ANALYSIS

Hydrologic systems are sometimes impacted by extreme events, such as severe storms, floods, and droughts. The magnitude of an extreme event is inversely related to its frequency of occurrence, very severe events occurring less frequently than more moderate events. The objective of frequency analysis of hydrologic data is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions. The hydrologic data analyzed are assumed to be independent and identically distributed, and the hydrologic system producing them (e.g., a storm rainfall system) is considered to be stochastic, space-independent, and time-independent in the classification scheme shown in Fig. 1.4.1. The hydrologic data employed should be carefully selected so that the assumptions of independence and identical distribution are satisfied. In practice, this is often achieved by selecting the annual maximum of the variable being analyzed (e.g., the annual maximum discharge, which is the largest instantaneous peak flow occurring at any time during the year) with the expectation that successive observations of this variable from year to year will be independent.

The results of flood flow frequency analysis can be used for many engineering purposes: for the design of dams, bridges, culverts, and flood control structures; to determine the economic value of flood control projects; and to delineate flood plains and determine the effect of encroachments on the flood plain.

12.1 RETURN PERIOD

Suppose that an extreme event is defined to have occurred if a random variable X is greater than or equal to some level x_T. The *recurrence interval* τ is the time

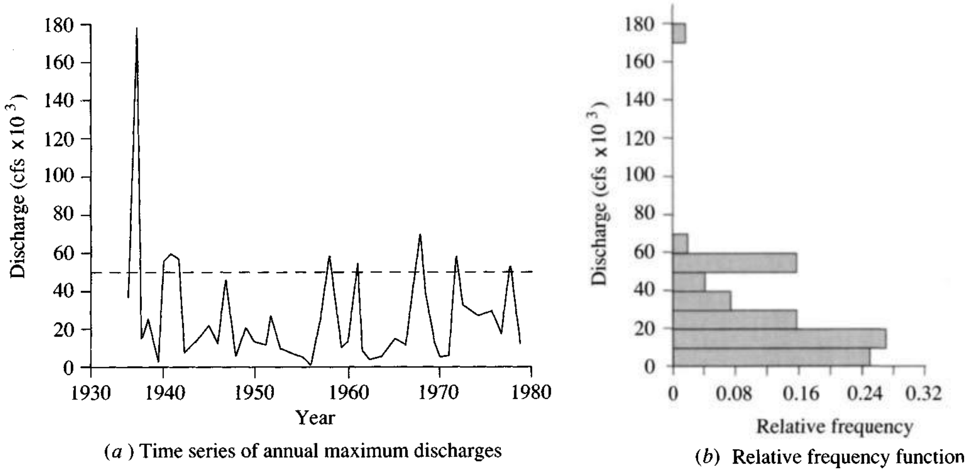

FIGURE 12.1.1
Annual maximum discharge of the Guadalupe River near Victoria, Texas.

between occurrences of $X \geq x_T$. For example, Fig. 12.1.1 shows the record of annual maximum discharges of the Guadalupe River near Victoria, Texas, from 1935 to 1978, plotted from the data given in Table 12.1.1. If $x_T = 50,000$ cfs, it can be seen that the maximum discharge exceeded this level nine times during the period of record, with recurrence intervals ranging from 1 year to 16 years, as shown in Table 12.1.2.

The *return period T* of the event $X \geq x_T$ is the expected value of τ, $E(\tau)$, its average value measured over a very large number of occurrences. For the Guadalupe River data, there are 8 recurrence intervals covering a total period of 41 years between the first and last exceedences of 50,000 cfs, so the return period of a 50,000 cfs annual maximum discharge on the Guadalupe River is

TABLE 12.1.1
Annual maximum discharges of the Guadalupe River near Victoria, Texas, 1935–1978, in cfs

Year	1930	1940	1950	1960	1970
0		55,900	13,300	23,700	9,190
1		58,000	12,300	55,800	9,740
2		56,000	28,400	10,800	58,500
3		7,710	11,600	4,100	33,100
4		12,300	8,560	5,720	25,200
5	38,500	22,000	4,950	15,000	30,200
6	179,000	17,900	1,730	9,790	14,100
7	17,200	46,000	25,300	70,000	54,500
8	25,400	6,970	58,300	44,300	12,700
9	4,940	20,600	10,100	15,200	

TABLE 12.1.2
Years with annual maximum discharge equaling or exceeding 50,000 cfs on the Guadalupe River near Victoria, Texas, and corresponding recurrence intervals

Exceedence year	1936	1940	1941	1942	1958	1961	1967	1972	1977	Average
Recurrence interval (years)		4	1	1	16	3	6	5	5	5.1

approximately $\bar{\tau} = 41/8 = 5.1$ years. Thus the return period of an event of a given magnitude may be defined as the *average recurrence interval* between events *equalling or exceeding* a specified magnitude.

The probability $p = P(X \geq x_T)$ of occurrence of the event $X \geq x_T$ in any observation may be related to the return period in the following way. For each observation, there are two possible outcomes: either "success" $X \geq x_T$ (probability p) or "failure" $X < x_T$ (probability $1 - p$). Since the observations are independent, the probability of a recurrence interval of duration τ is the product of the probabilities of $\tau - 1$ failures followed by one success, that is, $(1 - p)^{\tau-1}p$, and the expected value of τ is given by

$$
\begin{aligned}
E(\tau) &= \sum_{\tau=1}^{\infty} \tau(1 - p)^{\tau-1}p \\
&= p + 2(1 - p)p + 3(1 - p)^2p + 4(1 - p)^3p + \ldots \\
&= p[1 + 2(1 - p) + 3(1 - p)^2 + 4(1 - p)^3 + \ldots]
\end{aligned}
\tag{12.1.1a}
$$

The expression within the brackets has the form of the power series expansion $(1 + x)^n = 1 + nx + [n(n-1)/2]x^2 + [n(n-1)(n-2)/6]x^3 + \ldots$, with $x = -(1-p)$ and $n = -2$, so (12.1.1*a*) may be rewritten

$$
\begin{aligned}
E(\tau) &= \frac{p}{[1 - (1 - p)]^2} \\
&= \frac{1}{p}
\end{aligned}
\tag{12.1.1b}
$$

Hence $E(\tau) = T = 1/p$; that is, the probability of occurrence of an event in any observation is the inverse of its return period:

$$
P(X \geq x_T) = \frac{1}{T} \tag{12.1.2}
$$

For example, the probability that the maximum discharge in the Guadalupe River will equal or exceed 50,000 cfs in any year is approximately $p = 1/\bar{\tau} = 1/5.1 = 0.195$.

What is the probability that a T-year return period event will occur at least once in N years? To calculate this, first consider the situation where no T-year event occurs in N years. This would require a sequence of N successive "failures," so that

$$P(X < x_T \text{ each year for } N \text{ years}) = (1 - p)^N$$

The complement of this situation is the case required, so by (11.1.3)

$$P(X \geq x_T \text{ at least once in } N \text{ years}) = 1 - (1 - p)^N \qquad (12.1.3)$$

Since $p = 1/T$,

$$P(X \geq x_T \text{ at least once in } N \text{ years}) = 1 - \left(1 - \frac{1}{T}\right)^N \qquad (12.1.4)$$

Example 12.1.1. Estimate the probability that the annual maximum discharge Q on the Guadalupe River will exceed 50,000 cfs at least once during the next three years.

Solution. From the discussion above, $P(Q \geq 50{,}000$ cfs in any year$) \approx 0.195$, so from Eq. (12.1.3)

$$\begin{aligned} P(Q \geq 50{,}000 \text{ cfs at least once during the next 3 years}) &= 1 - (1 - 0.195)^3 \\ &= 0.48 \end{aligned}$$

The problem in Example 12.1.1 could have been phrased, "What is the probability that the discharge on the Guadalupe River will exceed 50,000 cfs at least once during the next three years?" The calculation given used only the annual maximum data, but, alternatively, all exceedences of 50,000 cfs contained in the Guadalupe River record could have been considered. This set of data is called the *partial duration series*. It will contain more than the nine exceedences shown in Table 12.1.2 if there were two or more exceedences of 50,000 cfs within some single year of record.

Hydrologic Data Series

A *complete duration series* consists of all the data available as shown in Fig. 12.1.2(a). A *partial duration series* is a series of data which are selected so that their magnitude is greater than a predefined *base value*. If the base value is selected so that the number of values in the series is equal to the number of years of the record, the series is called an *annual exceedence series;* an example is shown in Fig. 12.1.2(b). An *extreme value series* includes the largest or smallest values occurring in each of the equally-long time intervals of the record. The time interval length is usually taken as one year, and a series so selected is called an *annual series*. Using largest annual values, it is an *annual maximum series* as shown in Fig. 12.1.2(c). Selecting the smallest annual values produces an *annual minimum series*.

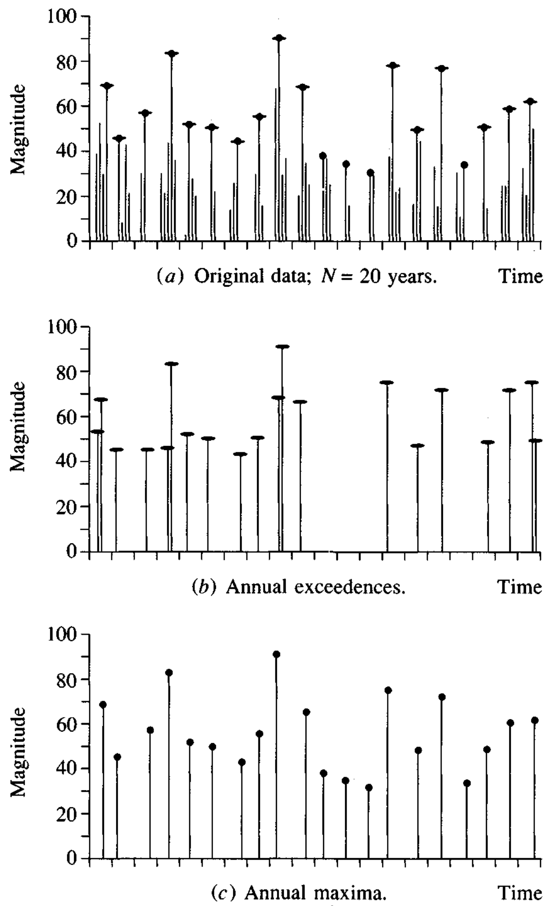

FIGURE 12.1.2
Hydrologic data arranged by time of occurrence. (*Source:* Chow, 1964. Used with permission.)

The annual maximum values and the annual exceedence values of the hypothetical data in Fig. 12.1.3(*a*) are arranged graphically in Fig. 12.1.3(*b*) in order of magnitude. In this particular example, only 16 of the 20 annual maxima appear in the annual exceedence series; the second largest value in several years outranks some annual maxima in magnitude. However, in the annual maximum series, these second largest values are excluded, resulting in the neglect of their effect in the analysis.

The return period T_E of event magnitudes developed from an annual exceedence series is related to the corresponding return period T for magnitudes derived from an annual maximum series by (Chow, 1964)

$$T_E = \left[\ln\left(\frac{T}{T-1}\right)\right]^{-1} \tag{12.1.5}$$

Although the annual exceedence series is useful for some purposes, it is limited by the fact that it may be difficult to verify that all the observations are

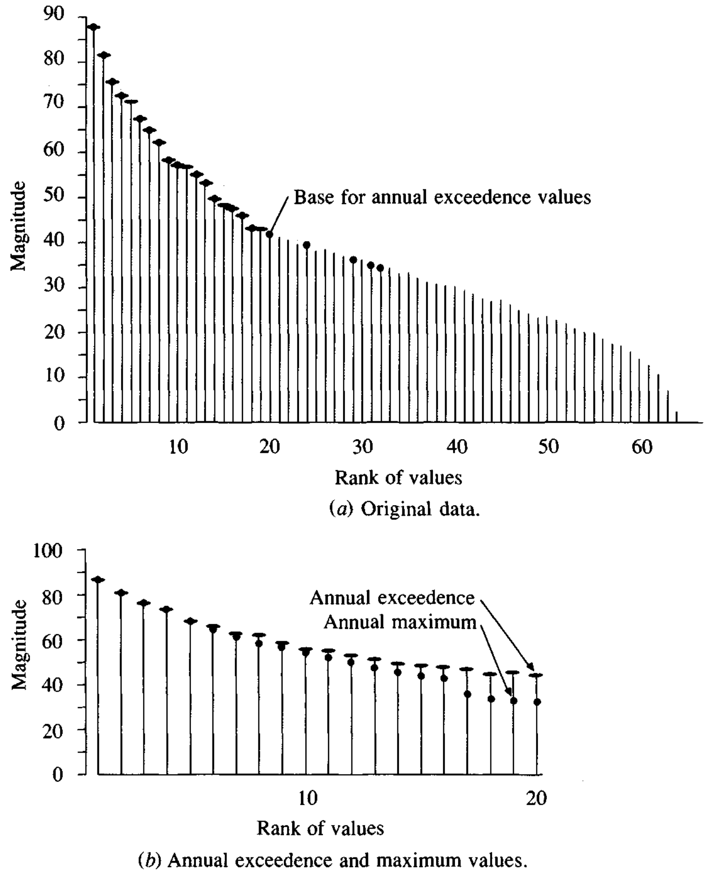

FIGURE 12.1.3
Hydrologic data arranged in the order of magnitude. (*Source:* Chow, 1964. Used with permission.)

independent—the occurrence of a large flood could well be related to saturated soil conditions produced during another large flood occurring a short time earlier. As a result, it is usually better to use the annual maximum series for analysis. In any case, as the return period of the event being considered becomes large, the results from the two approaches become very similar because the chance that two such events will occur within any year is very small.

12.2 EXTREME VALUE DISTRIBUTIONS

The study of extreme hydrologic events involves the selection of a sequence of the largest or smallest observations from sets of data. For example, the study of peak flows uses just the largest flow recorded each year at a gaging station out of the many thousands of values recorded. In fact, water level is usually recorded every 15 minutes, so there are $4 \times 24 = 96$ values recorded each day, and 365 × 96 = 35,040 values recorded each year; so the annual maximum flow event used for flood flow frequency analysis is the largest of more than 35,000 observations during that year. And this exercise is carried out for each year of historical data.

Since these observations are located in the extreme tail of the probability distribution of all observations from which they are drawn (the parent population), it is not surprising that their probability distribution is different from that of the parent population. As described in Sec. 11.5, there are three asymptotic forms of the distributions of extreme values, named Type I, Type II, and Type III, respectively.

The Extreme Value Type I (EVI) probability distribution function is

$$F(x) = \exp\left[-\exp\left(-\frac{x-u}{\alpha}\right)\right] \qquad -\infty \le x \le \infty \tag{12.2.1}$$

The parameters are estimated, as given in Table 11.5.1, by

$$\alpha = \frac{\sqrt{6}\, s}{\pi} \tag{12.2.2}$$

$$u = \bar{x} - 0.5772\alpha \tag{12.2.3}$$

The parameter u is the mode of the distribution (point of maximum probability density). A *reduced variate* y can be defined as

$$y = \frac{x-u}{\alpha} \tag{12.2.4}$$

Substituting the reduced variate into (12.2.1) yields

$$F(x) = \exp\left[-\exp(-y)\right] \tag{12.2.5}$$

Solving for y:

$$y = -\ln\left[\ln\left(\frac{1}{F(x)}\right)\right] \tag{12.2.6}$$

Let (12.2.6) be used to define y for the Type II and Type III distributions. The values of x and y can be plotted as shown in Fig. 12.2.1. For the EVI distribution the plot is a straight line while, for large values of y, the corresponding curve for the EVII distribution slopes more steeply than for EVI, and the curve for the EVIII distribution slopes less steeply, being bounded from above. Figure 12.2.1 also shows values of the return period T as an alternate axis to y. As shown by Eq. (12.1.2),

$$\begin{aligned}\frac{1}{T} &= P(x \ge x_T)\\ &= 1 - P(x < x_T)\\ &= 1 - F(x_T)\end{aligned}$$

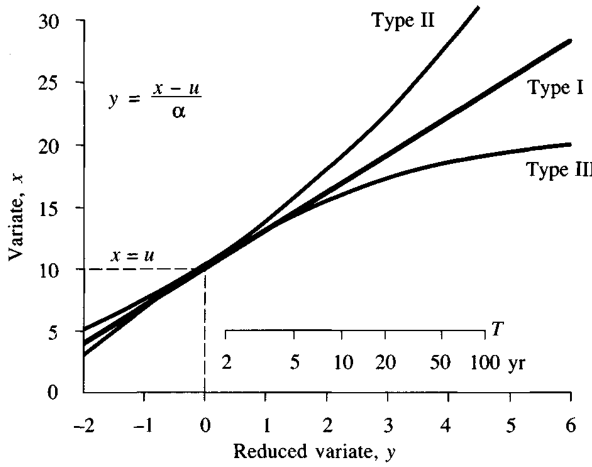

FIGURE 12.2.1
For each of the three types of extreme value distributions the variate x is plotted against a reduced variate y calculated for the Extreme Value Type I distribution. The Type I distribution is unbounded in x, while the Type II distribution has a lower bound and the Type III distribution has an upper bound. (*Source:* Natural Environment Research Council, 1975, Fig. 1.10, p. 41. Used with permission.)

so

$$F(x_T) = \frac{T-1}{T}$$

and, substituting into (12.2.6),

$$y_T = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right] \tag{12.2.7}$$

For the EVI distribution, x_T is related to y_T by Eq. (12.2.4), or

$$x_T = u + \alpha y_T \tag{12.2.8}$$

Extreme value distributions have been widely used in hydrology. They form the basis for the standardized method of flood frequency analysis in Great Britain (Natural Environment Research Council, 1975). Storm rainfalls are most commonly modeled by the Extreme Value Type I distribution (Chow, 1953; Tomlinson, 1980), and drought flows by the Weibull distribution, that is, the EVIII distribution applied to $-x$ (Gumbel, 1954, 1963).

Example 12.2.1. Annual maximum values of 10-minute-duration rainfall at Chicago, Illinois, from 1913 to 1947 are presented in Table 12.2.1. Develop a model for storm rainfall frequency analysis using the Extreme Value Type I distribution and calculate the 5-, 10-, and 50-year return period maximum values of 10-minute rainfall at Chicago.

Solution. The sample moments calculated from the data in Table 12.2.1 are $\bar{x}$ = 0.649 in and s = 0.177 in. Substituting into Eqs. (12.2.2) and (12.2.3) yields

$$\alpha = \frac{\sqrt{6}\, s}{\pi}$$

$$= \frac{\sqrt{6} \times 0.177}{\pi}$$

$$= 0.138$$

$$u = \bar{x} - 0.5772\alpha$$

$$= 0.649 - 0.5772 \times 0.138$$

$$= 0.569$$

The probability model is

$$F(x) = \exp\left[-\exp\left(-\frac{x - 0.569}{0.138}\right)\right]$$

To determine the values of x_T for various values of return period T, it is convenient to use the reduced variate y_T. For $T = 5$ years, Eq. (12.2.7) gives

$$y_T = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right]$$

$$= -\ln\left[\ln\left(\frac{5}{5-1}\right)\right]$$

$$= 1.500$$

and Eq. (12.2.8) yields

$$x_T = u + \alpha y_T$$

$$= 0.569 + 0.138 \times 1.500$$

$$= 0.78 \text{ in}$$

So the 10-minute, 5 year storm rainfall magnitude at Chicago is 0.78 in. By the same method, the 10- and 50-year values can be shown to be 0.88 in and 1.11 in,

TABLE 12.2.1

Annual maximum 10-minute rainfall in inches at Chicago, Illinois, 1913–1947

Year	1910	1920	1930	1940
0		0.53	0.33	0.34
1		0.76	0.96	0.70
2		0.57	0.94	0.57
3	0.49	0.80	0.80	0.92
4	0.66	0.66	0.62	0.66
5	0.36	0.68	0.71	0.65
6	0.58	0.68	1.11	0.63
7	0.41	0.61	0.64	0.60
8	0.47	0.88	0.52	
9	0.74	0.49	0.64	

Mean = 0.649 in
Standard deviation = 0.177 in

respectively. It may be noted from the data in Table 12.2.1 that the 50-year return period rainfall was equaled once in the 35 years of data (in 1936), and that the 10-year return period rainfall was equaled or exceeded four times during this period, so the frequency of occurrence of observed extreme rainfalls is approximately as predicted by the model.

12.3 FREQUENCY ANALYSIS USING FREQUENCY FACTORS

Calculating the magnitudes of extreme events by the method outlined in Example 12.2.1 requires that the probability distribution function be invertible, that is, given a value for T or $[F(x_T) = T/(T-1)]$, the corresponding value of x_T can be determined. Some probability distribution functions are not readily invertible, including the Normal and Pearson Type III distributions, and an alternative method of calculating the magnitudes of extreme events is required for these distributions.

The magnitude x_T of a hydrologic event may be represented as the mean μ plus the departure Δx_T of the variate from the mean (see Fig. 12.3.1):

$$x_T = \mu + \Delta x_T \tag{12.3.1}$$

The departure may be taken as equal to the product of the standard deviation σ and a *frequency factor* K_T; that is, $\Delta x_T = K_T\sigma$. The departure Δx_T and the frequency factor K_T are functions of the return period and the type of probability distribution to be used in the analysis. Equation (12.3.1) may therefore be expressed as

$$x_T = \mu + K_T\sigma \tag{12.3.2}$$

which may be approximated by

$$x_T = \bar{x} + K_T s \tag{12.3.3}$$

In the event that the variable analyzed is $y = \log x$, then the same method is applied to the statistics for the logarithms of the data, using

$$y_T = \bar{y} + K_T s_y \tag{12.3.4}$$

and the required value of x_T is found by taking the antilog of y_T.

The frequency factor equation (12.3.2) was proposed by Chow (1951), and it is applicable to many probability distributions used in hydrologic frequency analysis. For a given distribution, a K–T relationship can be determined between the frequency factor and the corresponding return period. This relationship can be expressed in mathematical terms or by a table.

Frequency analysis begins with the calculation of the statistical parameters required for a proposed probability distribution by the method of moments from the given data. For a given return period, the frequency factor can be determined from the K–T relationship for the proposed distribution, and the magnitude x_T computed by Eq. (12.3.3), or (12.3.4).

The theoretical K–T relationships for several probability distributions commonly used in hydrologic frequency analysis are now described.

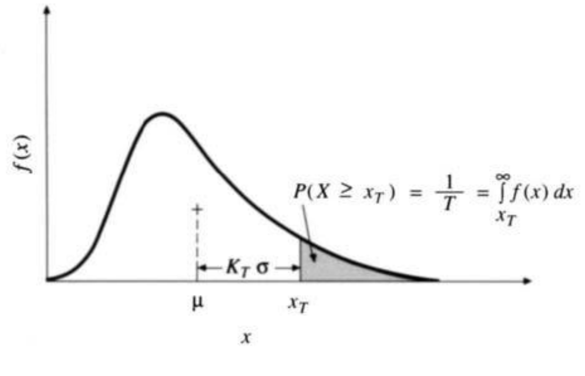

FIGURE 12.3.1
The magnitude of an extreme event x_T expressed as a deviation $K_T\sigma$ from the mean μ, where K_T is the frequency factor.

NORMAL DISTRIBUTION. The frequency factor can be expressed from Eq. (12.3.2) as

$$K_T = \frac{x_T - \mu}{\sigma} \tag{12.3.5}$$

This is the same as the standard normal variable z defined in Eq. (11.2.9).

The value of z corresponding to an exceedence probability of p ($p = 1/T$) can be calculated by finding the value of an intermediate variable w:

$$w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2} \qquad (0 < p \leq 0.5) \tag{12.3.6}$$

then calculating z using the approximation

$$z = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3} \tag{12.3.7}$$

When $p > 0.5$, $1 - p$ is substituted for p in (12.3.6) and the value of z computed by (12.3.7) is given a negative sign. The error in this formula is less than 0.00045 in z (Abramowitz and Stegun, 1965). The frequency factor K_T for the normal distribution is equal to z, as mentioned above.

For the lognormal distribution, the same procedure applies except that it is applied to the logarithms of the variables, and their mean and standard deviation are used in Eq. (12.3.4).

Example 12.3.1. Calculate the frequency factor for the normal distribution for an event with a return period of 50 years.

Solution. For $T = 50$ years, $p = 1/50 = 0.02$. From Eq. (12.3.6)

$$\begin{aligned} w &= \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2} \\ &= \left[\ln\left(\frac{1}{0.02^2}\right)\right]^{1/2} \\ &= 2.7971 \end{aligned}$$

Then, substituting w into (12.3.7)

$$
\begin{aligned}
K_T &= z \\
&= 2.7971 - \frac{2.51557 + 0.80285 \times 2.7971 + 0.01033 \times (2.7971)^2}{1 + 1.43279 \times 2.7971 + 0.18927 \times (2.7971)^2 + 0.00131 \times (2.7971)^3} \\
&= 2.054
\end{aligned}
$$

EXTREME VALUE DISTRIBUTIONS. For the Extreme Value Type I distribution, Chow (1953) derived the expression

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \tag{12.3.8}$$

To express T in terms of K_T, the above equation can be written as

$$T = \frac{1}{1 - \exp\left\{-\exp\left[-\left(\gamma + \frac{\pi K_T}{\sqrt{6}}\right)\right]\right\}} \tag{12.3.9}$$

where $\gamma = 0.5772$. When $x_T = \mu$, Eq. (12.3.5) gives $K_T = 0$ and Eq. (12.3.8) gives $T = 2.33$ years. This is the return period of the mean of the Extreme Value Type I distribution. For the Extreme Value Type II distribution, the logarithm of the variate follows the EVI distribution. For this case, (12.3.4) is used to calculate y_T, using the value of K_T from (12.3.8).

Example 12.3.2. Determine the 5-year return period rainfall for Chicago using the frequency factor method and the annual maximum rainfall data given in Table 12.2.1.

Solution. The mean and standard deviation of annual maximum rainfalls at Chicago are $\bar{x} = 0.649$ in and $s = 0.177$ in, respectively. For T=5, Eq. (12.3.8) gives

$$
\begin{aligned}
K_T &= -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \\
&= -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{5}{5-1}\right)\right]\right\} \\
&= 0.719
\end{aligned}
$$

By (12.3.3),

$$
\begin{aligned}
x_T &= \bar{x} + K_T s \\
&= 0.649 + 0.719 \times 0.177 \\
&= 0.78 \text{ in}
\end{aligned}
$$

as determined in Example 12.2.1.

LOG–PEARSON TYPE III DISTRIBUTION. For this distribution, the first step is to take the logarithms of the hydrologic data, $y = \log x$. Usually logarithms to base 10 are used. The mean $\bar{y}$, standard deviation s_y, and coefficient of skewness C_s are calculated for the logarithms of the data. The frequency factor depends on the return period T and the coefficient of skewness C_s. When $C_s = 0$, the frequency factor is equal to the standard normal variable z. When $C_s \neq 0, K_T$ is approximated by Kite (1977) as

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5 \qquad (12.3.10)$$

where $k = C_s/6$.

TABLE 12.3.1
K_T values for Pearson Type III distribution (positive skew)

	Return period in years						
	2	5	10	25	50	100	200
Skew coefficient	Exceedence probability						
C_s or C_w	**0.50**	**0.20**	**0.10**	**0.04**	**0.02**	**0.01**	**0.005**
3.0	−0.396	0.420	1.180	2.278	3.152	4.051	4.970
2.9	−0.390	0.440	1.195	2.277	3.134	4.013	4.909
2.8	−0.384	0.460	1.210	2.275	3.114	3.973	4.847
2.7	−0.376	0.479	1.224	2.272	3.093	3.932	4.783
2.6	−0.368	0.499	1.238	2.267	3.071	3.889	4.718
2.5	−0.360	0.518	1.250	2.262	3.048	3.845	4.652
2.4	−0.351	0.537	1.262	2.256	3.023	3.800	4.584
2.3	−0.341	0.555	1.274	2.248	2.997	3.753	4.515
2.2	−0.330	0.574	1.284	2.240	2.970	3.705	4.444
2.1	−0.319	0.592	1.294	2.230	2.942	3.656	4.372
2.0	−0.307	0.609	1.302	2.219	2.912	3.605	4.298
1.9	−0.294	0.627	1.310	2.207	2.881	3.553	4.223
1.8	−0.282	0.643	1.318	2.193	2.848	3.499	4.147
1.7	−0.268	0.660	1.324	2.179	2.815	3.444	4.069
1.6	−0.254	0.675	1.329	2.163	2.780	3.388	3.990
1.5	−0.240	0.690	1.333	2.146	2.743	3.330	3.910
1.4	−0.225	0.705	1.337	2.128	2.706	3.271	3.828
1.3	−0.210	0.719	1.339	2.108	2.666	3.211	3.745
1.2	−0.195	0.732	1.340	2.087	2.626	3.149	3.661
1.1	−0.180	0.745	1.341	2.066	2.585	3.087	3.575
1.0	−0.164	0.758	1.340	2.043	2.542	3.022	3.489
0.9	−0.148	0.769	1.339	2.018	2.498	2.957	3.401
0.8	−0.132	0.780	1.336	1.993	2.453	2.891	3.312
0.7	−0.116	0.790	1.333	1.967	2.407	2.824	3.223
0.6	−0.099	0.800	1.328	1.939	2.359	2.755	3.132
0.5	−0.083	0.808	1.323	1.910	2.311	2.686	3.041
0.4	−0.066	0.816	1.317	1.880	2.261	2.615	2.949
0.3	−0.050	0.824	1.309	1.849	2.211	2.544	2.856
0.2	−0.033	0.830	1.301	1.818	2.159	2.472	2.763
0.1	−0.017	0.836	1.292	1.785	2.107	2.400	2.670
0.0	0	0.842	1.282	1.751	2.054	2.326	2.576

The value of z for a given return period can be calculated by the procedure used in Example 12.3.1. Table 12.3.1 gives values of the frequency factor for the Pearson Type III (and log–Pearson Type III) distribution for various values of the return period and coefficient of skewness.

Example 12.3.3. Calculate the 5- and 50-year return period annual maximum discharges of the Guadalupe River near Victoria, Texas, using the lognormal and log–Pearson Type III distributions. The data from 1935 to 1978 are given in Table 12.1.1.

TABLE 12.3.1 *(cont.)*
K_T values for Pearson Type III distribution (negative skew)

	Return period in years						
	2	**5**	**10**	**25**	**50**	**100**	**200**
Skew coefficient	Exceedence probability						
C_s or C_w	**0.50**	**0.20**	**0.10**	**0.04**	**0.02**	**0.01**	**0.005**
–0.1	0.017	0.846	1.270	1.716	2.000	2.252	2.482
–0.2	0.033	0.850	1.258	1.680	1.945	2.178	2.388
–0.3	0.050	0.853	1.245	1.643	1.890	2.104	2.294
–0.4	0.066	0.855	1.231	1.606	1.834	2.029	2.201
–0.5	0.083	0.856	1.216	1.567	1.777	1.955	2.108
–0.6	0.099	0.857	1.200	1.528	1.720	1.880	2.016
–0.7	0.116	0.857	1.183	1.488	1.663	1.806	1.926
–0.8	0.132	0.856	1.166	1.448	1.606	1.733	1.837
–0.9	0.148	0.854	1.147	1.407	1.549	1.660	1.749
–1.0	0.164	0.852	1.128	1.366	1.492	1.588	1.664
–1.1	0.180	0.848	1.107	1.324	1.435	1.518	1.581
–1.2	0.195	0.844	1.086	1.282	1.379	1.449	1.501
–1.3	0.210	0.838	1.064	1.240	1.324	1.383	1.424
–1.4	0.225	0.832	1.041	1.198	1.270	1.318	1.351
–1.5	0.240	0.825	1.018	1.157	1.217	1.256	1.282
–1.6	0.254	0.817	0.994	1.116	1.166	1.197	1.216
–1.7	0.268	0.808	0.970	1.075	1.116	1.140	1.155
–1.8	0.282	0.799	0.945	1.035	1.069	1.087	1.097
–1.9	0.294	0.788	0.920	0.996	1.023	1.037	1.044
–2.0	0.307	0.777	0.895	0.959	0.980	0.990	0.995
–2.1	0.319	0.765	0.869	0.923	0.939	0.946	0.949
–2.2	0.330	0.752	0.844	0.888	0.900	0.905	0.907
–2.3	0.341	0.739	0.819	0.855	0.864	0.867	0.869
–2.4	0.351	0.725	0.795	0.823	0.830	0.832	0.833
–2.5	0.360	0.711	0.771	0.793	0.798	0.799	0.800
–2.6	0.368	0.696	0.747	0.764	0.768	0.769	0.769
–2.7	0.376	0.681	0.724	0.738	0.740	0.740	0.741
–2.8	0.384	0.666	0.702	0.712	0.714	0.714	0.714
–2.9	0.390	0.651	0.681	0.683	0.689	0.690	0.690
–3.0	0.396	0.636	0.666	0.666	0.666	0.667	0.667

Source: U. S. Water Resources Council (1981).

Solution. The logarithms of the discharge values are taken and their statistics calculated: $\bar{y} = 4.2743, s_y = 0.4027, C_s = -0.0696$.

Lognormal distribution. The frequency factor can be obtained from Eq. (12.3.7), or from Table 12.3.1 for coefficient of skewness 0. For $T = 50$ years, K_T was computed in Example 12.3.1 as $K_{50} = 2.054$; the same value can be obtained from Table 12.3.1. By (12.3.4)

$$\begin{aligned} y_T &= \bar{y} + K_T s_y \\ y_{50} &= 4.2743 + 2.054 \times 0.4027 \\ &= 5.101 \end{aligned}$$

Then

$$\begin{aligned} x_{50} &= (10)^{5.101} \\ &= 126,300 \text{ cfs} \end{aligned}$$

Similarly, $K_5 = 0.842$ from Table 12.3.1, $y_5 = 4.2743 + 0.842 \times 0.4027 = 4.6134$, and $x_5 = (10)^{4.6134} = 41,060$ cfs.

Log–Pearson Type III distribution. For $C_s = -0.0696$, the value of K_{50} is obtained by interpolation from Table 12.3.1 or by Eq. (12.3.10). By interpolation with T = 50 yrs:

$$K_{50} = 2.054 + \frac{(2.00 - 2.054)}{(-0.1 - 0)}(-0.0696 - 0) = 2.016$$

So $y_{50} = \bar{y} + K_{50}s_y = 4.2743 + 2.016 \times 0.4027 = 5.0863$ and $x_{50} = (10)^{5.0863} =$ 121,990 cfs. By a similar calculation, $K_5 = 0.845, y_5 = 4.6146$, and $x_5 = 41,170$ cfs.

The results for estimated annual maximum discharges are:

	Return Period	
	5 years	**50 years**
Lognormal ($C_s = 0$)	41,060	126,300
Log-Pearson Type III ($C_s = -0.07$)	41,170	121,990

It can be seen that the effect of including the small negative coefficient of skewness in the calculations is to alter slightly the estimated flow with that effect being more pronounced at $T = 50$ years than at $T = 5$ years. Another feature of the results is that the 50-year return period estimates are about three times as large as the 5-year return period estimates; for this example, the increase in the estimated flood discharges is less than proportional to the increase in return period.

12.4 PROBABILITY PLOTTING

As a check that a probability distribution fits a set of hydrologic data, the data may be plotted on specially designed *probability paper*, or using a plotting scale that linearizes the distribution function. The plotted data are then fitted with a straight line for interpolation and extrapolation purposes.

Probability Paper

The cumulative probability of a theoretical distribution may be represented graphically on probability paper designed for the distribution. On such paper the ordinate usually represents the value of x in a certain scale and the abscissa represents the probability $P(X \geq x)$ or $P(X < x)$, the return period T, or the reduced variate y_T. The ordinate and abscissa scales are so designed that the data to be fitted are expected to appear close to a straight line. The purpose of using the probability paper is to linearize the probability relationship so that the plotted data can be easily used for interpolation, extrapolation, or comparison purposes. In the case of extrapolation, however, the effect of various errors is often magnified; therefore, hydrologists should be warned against such practice if no consideration is given to this effect.

Plotting Positions

Plotting position refers to the probability value assigned to each piece of data to be plotted. Numerous methods have been proposed for the determination of plotting positions, most of which are empirical. If n is the total number of values to be plotted and m is the rank of a value in a list ordered by descending magnitude, the exceedence probability of the mth largest value, x_m, is, for large n,

$$P(X \geq x_m) = \frac{m}{n} \tag{12.4.1}$$

However, this simple formula (known as California's formula) produces a probability of 100 percent for $m = n$, which may not be easily plotted on a probability scale. As an adjustment, the above formula may be modified to

$$P(X \geq x_m) = \frac{m-1}{n} \tag{12.4.2}$$

While this formula does not produce a probability of 100 percent, it yields a zero probability (for $m = 1$), which may not be easily plotted on probability paper either.

The above two formulas represent the limits within which suitable plotting positions should lie. One compromise of the two formulas is

$$P(X \geq x_m) = \frac{m-0.5}{n} \tag{12.4.3}$$

which was first proposed by Hazen (1930). Another compromising formula (known as Chegodayev's) widely used in the U.S.S.R. and Eastern European countries is

$$P(X \geq x_m) = \frac{m-0.3}{n+0.4} \tag{12.4.4}$$

The Weibull formula is a compromise with more statistical justification. If the n values are distributed uniformly between 0 and 100 percent probability, then there must be $n + 1$ intervals, $n - 1$ between the data points and 2 at the ends. This simple plotting system is expressed by the Weibull formula:

$$P(X \geq x_m) = \frac{m}{n + 1} \tag{12.4.5}$$

indicating a return period one year longer than the period of record for the largest value.

In practice, for a complete duration series (employing all the data, not just selected extreme values), Eq. (12.4.1) is used, with n referring to the number of items in the data rather than to the number of years. For annual maximum series, Eq. (12.4.5), which is equivalent to the following formula for return period, was adopted as the standard plotting position method by the U. S. Water Resources Council (1981):

$$T = \frac{n + 1}{m} \tag{12.4.6}$$

where n refers to the number of years in the record.

Most plotting position formulas are represented by the following form:

$$P(X \geq x_m) = \frac{m - b}{n + 1 - 2b} \tag{12.4.7}$$

where b is a parameter. For example, $b = 0.5$ for Hazen's formula, $b = 0.3$ for Chegodayev's, and $b = 0$ for Weibull's. Also, for some other examples $b = 3/8$ for Blom's formula, 1/3 for Tukey's, and 0.44 for Gringorten's (see Chow, 1964).

Cunnane (1978) studied the various available plotting position methods using criteria of *unbiasedness* and *minimum variance*. An unbiased plotting method is one that, if used for plotting a large number of equally sized samples, will result in the average of the plotted points for each value of m falling on the theoretical distribution line. A minimum variance plotting method is one that minimizes the variance of the plotted points about the theoretical line. Cunnane concluded that the Weibull plotting formula is biased and plots the largest values of a sample at too small a return period. For normally distributed data, he found that the Blom (1958) plotting position ($b = 3/8$) is closest to being unbiased, while for data distributed according to the Extreme Value Type I distribution, the Gringorten (1963) formula ($b = 0.44$) is the best. For the log–Pearson Type III distribution, the optimal value of b depends on the value of the coefficient of skewness, being larger than 3/8 when the data are positively skewed and smaller than 3/8 when the data are negatively skewed. The same plotting positions can be applied to the logarithms of the data, when using the lognormal distribution, for example.

Once the data series is identified and ranked, and the plotting positions calculated, a graph of magnitude (x) vs. probability [$(P(X > x), P(X < x)$, or T)] can be plotted to graphically fit a distribution. Alternatively, an analytical fit can be made using the method of moments, and the resulting fitted line compared with the sample data.

Example 12.4.1. Perform a probability plotting analysis of the annual maximum discharges of the Guadalupe River near Victoria, Texas, given in Table 12.1.1. Compare the plotted data with the lognormal distribution fitted to them in Example 12.3.3.

Solution. First the data are ranked from largest ($m = 1$), to smallest ($m = n = 44$), as shown in columns 1 and 2 of Table 12.4.1. Blom's plotting formula is used, since the logarithms of the data are being fitted to a normal distribution. Blom's formula uses $b = 3/8$ in Eq. (12.4.7). For example, for $m = 1$, the exceedence probability $P(Q \geq 179{,}000 \text{ cfs}) \approx (m - 3/8)/(n + 1 - 6/8) = (1 - 3/8)/(44 + 1/4) = 0.014$, as shown in column 3 of Table 12.4.1. The corresponding value of the standard normal variable z is determined using $p = 0.014$ in Eqs. (12.3.6) and (12.3.7) in the manner shown in Example 12.3.1; the result, $z = 2.194$, is listed in column 4 of the table. The event magnitude with the same exceedence probability in the fitted lognormal distribution is found using the frequency factor method with $\bar{y} = 4.2743$, $s_y = 0.4027$, and $K_T = z = 2.194$; the result is $\log Q = 4.2743 + 2.194 \times 0.4027 = 5.158$ (column 5). This value is compared with $\log Q$ from the observed data, that is log (179,000) = 5.253, as shown in column 6. The observed data are plotted against the fitted curve in Fig. 12.4.1, in which the value of the standard normal variable is used as the horizontal axis to linearize the plot; this is equivalent to using normal probability plotting paper. The plot shows that the fitted line is consistent with the observed data, even including the largest value of 179,000 cfs, which looks quite different from the rest of the data in Fig. 12.1.1.

TABLE 12.4.1

Probability plotting using the normal distribution and Blom's formula for the annual maximum discharges of the Guadalupe River near Victoria, Texas (Example 12.4.1)

Column:	1	2	3	4	5	6
	Discharge Q (cfs)	Rank m	Exceedence probability $\frac{m - 3/8}{n + 1/4}$	Standard normal variable z	Log Q from lognormal distribution	Log Q from data
	179,000	1	0.014	2.194	5.158	5.253
	70,000	2	0.037	1.790	4.995	4.845
	58,500	3	0.059	1.561	4.903	4.767
	58,300	4	0.082	1.393	4.835	4.766
	58,000	5	0.105	1.256	4.780	4.763
	·	·	·	·	·	·
	·	·	·	·	·	·
	·	·	·	·	·	·
	5,720	40	0.895	−1.256	3.768	3.757
	4,950	41	0.918	−1.393	3.714	3.695
	4,940	42	0.941	−1.561	3.646	3.694
	4,100	43	0.963	−1.790	3.553	3.613
	1,730	44	0.986	−2.194	3.391	3.238

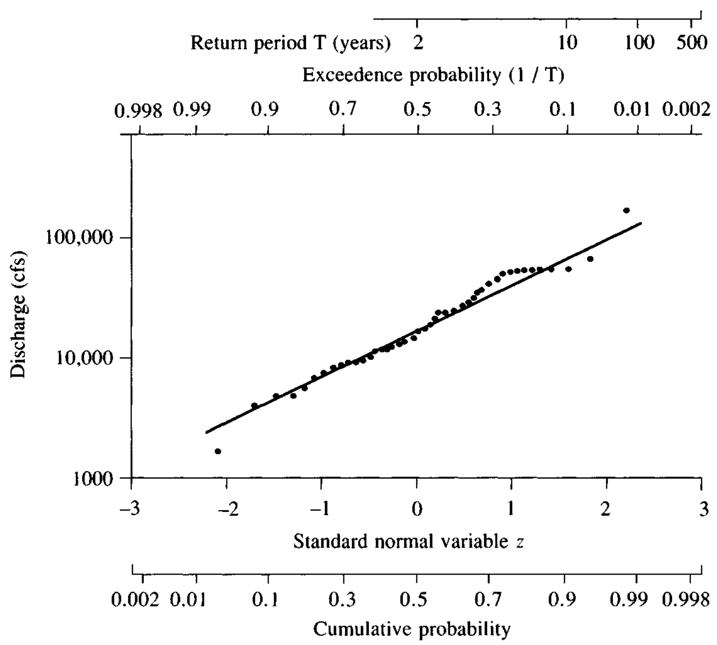

FIGURE 12.4.1
Annual maximum discharge for the Guadalupe River near Victoria, Texas, plotted using Blom's formula on a probability scale for the lognormal distribution.

12.5 WATER RESOURCES COUNCIL METHOD

The U. S. Water Resources Council* recommended that the log–Pearson Type III distribution be used as a base distribution for flood flow frequency studies (U. S. Water Resources Council, 1967, 1976, 1977, and 1981; Benson, 1968). Their decision was an attempt to promote a consistent, uniform approach to flood flow frequency determination for use in all federal planning involving water and related land resources. The choice of the log–Pearson Type III distribution is, however, subjective to some extent, in that there are several criteria that may be employed to select the best distribution, and no single probability distribution is the best under all criteria.

Determination of the Coefficient of Skewness

The coefficient of skewness used in fitting the log–Pearson Type III distribution is very sensitive to the size of the sample and, in particular, is difficult to estimate

*The U.S. Water Resources Council was abolished in 1981. The Council's work on guidelines for determining flood flow frequency was taken over by the Interagency Advisory Committee on Water Data, U.S. Geological Survey, Reston, Virginia.

accurately from small samples. Because of this, the Water Resources Council recommended using a generalized estimate of the coefficient of skewness, C_w, based upon the equation

$$C_w = WC_s + (1 - W)C_m \tag{12.5.1}$$

where W is a weighting factor, C_s is the coefficient of skewness computed using the sample data, and C_m is a map skewness, which is read from a map such as Fig. 12.5.1. The weighting factor W is calculated so as to minimize the variance of C_w, as explained next.

The estimates of the sample skew coefficient and the map skew coefficient in Eq. (12.5.1) are assumed to be independent with the same mean and different variances, $V(C_s)$ and $V(C_m)$. The variance of the weighted skew, $V(C_w)$, can be expressed as

$$V(C_w) = W^2V(C_s) + (1 - W)^2V(C_m) \tag{12.5.2}$$

The value of W that minimizes the variance C_w can be determined by differentiating (12.5.2) with respect to W and solving $d[V(C_w)]/dW = 0$ for W to obtain

$$W = \frac{V(C_m)}{V(C_s) + V(C_m)} \tag{12.5.3}$$

The second derivative

$$\frac{d^2V(C_w)}{dW^2} = 2[V(C_s) + V(C_m)] \tag{12.5.4}$$

is greater than zero, confirming that the weight given by (12.5.3) minimizes the variance of the skew, $V(C_w)$.

Determination of W using Eq. (12.5.3) requires knowledge of $V(C_m)$ and $V(C_s)$. $V(C_m)$ is estimated from the map of skew coefficients for the United States as 0.3025. Alternatively, $V(C_m)$ can be derived from a regression study relating the skew to physiographical and meteorological characteristics of the basins (Tung and Mays, 1981).

By substituting Eq. (12.5.3) into Eq. (12.5.1), the weighted skew C_w can be written

$$C_w = \frac{V(C_m)C_s + V(C_s)C_m}{V(C_m) + V(C_s)} \tag{12.5.5}$$

The variance of the station skew C_s for log–Pearson Type III random variables can be obtained from the results of Monte Carlo experiments by Wallis, Matalas, and Slack (1974). They showed that $V(C_s)$ of the logarithmic station skew is a function of record length and population skew. For use in calculating C_w, this function can be approximated with sufficient accuracy as

$$V(C_s) = 10^{A-B\log_{10}(n/10)} \tag{12.5.6}$$

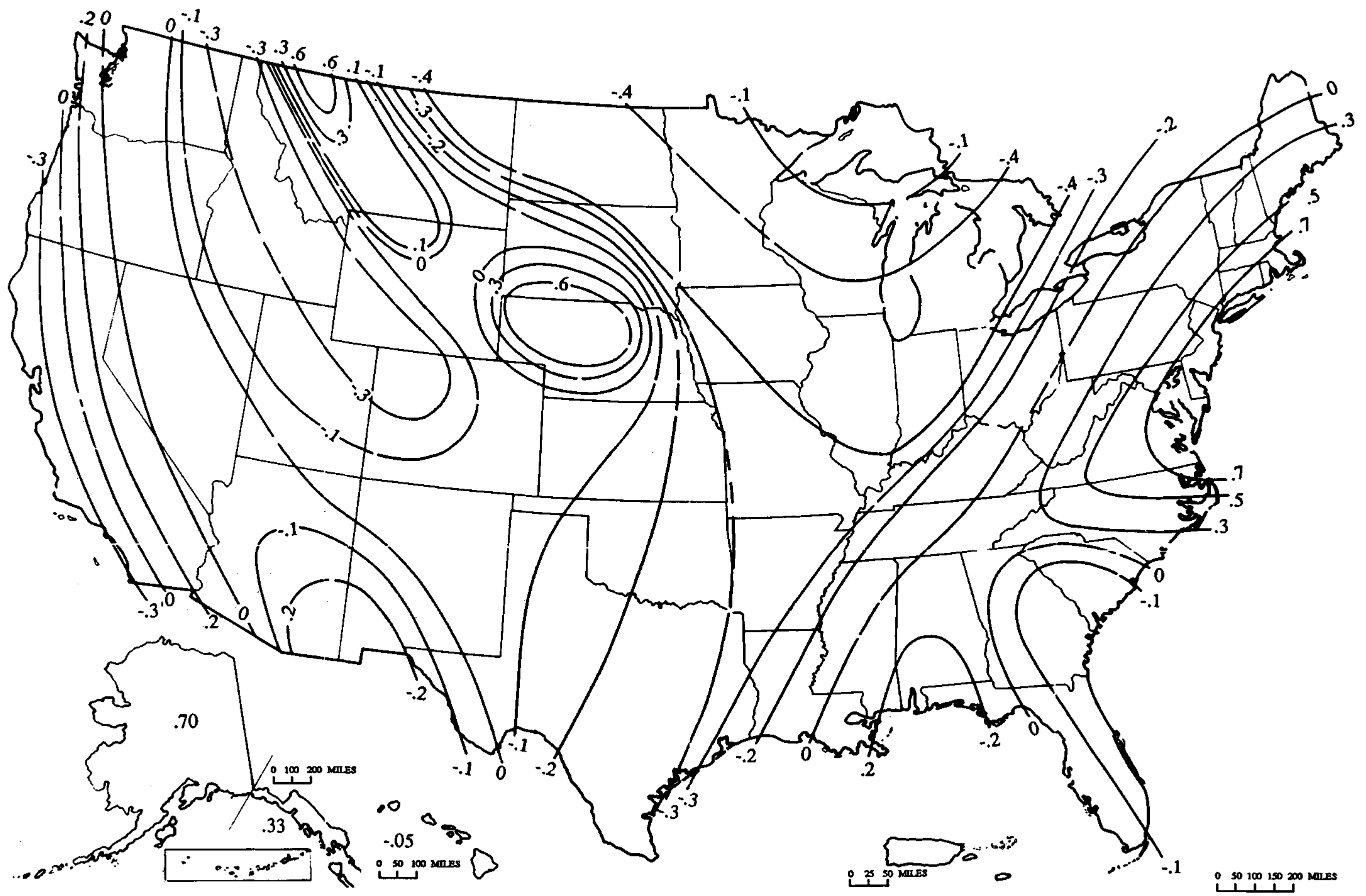

FIGURE 12.5.1
Generalized skew coefficients of annual maximum streamflow. (*Source:* Guidelines for determining flood flow frequency, Bulletin 17B, Hydrology Subcommittee, Interagency Advisory Committee on Water Data, U. S. Geological Survey, Reston, Va. Revised with corrections March 1982.)

where

$$A = -0.33 + 0.08|C_s| \quad \text{if} \quad |C_s| \le 0.90 \qquad (12.5.7a)$$

or

$$A = -0.52 + 0.30|C_s| \quad \text{if} \quad |C_s| > 0.90 \qquad (12.5.7b)$$

$$B = 0.94 - 0.26|C_s| \quad \text{if} \quad |C_s| \le 1.50 \qquad (12.5.7c)$$

or

$$B = 0.55 \quad \text{if} \quad |C_s| > 1.50 \qquad (12.5.7d)$$

in which $|C_s|$ is the absolute value of the station skew (used as an estimate of population skew) and n is the record length in years.

Example 12.5.1. Determine the frequency curve comprising the estimated flood magnitudes for return periods of 2, 5, 10, 25, 50, and 100 years using the Water Resources Council method for data from Walnut Creek at Martin Luther King Blvd. in Austin, Texas, as listed in Table 12.5.1.

Solution. The sample data shown in columns 1 and 2 of Table 12.5.1 cover $n = 16$ years, from 1967 to 1982.

Step 1. Transform the sample data, x_i, to their logarithmic values, y_i; that is, let $y_i = \log x_i$ for $i = 1, \ldots, n$, as shown in column 3 of the table.

TABLE 12.5.1
Calculation of statistics for logarithms of annual maximum discharges for Walnut Creek (Example 12.5.1)

Column:	1	2	3	4	5
	Year	Flow x (cfs)	$y = \log x$	$(y - \bar{y})^2$	$(y - \bar{y})^3$
	1967	303	2.4814	1.3395	−1.5502
	1968	5,640	3.7513	0.0127	0.0014
	1969	1,050	3.0212	0.3814	−0.2356
	1970	6,020	3.7796	0.0198	0.0028
	1971	3,740	3.5729	0.0043	−0.0003
	1972	4,580	3.6609	0.0005	0.0000
	1973	5,140	3.7110	0.0052	0.0004
	1974	10,560	4.0237	0.1481	0.0570
	1975	12,840	4.1086	0.2207	0.1037
	1976	5,140	3.7110	0.0052	0.0004
	1977	2,520	3.4014	0.0564	−0.0134
	1978	1,730	3.2380	0.1606	−0.0644
	1979	12,400	4.0934	0.2067	0.0940
	1980	3,400	3.5315	0.0115	−0.0012
	1981	14,300	4.1553	0.2668	0.1378
	1982	9,540	3.9795	0.1161	0.0396
	Total		58.2206	2.9555	−1.4280
	$n = 16$	$\bar{y} = 3.6388$			

Step 2. Compute the sample statistics. The mean of log-transformed values is

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{58.22}{16} = 3.639$$

Using column 4 of the table, the standard deviation is

$$s_y = \left(\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{1/2}$$

$$= \left(\frac{1}{15}2.9555\right)^{1/2}$$

$$= 0.4439$$

Using column 5 of the table, the skew coefficient is

$$C_s = \frac{n\sum_{i=1}^{n}(y_i - \bar{y})^3}{(n-1)(n-2)s_y^3} = \frac{16 \times (-1.4280)}{15 \times 14 \times (0.4439)^3} = -1.244$$

Step 3. Compute the weighted skew. The map skew is -0.3 from Fig. 12.5.1 at Austin, Texas. The variance of the station skew can be computed by Eq. (12.5.6) as follows. From (12.5.7*b*) with $|C_s| > 0.90$

$$A = -0.52 + 0.30|-1.244| = -0.147$$

From (12.5.7*c*) with $|C_s| < 1.50$

$$B = 0.94 - 0.26|-1.244| = 0.617$$

Then using (12.5.6)

$$V(C_s) = (10)^{-0.147-0.617\log(16/10)} = 0.533$$

The variance of the generalized skew is $V(C_m) = 0.303$. The weight to be applied to C_s is $W = V(C_m)/[V(C_m) + V(C_s)] = 0.303/(0.303 + 0.533) = 0.362$, and the complementary weight to be applied to C_m is $1 - W = 1 - 0.362 = 0.638$. Then, from (12.5.1)

$$C_w = WC_s + (1 - W)C_m$$

$$= 0.362 \times (-1.244) + 0.638 \times (-0.3)$$

$$= -0.64$$

Step 4. Compute the frequency curve coordinates. The log-Pearson Type III frequency factors K_T for skew coefficient values of -0.6 and -0.7 are found in Table 12.3.1. The values for $C_w = -0.64$ are found by linear interpolation as in Example 12.3.3, with results presented in column 2 of Table 12.5.2. The corresponding value of y_T is found from Eq. (12.3.4), and its antilogarithm is taken to determine the estimated flood magnitude. For example, for $T = 100$ years, $K_T = 1.850$ and

TABLE 12.5.2
Results of frequency analysis using the Water Resources Council method (Examples 12.5.1 and 12.5.2)

Column:	1 Return period T (years)	2 Frequency factor K_T	3 $\log Q_T$	4 Flood Estimates Q_T (cfs)	5 Q'_T (cfs)
	2	0.106	3.686	4,900	5,500
	5	0.857	4.019	10,500	10,000
	10	1.193	4.169	14,700	13,200
	25	1.512	4.310	20,400	17,600
	50	1.697	4.392	24,700	20,900
	100	1.850	4.460	28,900	24,200

The values in column 4 are those computed without adjustment for outliers and those in column 5 after outlier adjustment.

$$y_T = \bar{y} + K_T s_y$$
$$= 3.639 + 1.850 \times 0.4439$$
$$= 4.460$$

and $Q_T = (10)^{4.460} = 28,900$ cfs, as shown in columns 3 and 4 of the table. Similarly computed flood estimates for the other required return periods are also shown.

As was shown in Example (12.3.3), the increase in flood magnitude is less than directly proportional to the increase in return period. For example, increasing the return period from 10 years to 100 years approximately doubles the estimated flood magnitude in the table. As stated previously, flood magnitudes estimated using the log–Pearson Type III distribution are very sensitive to the value of the skew coefficient. The flood magnitudes for the longer return periods (50 and 100 years) are difficult to estimate reliably from only 16 years of data.

Testing for Outliers

The Water Resources Council method recommends that adjustments be made for outliers. *Outliers* are data points that depart significantly from the trend of the remaining data. The retention or deletion of these outliers can significantly affect the magnitude of statistical parameters computed from the data, especially for small samples. Procedures for treating outliers require judgment involving both mathematical and hydrologic considerations. According to the Water Resources Council (1981), if the station skew is greater than $+0.4$, tests for high outliers are considered first; if the station skew is less than -0.4, tests for low outliers are considered first. Where the station skew is between ± 0.4, tests for both high and low outliers should be applied before eliminating any outliers from the data set.

The following frequency equation can be used to detect high outliers:

$$y_H = \bar{y} + K_n s_y \tag{12.5.8}$$

where y_H is the high outlier threshold in log units and K_n is as given in Table 12.5.3 for sample size n. The K_n values in Table 12.5.3 are used in *one-sided tests* that detect outliers at the 10-percent level of significance in normally distributed data. If the logarithms of the values in a sample are greater than y_H in the above equation, then they are considered high outliers. Flood peaks considered high outliers should be compared with historic flood data and flood information at nearby sites. Historic flood data comprise information on unusually extreme events outside of the systematic record. According to the Water Resources Council (1981), if information is available that indicates a high outlier is the maximum over an extended period of time, the outlier is treated as historic flood data and excluded from analysis. If useful historic information is not available to compare to high outliers, then the outliers should be retained as part of the systematic record.

A similar equation can be used to detect low outliers:

$$y_L = \bar{y} - K_n s_y \tag{12.5.9}$$

where y_L is the low outlier threshold in log units. Flood peaks considered low outliers are deleted from the record and a conditional probability adjustment described by the Water Resources Council (1981) can be applied.

Example 12.5.2. Using the data for the Walnut Creek example (Table 12.5.1), determine if there are any high or low outliers for the sample. If so, omit them from the data set and recalculate the flood frequency curve.

TABLE 12.5.3
Outlier test K_n values

Sample size n	K_n	Sample size n	K_n	Sample size n	K_n	Sample size n	K_n
10	2.036	24	2.467	38	2.661	60	2.837
11	2.088	25	2.486	39	2.671	65	2.866
12	2.134	26	2.502	40	2.682	70	2.893
13	2.175	27	2.519	41	2.692	75	2.917
14	2.213	28	2.534	42	2.700	80	2.940
15	2.247	29	2.549	43	2.710	85	2.961
16	2.279	30	2.563	44	2.719	90	2.981
17	2.309	31	2.577	45	2.727	95	3.000
18	2.335	32	2.591	46	2.736	100	3.017
19	2.361	33	2.604	47	2.744	110	3.049
20	2.385	34	2.616	48	2.753	120	3.078
21	2.408	35	2.628	49	2.760	130	3.104
22	2.429	36	2.639	50	2.768	140	3.129
23	2.448	37	2.650	55	2.804		

Source: U.S. Water Resources Council, 1981. This table contains one-sided 10-percent significance level K_n values for the normal distribution.

Solution.

Step 1. Determine the threshold value for high outliers. From Table 12.5.3, $K_n = 2.279$ for $n = 16$ data. From Eq. (12.5.8) using $\bar{y}$ and s_y from Example 12.5.1

$$y_H = \bar{y} + K_n s_y = 3.639 + 2.279(0.4439) = 4.651$$

Then

$$Q_H = (10)^{4.651} = 44,735 \text{ cfs}$$

The largest recorded value (14,300 cfs in Table 12.5.1) does not exceed the threshold value, so there are no high outliers in this sample.

Step 2. Determine the threshold value for low outliers. The same K_n value is used:

$$y_L = \bar{y} - K_n s_y = 3.639 - 2.279(0.4439) = 2.627$$

$$Q_L = (10)^{2.627} = 424 \text{ cfs}$$

The 1967 peak flow of 303 cfs is less than Q_L and so is considered a low outlier.

Step 3. The low outlier is deleted from the sample and the frequency analysis is repeated using the same procedure as in Example 12.5.1. The statistics for the logarithms of the new data set, now reduced to 15 values, are $\bar{y} = 3.716$, $s_y = 0.3302$, and $C_s = -0.545$. It can be seen that the omission of the 303 cfs value has significantly altered the calculated skewness value (from the -1.24 found in Example 12.5.1). The map skewness remains at -0.3 for Austin, Texas, and the revised weighted skewness is $C_w = -0.41$. Values of K_T are interpolated from Table 12.3.1 at the required return periods, and the corresponding flood flow estimates computed as Q_T', listed in column 5 of Table 12.5.2. By comparing these values with those given in column 4 for the full data set, it can be seen that the effect of removing the low outlier in this example is to decrease the flood estimates for the longer return periods.

Computer Program HECWRC

The computer program HECWRC (U. S. Army Corps of Engineers, 1982) performs flood flow frequency analysis of annual maximum flood series according to the U. S. Water Resources Council Bulletin 17B (1981). This program is available from the U. S. Army Corps of Engineers Hydrologic Engineering Center in Davis, California, in both a mainframe computer version and a microcomputer version.

12.6 RELIABILITY OF ANALYSIS

The reliability of the results of frequency analysis depends on how well the assumed probabilistic model applies to a given set of hydrologic data.

Confidence Limits

Statistical estimates are often presented with a range, or *confidence interval*, within which the true value can reasonably be expected to lie. The size of the confidence interval depends on the *confidence level* β. The upper and lower boundary values of the confidence interval are called *confidence limits* (Fig. 12.6.1).

Corresponding to the confidence level β is a *significance level* α, given by

$$\alpha = \frac{1-\beta}{2} \tag{12.6.1}$$

For example, if β = 90 percent, then $\alpha = (1 - 0.9)/2 = 0.05$, or 5 percent.

For estimating the event magnitude for return period T, the upper limit $U_{T,\alpha}$ and lower limit $L_{T,\alpha}$ may be specified by adjustment of the frequency factor equation:

$$U_{T,\alpha} = \bar{y} + s_y K^U_{T,\alpha} \tag{12.6.2}$$

and

$$L_{T,\alpha} = \bar{y} + s_y K^L_{T,\alpha} \tag{12.6.3}$$

where $K^U_{T,\alpha}$ and $K^L_{T,\alpha}$ are the upper and lower confidence limit factors, which can be determined for normally distributed data using the noncentral t distribution (Kendall and Stuart, 1967). The same factors are used to construct approximate confidence limits for the Pearson Type III distribution. Approximate values for these factors are given by the following formulas (Natrella, 1963; U. S. Water Resources Council, 1981):

$$K^U_{T,\alpha} = \frac{K_T + \sqrt{K_T^2 - ab}}{a} \tag{12.6.4}$$

$$K^L_{T,\alpha} = \frac{K_T - \sqrt{K_T^2 - ab}}{a} \tag{12.6.5}$$

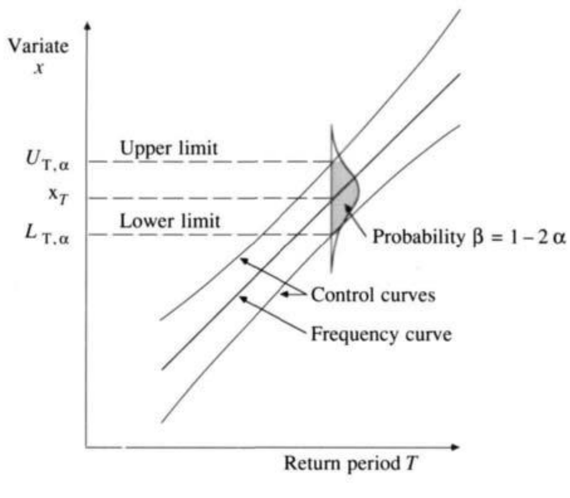

FIGURE 12.6.1
Definition of confidence limits.

in which

$$a = 1 - \frac{z_\alpha^2}{2(n-1)} \tag{12.6.6}$$

and

$$b = K_T^2 - \frac{z_\alpha^2}{n} \tag{12.6.7}$$

The quantity z_α is the standard normal variable with exceedence probability α.

Example 12.6.1. Determine the 90-percent confidence limits for the 100-year discharge for Walnut Creek, using the data presented in Example 12.5.1. The logarithmic mean, standard deviation, and skew coefficient are 3.639, 0.4439, and −0.64, respectively, for 16 years of data.

Solution. For $\beta = 0.9$, $\alpha = 0.05$ and the required standard normal variable z_α has exceedence probability 0.05, or cumulative probability 0.95. From Table 11.2.1, the required value is $z_\alpha = 1.645$. The frequency factor K_T for $T = 100$ years was calculated in Example 12.5.1 as $K_{100} = 1.850$. Hence, by Eqs. (12.6.4) to (12.6.7)

$$a = 1 - \frac{z_\alpha^2}{2(n-1)} = 1 - \frac{(1.645)^2}{2(16-1)} = 0.9098$$

$$b = K_T^2 - \frac{z_\alpha^2}{n} = (1.850)^2 - \frac{(1.645)^2}{16} = 3.253$$

$$K_{100,0.05}^U = \frac{K_T + \sqrt{K_T^2 - ab}}{a} = \frac{1.850 + [(1.850)^2 - 0.9098 \times 3.253]^{1/2}}{0.9098}$$
$$= 2.781$$

$$K_{100,0.05}^L = \frac{K_T - \sqrt{K_T^2 - ab}}{a} = \frac{1.850 - [(1.850)^2 - 0.9098 \times 3.253]^{1/2}}{0.9098}$$
$$= 1.286$$

The confidence limits are computed using Eqs. (12.6.2) and (12.6.3):

$$U_{100,0.05} = \bar{y} + s_y K_{100,0.05}^U$$
$$= 3.639 + 0.4439 \times 2.781$$
$$= 4.874$$
$$L_{100,0.05} = \bar{y} + s_y K_{100,0.05}^L$$
$$= 3.639 + 0.4439 \times 1.286$$
$$= 4.210$$

The corresponding discharges for the upper and lower limits are $(10)^{4.874} = 74,820$ cfs, and $(10)^{4.210} = 16,200$ cfs, respectively, as compared to an estimated event magnitude of 28,900 cfs from Table 12.5.2. The confidence interval is quite wide in this case because the sample size is small. As the sample size increases, the width of the confidence interval around the estimated flood magnitude will diminish.

Standard Error

The *standard error of estimate* s_e is a measure of the standard deviation of event magnitudes computed from samples about the true event magnitude. Formulas for the standard error of estimate for the normal and Extreme Value Type I distributions are (Kite, 1977):
Normal

$$s_e = \left(\frac{2 + z^2}{n}\right)^{1/2} s \tag{12.6.8}$$

Extreme Value Type I

$$s_e = \left[\frac{1}{n}(1 + 1.1396K_T + 1.1000K_T^2)\right]^{1/2} s \tag{12.6.9}$$

where s is the standard deviation of the original sample of size n. Standard errors may be used to construct confidence limits in a similar manner to that illustrated in Example 12.6.1, except that in this case the confidence limits for significance level α are defined as $x_T \pm s_e z_\alpha$.

Example 12.6.2. Determine the standard error of estimate and the 90 percent confidence limits of the 5-year-return-period, 10-minute-duration rainfall at Chicago, Illinois. From Example 12.3.2, the estimated 5-year depth is $x_T = 0.78$ in; also, $s = 0.177$ in, $K_T = 0.719$, and $n = 35$.

Solution. The standard error is computed for the Extreme Value Type I distribution using Eq. (12.6.9)

$$\begin{aligned} s_e &= \left[\frac{1}{n}(1 + 1.1396K_T + 1.1000K_T^2)\right]^{1/2} s \\ &= \left\{\frac{1}{35}\left[1 + 1.1396 \times 0.719 + 1.1000 \times (0.719)^2\right]\right\}^{1/2} \times 0.177 \\ &= 0.046 \text{ in} \end{aligned}$$

The 90 percent confidence limits, with $z_\alpha = 1.645$ for $\alpha = 0.05$, are $x_T \pm s_e z_\alpha = 0.78 \pm 0.046 \times 1.645 = 0.70$ and 0.86 in. Thus the 5 year, 10-minute rainfall estimate in Chicago is 0.78 in with 90 percent confidence limits [0.70, 0.86] in.

Expected Probability

Expected probability is defined as the average of the true exceedence probabilities of all magnitude estimates that might be made from successive samples of a specified size for a specified flood frequency (Beard, 1960; U. S. Water Resources

Council, 1981). The flood magnitude estimate computed for a given sample is approximately the median of all possible estimates; that is, there is an approximately equal chance that the true magnitude will be either above or below the estimated magnitude. But the probability distribution of the estimate is positively skewed, so the average of the magnitudes computed from many samples is larger than the median. The skewness arises because flood magnitude has a lower bound at zero but no upper bound.

The consequence of the discrepancy between the median and the mean flood estimate is that, if a very large number of estimates of flood magnitude are made over a region, on average more 100-year floods will occur than expected (Beard, 1978). The expected probability of occurence of flood events in any year can be estimated for events of nominal return period T by the following formulas, which are derived for the normal distribution, and apply approximately to the Pearson Type III distribution (Beard, 1960; Hardison and Jennings, 1972).

The expected probability for the normal distribution is expressed for a sample size of n as

$$E(P_n) = P\left[t_{n-1} > z\left(\frac{n}{n+1}\right)^{1/2}\right] \tag{12.6.10}$$

where z is the standard normal variable for the desired probability of exceedence and t_{n-1} is the student's t-statistic with $n - 1$ degrees of freedom. Calculation can be performed using the appropriate tables for t_{n-1} and z. These computations can also be carried out using the following equations (U. S. Water Resources Council, 1981; U. S. Army Corps of Engineers, 1972).

T (years)	**Exceedence probability**	**Expected probability $E(P_n)$**	
1000	0.001	$0.001\left(1.0+\dfrac{280}{n^{1.55}}\right)$	(12.6.11*a*)
100	0.01	$0.01\left(1.0+\dfrac{26}{n^{1.16}}\right)$	(12.6.11*b*)
20	0.05	$0.05\left(1.0+\dfrac{6}{n^{1.04}}\right)$	(12.6.11*c*)
10	0.10	$0.10\left(1.0+\dfrac{3}{n^{1.04}}\right)$	(12.6.11*d*)
3.33	0.30	$0.30\left(1.0+\dfrac{0.46}{n^{0.925}}\right)$	(12.6.11*e*)

Example 12.6.3. Determine the expected probability for the 100-year discharge for the Walnut Creek data given in Example 12.5.1 ($n = 16$).

Solution. For $T = 100$ years, use Eq. (12.6.11*b*) to obtain

$$E(P_n) = 0.01\left(1 + \frac{26}{n^{1.16}}\right)$$

$$= 0.01\left(1 + \frac{26}{(16)^{1.16}}\right)$$

$$= 0.020$$

The 100-year discharge according to the above adjustment has an expected probability of 0.02 (not 0.01) or a return period of 1/0.02 = 50 years.

REFERENCES

Abramowitz, M., and Stegun, I. A., *Handbook of Mathematical Functions,* Dover, New York, 1965.

Beard, L. R., Probability estimates based on small normal distribution samples, *J. Geophys. Res.,* vol. 65, no. 7, pp. 2143–2148, 1960.

Beard, L. R., Statistical methods in hydrology, U. S. Army Corps of Engineers, January 1962.

Beard, L. R., Impact of hydrologic uncertainties on flood insurance, *J. Hyd. Div., Am. Soc. Civ. Eng.,* vol. 104, no. HY11, pp. 1473–1484, 1978.

Benson, M. A., Uniform flood-frequency estimating methods for federal agencies, *Water Resour. Res.,* vol. 4, no. 5, pp. 891–908, 1968.

Blom, G., *Statistical Estimates and Transformed Beta Variables,* Wiley, New York, pp. 68–75 and 143–146, 1958.

Chow, V. T., A general formula for hydrologic frequency analysis, *Trans. Am. Geophysical Union,* vol. 32, no. 2, pp. 231–237, 1951.

Chow, V. T., Frequency analysis of hydrologic data with special application to rainfall intensities, bulletin no. 414, University of Illinois Eng. Expt. Station, 1953.

Chow, V. T., The log-probability law and its engineering applications, *Proc. Am. Soc. Civ. Eng.,* vol. 80, no. 536, pp. 1–25, September 1954.

Chow, V. T., Statistical and probability analysis of hydrologic data, sec. 8-I in *Handbook of Applied Hydrology,* ed. by V. T. Chow, McGraw-Hill, New York, 1964.

Cunnane, C., Unbiased plotting positions—a review, *J. Hydrol.,* vol. 37, pp. 205–222, 1978.

Gringorten, I. I., A plotting rule for extreme probability paper, *J. Geophys. Res.,* vol. 68, no. 3, pp. 813–814, 1963.

Gumbel, E. J., The statistical theory of droughts, *Proc. Am. Soc. Civ. Eng.,* vol. 80, pp. 439-1 to 439-19, 1954.

Gumbel, E. J., Statistical forecast of droughts, *Bull. Int. Assn. Sci. Hydrol.,* 8th year, no. 1, pp. 5–23, 1963.

Hardison, C. H., and M. E. Jennings, Bias in computed flood risk, *J. Hyd. Div., Am. Soc. Civ. Eng.,* vol. 98, no. HY3, pp. 415–427, March 1972.

Hazen, A., *Flood Flows, a Study of Frequencies and Magnitudes,* Wiley, New York, 1930.

Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics,* vol. 2, 2nd ed., Hafner, New York, 1967.

Kite, G. W., *Frequency and Risk Analysis in Hydrology,* Water Resources Publications, Fort Collins, Colo., 1977.

Natrella, M. G., *Experimental Statistics,* National Bureau of Standards Handbook 91, 1963.

Natural Environment Research Council, Flood studies report, vol. 1, hydrological studies, Natural Environment Research Council, London (available from Institute of Hydrology, Wallingford, Oxon, England), 1975.

Tomlinson, A. I., The frequency of high intensity rainfalls in New Zealand, *Water and Soil Tech. Publ.*, no. 19, Ministry of Works and Development, Wellington, New Zealand, 1980.

Tung, Y.-K., and L. W. Mays, Reducing hydrologic parameter uncertainty, *J. Water Res. Planning and Management Div., Am. Soc. Civ. Eng.*, vol. 107, no. WR1, pp. 245–262, March 1981.

U. S. Army Corps of Engineers Hydrologic Engineering Center, Regional frequency computation, generalized computer program, Davis, California, July 1972.

U. S. Army Corps of Engineers Hydrologic Engineering Center, Flood flow frequency analysis, computer program 723-X6-L7550 user's manual, Davis, California, February 1982.

U. S. Water Resources Council, A uniform technique for determining flood flow frequencies, bulletin 15, Washington, D.C., 1967.

U. S. Water Resources Council, Guidelines for determining flood flow frequency, bulletin 17, Washington, D.C., 1976.

U. S. Water Resources Council, Guidelines for determining flood flow frequency, bulletin 17A, Washington, D.C., 1977.

U. S. Water Resources Council (now called Interagency Advisory Committee on Water Data), Guidelines for determining flood flow frequency, bulletin 17B, available from Office of Water Data Coordination, U.S. Geological Survey, Reston, VA 22092, 1981.

Wallis, J. R., N. C. Matalas, and J. R. Slack, Just a moment, *Water Resour. Res.*, vol. 10, no. 2, pp. 211–219, April 1974.

PROBLEMS

12.1.1 Estimate the return period of an annual maximum discharge of 40,000 cfs from the data given in Table 12.1.1.

12.1.2 Estimate the return period of annual maximum discharges of 10,000, 20,000, 30,000, 40,000 and 50,000 cfs for the Guadalupe River at Victoria, Texas, from the data given in Table 12.1.1. Plot a graph of flood discharge vs. return period from the results.

12.1.3 Calculate the probability that a 100-year flood will occur at a given site at least once during the next 5, 10, 50, and 100 years. What is the chance that a 100-year flood will not occur at this site during the next 100 years?

12.1.4 What is the probability that a five-year flood will occur (*a*) in the next year, (*b*) at least once during the next five years, and (*c*) at least once during the next 50 years?

12.2.1 Calculate the 20-year and 100-year return period rainfall of 10 minutes duration at Chicago using the data given in Table 12.2.1. Use the Extreme Value Type I distribution.

12.3.1 (*a*) For the annual maximum series given below, determine the 25-, 50-, and 100-year peak discharges using the Extreme Value Type I distribution.

Year	1	2	3	4	5	6	7
Peak discharge (cfs)	4,780	1,520	9,260	17,600	4,300	21,200	12,000
Year	8	9	10	11	12	13	14
Peak discharge	2,840	2,120	3,170	3,490	3,920	3,310	13,200
Year	15	16	17	18	19	20	21
Peak discharge	9,700	3,380	9,540	12,200	20,400	7,960	15,000
Year	22	23	24	25	26	27	
Peak discharge	3,930	3,840	4,470	16,000	6,540	4,130	

(*b*) Determine the risk that a flow equaling or exceeding 25,000 cfs will occur at this site during the next 15 years.

(*c*) Determine the return period for a flow rate of 15,000 cfs.

12.3.2 The maximum discharges as recorded at a river gaging station are as follows:

Date of Occurrence	Discharge (cfs)	Date of Occurrence	Discharge (cfs)
1940 June 23	908	1944 Feb. 26	1610
1941 Feb. 13	1930	1944 March 13	4160
1941 March 20	3010	1945 May 14	770
1941 May 31	2670	1946 Jan. 5	5980
1941 June 3	2720	1946 Jan. 9	2410
1941 June 28	2570	1946 March 5	1650
1941 Sept. 8	1930	1947 March 13	1260
1941 Oct. 23	2270	1948 Feb. 28	4630
1942 June 3	1770	1948 March 15	2690
1942 June 10	1770	1948 March 19	4160
1942 June 11	1970	1949 Jan. 4	1680
1942 Sept. 3	1570	1949 Jan. 15	1640
1942 Dec. 27	3850	1949 Feb. 13	2310
1943 Feb. 20	2650	1949 Feb. 18	3300
1943 March 15	2450	1949 Feb. 24	3460
1943 June 2	1290	1950 Jan. 25	3050
1943 June 20	1200	1950 March 5	2880
1943 Aug. 2	1200	1950 June 2	1450
1944 Feb. 23	1490		

Select the annual maximum series from this data set. By fitting the annual maximum data to an Extreme Value Type I distribution, determine the flood flow for 10-, 50-, and 100-year return periods.

12.3.3 Select the annual exceedence series from the data set given in Prob. 12.3.2 and calculate the 10-, 50-, and 100-year discharge values from these data using the Extreme Value Type I distribution. Compare the computed values with those obtained in Prob. 12.3.2.

12.3.4 Solve Prob. 12.3.2 using the lognormal distribution.

12.3.5 Solve Prob. 12.3.2 using the log–Pearson Type III distribution.

12.3.6 The record of annual peak discharges at a stream gaging station is as follows:

Year	1961	1962	1963	1964	1965	1966	1967	1968	1969
Discharge (m^3/s)	45.3	27.5	16.9	41.1	31.2	19.9	22.7	59.0	35.4

Determine using the lognormal distribution

(*a*) The probability that an annual flood peak of 42.5 m^3/s will not be exceeded.

(*b*) The return period of a discharge of 42.5 m^3/s.

(*c*) The magnitude of a 20-year flood.

12.3.7 Show that the frequency factor for the Extreme Value Type I distribution is given by

$$K_T = -\frac{\sqrt{6}}{\pi}\left[0.5772 + \ln\left(\ln\frac{T}{T-1}\right)\right]$$

12.4.1 Plot the annual maximum discharge data from Walnut Creek given in Table 12.5.1 on a lognormal probability scale using Blom's plotting formula.

12.4.2 Solve Prob. 12.4.1 using the Weibull plotting formula and compare the results of the two plotting formulas.

12.4.3 Plot the data given in Prob. 12.3.1 on an Extreme Value Type I probability scale using the reduced variate y as the horizontal axis and discharge as the vertical axis. Use the Gringorten plotting formula.

12.4.4 Solve Prob. 12.4.3 using the Weibull plotting formula and compare the results of the two plotting formulas.

12.5.1 Perform a frequency analysis for the annual maximum discharge of Walnut Creek using the data given in Table 12.5.1, employing the log–Pearson Type III distribution without the U. S. Water Resources Council corrections for skewness and outliers. Compare your results with those given in Table 12.5.2 for the 2-, 5-, 10-, 25-, 50-, and 100-year events.

12.5.2 Using the log–Pearson Type III distribution and the hydrologic data in the following table, compute the 2-, 5-, 10-, 25-, 50-, and 100-year annual maximum floods at Leaf River, Illinois. Use the U. S. Water Resources Council method for skewness and check for outliers. The map skew for Leaf River is −0.4.

Annual maximum discharges for Leaf River, Illinois

Year	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950
Discharge (cfs)	2160	3210	3070	4000	3830	978	6090	1150	6510	3070	3360

12.5.3 Using the annual maximum flows given below for Mills Creek near Los Molinos, California, determine the 2-, 10-, 25-, 50-, and 100-year flood peaks using the log–Pearson Type III distribution with the U. S. Water Resources Council skewness adjustment. The map skewness at Los Molinas is $C_m = 0$.

Year	1929	1930	1931	1932	1933	1934	1935	1936
Discharge (cfs)	1,520	6,000	1,500	5,440	1,080	2,630	4,010	4,380
Year	1937	1938	1939	1940	1941	1942	1943	1944
Discharge	3,310	23,000	1,260	11,400	12,200	11,000	6,970	3,220
Year	1945	1946	1947	1948	1949	1950	1951	1952
Discharge	3,230	6,180	4,070	7,320	3,870	4,430	3,870	5,280
Year	1953	1954	1955	1956	1957	1958		
Discharge	7,710	4,910	2,480	9,180	6,150	6,880		

The statistics of the logarithms to base 10 of these data are: mean 3.6656, standard deviation 0.3031, coefficient of skewness −0.165.

12.5.4 The station record for Fishkill Creek at Beacon, New York, has a mean of the tranformed flows (log Q) of 3.3684, a standard deviation of transformed flows of 0.2456, and a skew coefficient of the tranformed flows of 0.7300. The station record is in cfs and is based upon 24 values.

(*a*) Determine the flood discharge for 2-, 20-, and 100-year return periods using the lognormal distribution.

(*b*) Determine the flood discharges for the same return periods using the sample skew for the log–Pearson III distribution.

(*c*) Determine the flood discharges using the procedure as recommended by the U. S. Water Resources Council. The map skew is 0.6. Compare the results obtained in parts (*a*), (*b*), and (*c*).

12.5.5 Use the U. S. Water Resources Council method to determine the 2-, 10-, 25-, 50-, and 100-year peak discharges for the station record of the San Gabriel River at Georgetown, Texas. The map skew is −0.3.

Year	1935	1936	1937	1938	1939	1940	1941	1942
Discharge (cfs)	25,100	32,400	16,300	24,800	903	34,500	30,000	18,600
Year	1943	1944	1945	1946	1947	1948	1949	1950
Discharge	7,800	37,500	10,300	8,000	21,000	14,000	6,600	5,080
Year	1951	1952	1953	1954	1955	1956	1957	1958
Discharge	5,350	11,000	14,300	24,200	12,400	5,660	155,000	21,800
Year	1959	1960	1961	1962	1963	1964	1965	1966
Discharge	3,080	71,500	22,800	4,040	858	13,800	26,700	5,480
Year	1967	1968	1969	1970	1971	1972	1973	
Discharge	1,900	21,800	20,700	11,200	9,640	4,790	18,100	

12.5.6 Solve Prob. 12.5.5 using the U. S. Army Corps of Engineers computer program HECWRC for flood flow frequency analysis with the log–Pearson III distribution.

12.5.7 Use the U. S. Water Resources Council method to determine the 2-, 10-, 25-, 50-, and 100-year peak discharges for the station record (Table 12.1.1) for the Guadalupe River at Victoria, Texas. The map coefficient of skewness is −0.3.

12.5.8 Solve Prob. 12.5.7 using the U. S. Army Corps of Engineers computer program HECWRC for flood flow frequency analysis with the log–Pearson III distribution.

12.6.1 Plot the 90-percent confidence limits of the flood flow frequency curve for the Walnut Creek data given in Table 12.5.1. Consider the 2-, 10-, 25-, 50-, and 100-year return periods.

12.6.2 Plot the 90-percent confidence limits of the flood flow frequency curve for the Los Molinos, California station record (Prob. 12.5.3). Consider the 2-, 10-, 25-, 50-, and 100-year return periods.

12.6.3 Plot the 90-percent confidence limits of the flood flow frequency curve for the San Gabriel River at Georgetown, Texas (Prob. 12.5.5). Consider the 2-, 10-, 25-, 50-, and 100-year return periods.

12.6.4 Plot the 90-percent confidence limits of the flood flow frequency curve for the Guadalupe River at Victoria, Texas (Prob. 12.5.7).

12.6.5 Determine the expected probability of a 10-year event for the Walnut Creek data (Table 12.5.1).

12.6.6 Determine the expected probability of a 10-year and a 100-year flood on the Guadalupe River at Victoria, Texas (data given in Table 12.1.1).

12.6.7 Determine the expected probability of a 10-year and a 100-year flood discharge estimated for the San Gabriel River at Georgetown, Texas (Prob. 12.5.5).

CHAPTER 13

HYDROLOGIC DESIGN

Hydrologic design is the process of assessing the impact of hydrologic events on a water resource system and choosing values for the key variables of the system so that it will perform adequately. Hydrologic design may be used to develop plans for a new structure, such as a flood control levee, or to develop management programs for better control of an existing system, for example, by producing a flood plain map for limiting construction near a river. There are many factors besides hydrology that bear on the design of water resource systems; these include public welfare and safety, economics, aesthetics, legal issues, and engineering factors such as geotechnical and structural design. While the central concern of the hydrologist is on the flow of water through a system, he or she must also be aware of these other factors and of how the hydrologic operation of the system might affect them. In this sense hydrologic design is a much broader subject than hydrologic analysis as covered in previous chapters.

13.1 HYDROLOGIC DESIGN SCALE

The purposes of water resources planning and management may be grouped roughly into two categories. One is *water control,* such as drainage, flood control, pollution abatement, insect control, sediment control, and salinity control. The other is *water use* and management, such as domestic and industrial water supply, irrigation, hydropower generation, recreation, fish and wildlife improvement, low-flow augmentation for water quality management, and watershed management. In either case, the task of the hydrologist is the same, namely, to determine a design inflow, to route the flow through the system, and to check whether the output values are satisfactory. The difference between the two cases is that design for water control is usually concerned with extreme events of short duration, such as the instantaneous peak discharge during a flood, or the minimum flow over a period of a few days during a dry period, while design for water use is concerned with the complete flow hydrograph over a period of years.

The *hydrologic design scale* is the range in magnitude of the design variable (such as the design discharge) within which a value must be selected to determine the inflow to the system (see Fig. 13.1.1). The most important factors in selecting the design value are cost and safety. It is too costly to design small structures such as culverts for very large peak discharges; however, if a major hydraulic structure, such as the spillway on a large dam, is designed for too small a flood, the result might be a catastrophe, such as a dam's failure. The optimal magnitude for design is one that balances the conflicting considerations of cost and safety.

Estimated Limiting Value

The practical upper limit of the hydrologic design scale is not infinite, since the global hydrologic cycle is a closed system; that is, the total quantity of water on earth is essentially constant. Some hydrologists recognize no upper limit, but such a view is physically unrealistic. The lower limit of the design scale is zero in most cases, since the value of the design variable cannot be negative. Although the true upper limit is usually unknown, for practical purposes an estimated upper limit may be determined. This *estimated limiting value* (ELV) is defined as *the largest magnitude possible for a hydrologic event at a given location, based on the best available hydrologic information.* The range of uncertainty for the ELV

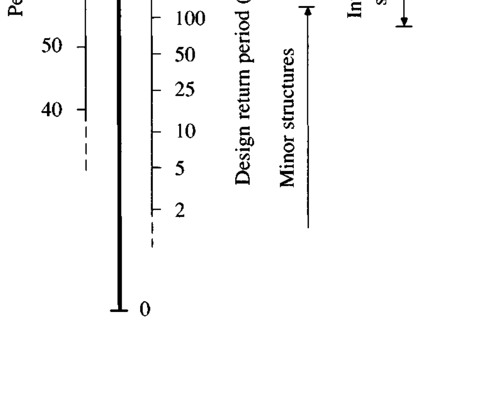

FIGURE 13.1.1
Hydrologic design scale. Approximate ranges of the design level for different types of structures are shown. Design may be based on a percentage of the ELV or on a design return period. The values for the two scales shown in the diagram are illustrative only and do not correspond directly with one another.

depends on the reliability of information, technical knowledge, and accuracy of analysis. As information, knowledge, and analysis improve, the estimate better approximates the true upper limit, and its range of uncertainty decreases. There have been cases in which observed hydrologic events exceeded their previously estimated limiting values.

The concept of an estimated limiting value is implicit in the commonly used *probable maximum precipitation* (PMP) and the corresponding *probable maximum flood* (PMF). The probable maximum precipitation is defined by the World Meteorological Organization (1983) as a "quantity of precipitation that is close to the physical upper limit for a given duration over a particular basin." Based on worldwide records, the PMP can have a return period of as long as 500,000,000 years, corresponding approximately to a frequency factor of 15. However, the return period varies geographically. Some would arbitrarily assign a return period, say 10,000 years, to the PMP or PMF, but this suggestion has no physical basis.

Probability-Based Limits

Because of its unknown probability, the estimated limiting value is used deterministically. Lower down on the design scale, a probability- or frequency-based approach is commonly adopted. The magnitudes of hydrologic events at this level are smaller, usually within or near the range of frequent observations. As a result, their probabilities of occurrence can be estimated adequately when hydrologic records of sufficient length are available for frequency analysis. The probabilistic approach is less subjective and more theoretically manageable than the deterministic approach. Probabilistic methods also lead to logical ways of determining optimum design levels, such as by hydroeconomic and risk analyses, which will be discussed in Sec. 13.2.

For a densely populated area, where the failure of water-control works would result in loss of life and extensive property damage, a design using the ELV might be justified. In a less populous area where failure would result only in minor damage, a design for a much smaller degree of protection is reasonable. Between these extremes on the hydrologic design scale, varying conditions exist and varying design values are required. When the probabilistic behavior of a hydrologic event can be determined, it is usually best to use the event magnitude for a specified return period as a design value.

Based on past experience and judgment, some generalized design criteria for water-control structures have been developed, as summarized in Table 13.1.1. According to the potential consequence of failure, structures are classified as *major*, *intermediate* and *minor*; the corresponding approximate ranges on the design scale are shown in Fig. 13.1.1. The criteria for dams in Table 13.1.1 pertain to the design of spillway capacities, and are taken from the National Academy of Sciences (1983). The Academy defines a small dam as having 50–1000 acre·ft of storage or being 25–40 ft high, an intermediate dam as having 1000–50,000 acre·ft of storage or being 40–100 ft high, and a large dam as having more than 50,000 acre·ft of storage or being more than 100 ft high. In general,

TABLE 13.1.1
Generalized design criteria for water-control structures

Type of structure	Return period (years)	ELV
Highway culverts		
Low traffic	5–10	—
Intermediate traffic	10–25	—
High traffic	50–100	—
Highway bridges		
Secondary system	10–50	—
Primary system	50–100	—
Farm drainage		
Culverts	5–50	—
Ditches	5–50	—
Urban drainage		
Storm sewers in small cities	2–25	—
Storm sewers in large cities	25–50	—
Airfields		
Low traffic	5–10	—
Intermediate traffic	10–25	—
High traffic	50–100	—
Levees		
On farms	2–50	—
Around cities	50–200	—
Dams with no likelihood of loss of life (low hazard)		
Small dams	50–100	—
Intermediate dams	100+	—
Large dams	—	50–100%
Dams with probable loss of life (significant hazard)		
Small dams	100+	50%
Intermediate dams	—	50–100%
Large dams	—	100%
Dams with high likelihood of considerable loss of life (high hazard)		
Small dams	—	50–100%
Intermediate dams	—	100%
Large dams	—	100%

there would be considerable loss of life and extensive damage if a major structure failed. In the case of an intermediate structure, a small loss of life would be possible and the damage would be within the financial capability of the owner. For minor structures, there generally would be no loss of life, and the damage would be of the same magnitude as the cost of replacing or repairing the structure.

Design for Water Use

The above discussion applies to the hydrologic design for the control of excessive waters, such as floods. Design for water use is handled similarly, except that insufficient rather than excessive water is the concern. Because of the long time span of droughts, there are fewer of them in historical hydrologic records than there are extreme floods. It is therefore more difficult to determine drought design levels through frequency analysis, especially if the design event lasts several years, as is sometimes the case in water supply design. A common basis for the design of municipal water supply systems is the *critical drought of record*, that is, the worst recorded drought. The design is considered satisfactory if it will supply water at the required rate throughout an equivalent critical period. The limitation of the critical-period approach is that the risk level associated with basing the design on this single historical event is unknown. To overcome this limitation, methods of synthetic streamflow generation have been developed using computers and random number generation to prepare synthetic streamflow records that are statistically equivalent to the historical record. Together with the historical record, the synthetic records provide a probabilistic basis for design against drought events (Hirsch, 1979; Salas, et al., 1980).

Hydrologic design for water use is closely regulated by the legal framework of water rights, especially in arid regions. The law specifies which users will have their allocations reduced in the event of a shortage. In an effort to protect the fish and wildlife of a stream, methods have been developed in recent years to quantify their need for *instream flow* (Milhous and Grenney, 1980). Unlike flood control and water supply, for which sufficient hydrologic information is provided by flow rate and water level, instream flow needs are influenced also by turbidity, temperature, and other water quality variables in a complex manner varying from one species to another. Water resources systems are subject to the demands of competing users, the need to maintain instream flow, and competing demands related to flood control. Hydrologic design must specify the appropriate design level for each of these factors.

13.2 SELECTION OF THE DESIGN LEVEL

A *hydrologic design level* on the design scale is the magnitude of the hydrologic event to be considered for the design of a structure or project. As it is not always economical to design structures and projects for the estimated limiting value, the ELV is often modified for specific design purposes. The final design value may be further modified according to engineering judgment and the experience of the designer or planner. Three approaches are commonly used to determine a hydrologic design value: an empirical approach, risk analysis, and hydroeconomic analysis.

Empirical Approach

During the early years of hydraulic engineering practice, around the early 1900s, a spillway designed to pass a flood 50 to 100 percent larger than the largest recorded in a period of perhaps 25 years was considered adequate. This design criterion is no more than a rule of thumb involving an arbitrary factor of safety. As an example of the inadequacies of this criterion, the Republican River in Nebraska in 1935 experienced a flood over 10 times as large as any that had occurred on that river during 40 prior years of record. This design practice was found to be entirely inadequate, and hydrologists and hydraulic engineers searched for better methods.

As an empirical approach the most extreme event among past observations is often selected as the design value. The probability that the most extreme event of the past N years will be equaled or exceeded once during the next n years can be estimated as

$$P(N, n) = \frac{n}{N + n} \tag{13.2.1}$$

Thus, for example, the probability that the largest flood observed in N years will be equaled or exceeded in N future years is 0.50.

If a drought lasting m years is the critical event of record over an N-year period, what is the probability $P(N, m, n)$ that a worse drought will occur within the next n years? The number of sequences of length m in N years of record is $N - m + 1$, and in n years of record $n - m + 1$. Thus the chance that the worst event over the past and future spans combined will be contained in the n future years is given approximately by

$$\begin{aligned} P(N, m, n) &= \frac{(n - m + 1)}{(N - m + 1) + (n - m + 1)} \\ &= \frac{n - m + 1}{N + n - 2m + 2} \qquad (n \geq m) \end{aligned} \tag{13.2.2}$$

which reduces to (13.2.1) when $m = 1$.

Example 13.2.1. If the critical drought of record, as determined from 40 years of hydrologic data, lasted 5 years, what is the chance that a more severe drought will occur during the next 20 years?

Solution. Using Eq. (13.2.2),

$$\begin{aligned} P(40, 5, 20) &= \frac{20 - 5 + 1}{40 + 20 - 2 \times 5 + 2} \\ &= 0.308 \end{aligned}$$

Risk Analysis

Water-control design involves consideration of risks. A water-control structure might fail if the magnitude for the design return period T is exceeded within the expected life of the structure. This *natural*, or *inherent*, hydrologic risk of failure can be calculated using Eq. (12.1.4):

$$\bar{R} = 1 - [1 - P(X \geq x_T)]^n \tag{13.2.3}$$

where $P(X \geq x_T) = 1/T$, and n is the expected life of the structure; $\bar{R}$ represents the probability that an event $x \geq x_T$ will occur at least once in n years. This

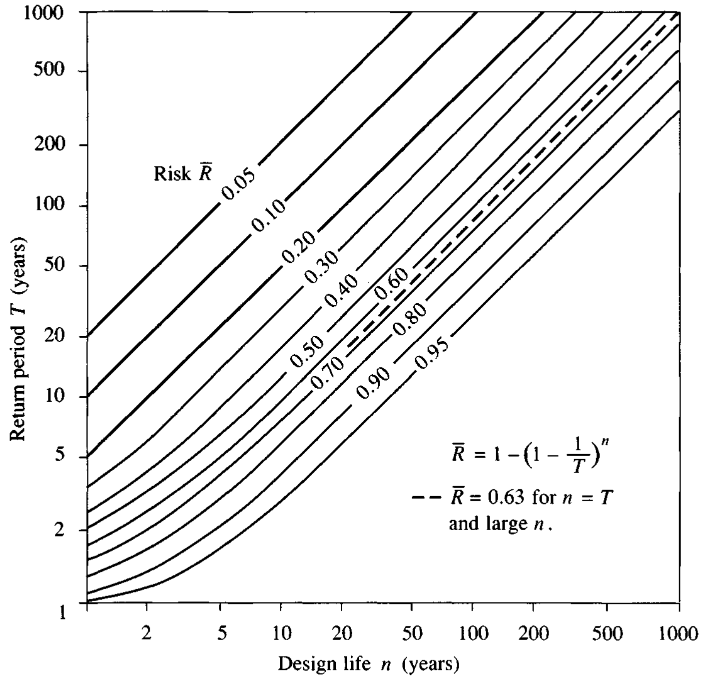

FIGURE 13.2.1
Risk of at least one exceedence of the design event during the design life.

relationship is plotted in Fig. 13.2.1. If, for example, a hydrologist wants to be approximately 90 percent certain that the design capacity of a culvert will not be exceeded during the structure's expected life of 10 years, he or she designs for the 100-year peak discharge of runoff. If a 40-percent risk of failure is acceptable, the design return period can be reduced to 20 years or the expected life extended to 50 years.

Example 13.2.2. A culvert has an expected life of 10 years. If the acceptable risk of at least one event exceeding the culvert capacity during the design life is 10 percent, what design return period should be used? What is the chance that a culvert designed for an event of this return period will not have its capacity exceeded for 50 years?

Solution. By Eq. (13.2.3)

$$\bar{R} = 1 - \left(1 - \frac{1}{T}\right)^n$$

or

$$0.10 = 1 - \left(1 - \frac{1}{T}\right)^{10}$$

and solving yields T = 95 years.

If T = 95 years, the risk of failure over n = 50 years is

$$\bar{R} = 1 - \left(1 - \frac{1}{95}\right)^{50}$$

$$= 0.41$$

So the probability that the capacity will not be exceeded during this 50-year period is $1 - 0.41 = 0.59$, or 59 percent.

It can be seen in Fig. 13.2.1 that, for a given risk of failure, the required design return period T increases linearly with the design life n, as T and n become large. Under these conditions, what is the risk of failure if the design return period is equal to the design life, that is, $T = n$? By expanding Eq. (13.2.3) as a power series, it can be shown that for large values of n, $1 - (1 - 1/T)^n \approx 1 - e^{-n/T}$, so, for $T = n$, the risk is $1 - e^{-1} = 0.632$. For example, there is approximately a 63-percent chance that a 100-year event will be exceeded at least once during the next 100 years.

Although natural hydrologic uncertainty can be accounted for as above, other kinds of uncertainty are difficult to calculate. These are often treated using a *safety factor,* SF, or a *safety margin,* SM. Letting the hydrologic design value be L and the actual capacity adopted for the project be C, the factor of safety is

$$\text{SF} = \frac{C}{L} \qquad (13.2.4)$$

and the safety margin is

$$\text{SM} = C - L \qquad (13.2.5)$$

The actual capacity is larger than the hydrologic design value because it has to allow for other kinds of uncertainty: technological (hydraulic, structural, construction, operation, etc.), socioeconomic, political, and environmental.

For a specified hydrologic risk $\bar{R}$ and design life n of a structure, Eq. (13.2.3) can be used to compute the relevant return period T. The hydrologic event magnitude L corresponding to this exceedence probability is found by a frequency analysis of hydrologic data. The design value C is then given by L multiplied by an assigned factor of safety, or by L plus an added margin of safety. For example, it is customary to design levees with a safety margin of one to three feet, that is, one to three feet of freeboard above the calculated maximum water surface elevation.

Hydroeconomic Analysis

The optimum design return period can be determined by hydroeconomic analysis if the probabilistic nature of a hydrologic event and the damage that will result if it occurs are both known over the feasible range of hydrologic events. As the design return period increases, the capital cost of a structure increases, but the expected damages decrease because of the better protection afforded. By summing the capital cost and the expected damage cost on an annual basis, a design return period having minimum total cost can be found.

Figure 13.2.2(*a*) shows the damage that would result if an event, such as a flood, having the specified return period were to occur. If the design event magnitude is x_T, the structure will prevent all damages for events with $x \leq x_T$ but none for $x > x_T$, so the *expected annual damage cost* is found by taking the product of the probability $f(x)\,dx$ that an event of magnitude x will occur in any given year, and the damage $D(x)$ that would result from that event, and

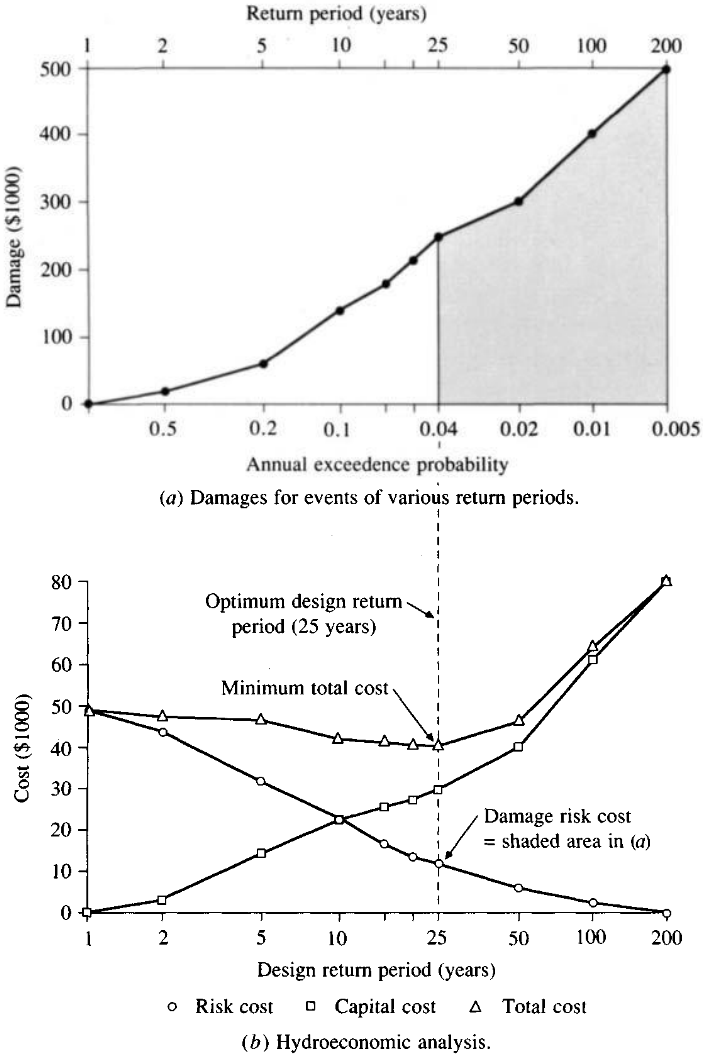

FIGURE 13.2.2
Determination of the optimum design return period by hydroeconomic analysis (Example 13.2.3).

integrating for $x > x_T$ (the design level). That is, the expected annual cost D_T is

$$D_T = \int_{x_T}^{\infty} D(x)f(x)\,dx \qquad (13.2.6)$$

which is the shaded area in Fig. 13.2.2(*a*).

The integral (13.2.6) is evaluated by breaking the range of $x > x_T$ into intervals and computing the expected annual damage cost for events in each interval. For $x_{i-1} \le x \le x_i$,

$$\Delta D_i = \int_{x_{i-1}}^{x_i} D(x)f(x)\,dx \qquad (13.2.7)$$

which is approximated by

$$\begin{aligned}\Delta D_i &= \left[\frac{D(x_{i-1}) + D(x_i)}{2}\right]\int_{x_{i-1}}^{x_i} f(x)\,dx \\ &= \frac{D(x_{i-1}) + D(x_i)}{2}\big[P(x \le x_i) - P(x \le x_{i-1})\big]\end{aligned} \qquad (13.2.8)$$

But $P(x \le x_i) - P(x \le x_{i-1}) = [1 - P(x \ge x_i)] - [1 - P(x \ge x_{i-1})] = P(x \ge x_{i-1}) - P(x \ge x_i)$, so (13.2.8) can be written

$$\Delta D_i = \frac{D(x_{i-1}) + D(x_i)}{2}\big[P(x \ge x_{i-1}) - P(x \ge x_i)\big] \qquad (13.2.9)$$

and the annual expected damage cost for a structure designed for return period T is given by

$$D_T = \sum_{i=1}^{\infty}\left[\frac{D(x_{i-1}) + D(x_i)}{2}\right]\big[P(x \ge x_{i-1}) - P(x \ge x_i)\big] \qquad (13.2.10)$$

By adding D_T to the annualized capital cost of the structure, the total cost can be found; the optimum design return period is the one having the minimum total cost.

Example 13.2.3. For events of various return periods at a given location, the damage costs and the annualized capital costs of structures designed to control the events, are shown in columns 4 and 7, respectively, of Table 13.2.1. Determine the expected annual damages if no structure is provided, and calculate the optimal design return period.

Solution. For each return period shown in column 2 of Table 13.2.1, the annual exceedence probability is $P(x \ge x_T) = 1/T$. The corresponding damage cost ΔD is found using Eq. (13.2.9). For example, for the interval $i = 1$ between $T = 1$ year and $T = 2$ years,

$$\Delta D_1 = \left[\frac{D(x_1) + D(x_2)}{2}\right]\big[P(x \ge x_1) - P(x \ge x_2)\big]$$

TABLE 13.2.1
Calculation of the optimum design return period by hydroeconomic analysis (Example 13.2.3)

Column: 1 Incre-ment i	2 Return period T (years)	3 Annual exceedence probability	4 Damage ($)	5 Incremental expected damage ($/year)	6 Damage risk cost ($/year)	7 Capital cost ($/year)	8 Total cost ($/year)
	1	1.000	0		49,098	0	49,098
1	2	0.500	20,000	5,000	44,098	3,000	47,098
2	5	0.200	60,000	12,000	32,098	14,000	46,098
3	10	0.100	140,000	10,000	22,098	23,000	45,098
4	15	0.067	177,000	5,283	16,815	25,000	41,815
5	20	0.050	213,000	3,250	13,565	27,000	40,565
6	25	0.040	250,000	2,315	11,250	29,000	*40,250*
7	50	0.020	300,000	5,500	5,750	40,000	45,750
8	100	0.010	400,000	3,500	2,250	60,000	62,250
9	200	0.005	500,000	2,250	0	80,000	80,000
		Annual expected damage = $49,098					

$$= \left(\frac{0 + 20,000}{2}\right)(1.0 - 0.5)$$

$$= \$5,000/\text{year}$$

as shown in column 5 of the table. Summing these incremental costs yields an annual expected damage cost of $49,098/year if no structure is built. This represents the average annual cost of flood damage over many years, assuming constant economic conditions. This amount is the damage risk cost corresponding to no structure, and is shown in the first line of column 6 of the table.

The damage risk costs diminish as the design return period of the control structure increases. For example, if $T = 2$ years were selected, the damage risk cost would be $49,098 - \Delta D_1 = 49,098 - 5,000 =$ $44,098/year. The values of damage risk cost and capital cost (column 7) are added to form the total cost (column 8); the three costs are plotted in Fig. 13.2.2(*b*). It can be seen from the table and the figure that the optimum design return period, the one having minimal total cost, is 25 years, for which the total cost is $40,250/year. Of this amount, $29,000/year (72 percent) is capital cost and $11,250/year (28 percent) is damage risk cost.

Hydroeconomic analysis has been applied to the design of flood control reservoirs, levees, channels, and highway stream crossings (Corry, Jones, and Thompson, 1980). For a flood damage study, the duration and extent of flooding must be determined for events of various return periods and economic surveys must be taken to quantify damages for each level of flooding. The social costs of flooding are difficult to quantify. The U. S. Army Corps of Engineers Hydrologic Engineering Center in Davis, California, has available the following computer programs for hydroeconomic analysis (U. S. Army Corps of Engineers, 1986):

DAMCAL (Damage Reach Stage–Damage Calculation), EAD (Expected Annual Flood Damage Computation), SID (Structure Inventory for Damage Analysis), AGDAM (Agricultural Flood Damage Analysis), and SIPP (Interactive Nonstructural Analysis Package).

13.3 FIRST ORDER ANALYSIS OF UNCERTAINTY

Many of the uncertainties associated with hydrologic systems are not quantifiable. For example, the conveyance capacity of a culvert with an unobstructed entrance can be calculated within a small margin of error, but during a flood, debris may become lodged around the entrance to the culvert, reducing its conveyance capacity by an amount that cannot be predetermined. Hydrologic uncertainty may be broken down into three categories: *natural*, or *inherent*, *uncertainty*, which arises from the random variability of hydrologic phenomena; *model uncertainty*, which results from the approximations made when representing phenomena by equations; and *parameter uncertainty*, which stems from the unknown nature of the coefficients in the equations, such as the bed roughness in Manning's equation. Inherent uncertainty in the magnitude of the design event is described by Eq. (13.2.3); in this section, model and parameter uncertainty will be considered.

The *first order analysis of uncertainty* is a procedure for quantifying the expected variability of a dependent variable calculated as a function of one or more independent variables (Ang and Tang, 1975; Kapur and Lamberson, 1977; Ang and Tang, 1984; Yen, 1986). Suppose w is expressed as a function of x:

$$w = f(x) \tag{13.3.1}$$

There are two sources of error in w: first, the function f, or model, may be incorrect; second, the measurement of x may be inaccurate. In the following analysis it is assumed that there is no model error, or *bias*. Kapur and Lamberson (1977) show how to extend the analysis when there is model error. Assuming, then, that $f(\cdot)$ is a correct model, a nominal value of x, denoted $\bar{x}$, is selected as a design input and the corresponding value of w calculated:

$$\bar{w} = f(\bar{x}) \tag{13.3.2}$$

If the true value of x differs from $\bar{x}$, the effect of this discrepancy on w can be estimated by expanding $f(x)$ as a Taylor series around $x = \bar{x}$:

$$w = f(\bar{x}) + \frac{df}{dx}(x - \bar{x}) + \frac{1}{2!}\frac{d^2f}{dx^2}(x - \bar{x})^2 + \dots \tag{13.3.3}$$

where the derivatives df/dx, d^2f/dx^2, . . . , are evaluated at $x = \bar{x}$. If second and higher order terms are neglected, the resulting *first order* expression for the error in w is

$$w - \bar{w} = \frac{df}{dx}(x - \bar{x}) \tag{13.3.4}$$

The variance of this error is $s_w^2 = E[(w - \bar{w})^2]$ where E is the expectation operator [see Eq. (11.3.3)]; that is,

$$s_w^2 = E\left\{\left[\frac{df}{dx}(x - \bar{x})\right]^2\right\}$$

or

$$s_w^2 = \left(\frac{df}{dx}\right)^2 s_x^2 \tag{13.3.5}$$

where s_x^2 is the variance of x.

Equation (13.3.5) gives the variance of a dependent variable w as a function of the variance of an independent variable x, assuming that the functional relationship $w = f(x)$ is correct. The value s_w is the *standard error of estimate* of w.

If w is dependent on several *mutually independent* variables $x_1, x_2, \ldots, x_n$, it can be shown by a procedure similar to the above that

$$s_w^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 s_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 s_{x_2}^2 + \ldots + \left(\frac{\partial f}{\partial x_n}\right)^2 s_{x_n}^2 \tag{13.3.6}$$

Kapur and Lamberson (1977) show how to extend (13.3.6) to account for the effect on s_w^2 of correlation between $x_1, x_2, \ldots, x_n$, if any exists.

First-Order Analysis of Manning's Equation: Depth as the Dependent Variable

Manning's equation is widely applied in hydrology to determine depths of flow for specified flow rates, or to determine discharges for specified depths of flow, taking into account the resistance to flow in channels arising from bed roughness. A common application, such as in channel design or flood plain delineation, is to calculate the depth of flow y in the channel, given the flow rate Q, roughness coefficient n, and the shape and slope of the channel as determined by design or by surveys. Once the depth of flow (or elevation of the water surface) is known, the values of the design variables are determined, such as the channel wall elevation or the flood plain extent. The hydrologist faced with this task is conscious of the uncertainties involved, especially in the selection of the design flow and Manning roughness. Although it is not so obvious, there is also uncertainty in the value of the friction slope S_f, depending on how it is calculated, ranging from the simplest case of uniform flow ($S_o = S_f$) to more complex cases of steady nonuniform flow or unsteady nonuniform flow [see Eq. (9.2.1)]. The first-order analysis of uncertainty can be used to estimate the effect on y of uncertainty in Q, n, and S_f.

Consider, first, the effect on flow depth of variation in the flow rate Q. Manning's equation is written in English units as

$$Q = \frac{1.49}{n} S_f^{1/2} A R^{2/3} \tag{13.3.7}$$

where A is the cross-sectional area and R the hydraulic radius, both dependent on the flow depth y. If variations in y are dependent only on variations in Q, then, by (13.3.5),

$$s_y^2 = \left(\frac{dy}{dQ}\right)^2 s_Q^2 \tag{13.3.8}$$

where dy/dQ is the rate at which the depth changes with changes in Q. Now, in Chap. 5, it was shown [Eq. (5.6.15)] that the inverse of this derivative, namely dQ/dy, is given for Manning's equation by

$$\frac{dQ}{dy} = Q\left[\frac{2}{3R}\frac{dR}{dy} + \frac{1}{A}\frac{dA}{dy}\right] \tag{13.3.9}$$

Table 5.6.1 gives formulas for the channel shape function $(2/3R)(dR/dy) + (1/A)(dA/dy)$ for common channel cross sections. Substituting into (13.3.8),

$$s_y^2 = \frac{s_Q^2}{Q^2\left(\dfrac{2}{3R}\dfrac{dR}{dy} + \dfrac{1}{A}\dfrac{dA}{dy}\right)^2} \tag{13.3.10}$$

But $s_Q/Q = \mathrm{CV}_Q$, the coefficient of variation of the flow rate (see Table 11.3.1), so (13.3.10) can be rewritten

$$s_y^2 = \frac{\mathrm{CV}_Q^2}{\left(\dfrac{2}{3R}\dfrac{dR}{dy} + \dfrac{1}{A}\dfrac{dA}{dy}\right)^2} \tag{13.3.11}$$

which specifies the variance of the flow depth as a function of the coefficient of variation of the flow rate and the value of the channel shape function. To take into account also the uncertainty in Manning's roughness n and the friction slope S_f, it may be similarly shown, using Eq. (13.3.6), that

$$s_y^2 = \frac{\mathrm{CV}_Q^2 + \mathrm{CV}_n^2 + (1/4)\mathrm{CV}_{S_f}^2}{\left(\dfrac{2}{3R}\dfrac{dR}{dy} + \dfrac{1}{A}\dfrac{dA}{dy}\right)^2} \tag{13.3.12}$$

giving the variance of the flow depth y as a function of the coefficients of variation of flow rate, Manning's n and friction slope, and the channel shape function.

Example 13.3.1. A 50-foot wide rectangular channel has a bed slope of one percent. A hydrologist estimates that the design flow rate is 5000 cfs and that the roughness is $n = 0.035$. If the coefficients of variation of the flow estimate and the roughness estimate are 30 percent and 15 percent, respectively, what is the standard error of estimate of the flow depth y? If houses are built next to this channel with floor elevation one foot above the water surface elevation calculated for the design event, estimate the chance that these houses will be flooded during the design event due to uncertainties involved in calculating the water level. Assume uniform flow.

Solution. For a width of 50 feet, $A = 50y$ and $R = 50y/(50 + 2y)$; the flow depth for the base case is calculated from Manning's equation:

$$Q = \frac{1.49}{n} S_f^{1/2} A R^{2/3}$$

$$5000 = \frac{1.49}{0.035}(0.01)^{1/2}(50y)\left(\frac{50y}{50 + 2y}\right)^{2/3}$$

which is solved using Newton's iteration technique (see Sec. 5.6) to yield

$$y = 7.37 \text{ ft}$$

The standard error of the estimate is s_y, calculated by Eq. (13.3.12) with $CV_Q = 0.30$, $CV_n = 0.15$, and $CV_{S_f} = 0$. From Table 5.6.1, for a rectangular channel,

$$\left(\frac{2}{3R}\frac{dR}{dy} + \frac{1}{A}\frac{dA}{dy}\right) = \frac{5B + 6y}{3y(B + 2y)}$$

$$= \frac{5 \times 50 + 6 \times 7.37}{3 \times 7.37(50 + 2 \times 7.37)}$$

$$= 0.206$$

So

$$s_y^2 = \frac{CV_Q^2 + CV_n^2 + (1/4)CV_{S_f}^2}{\left(\frac{2}{3R}\frac{dR}{dy} + \frac{1}{A}\frac{dA}{dy}\right)^2}$$

$$= \frac{(0.30)^2 + (0.15)^2}{(0.206)^2}$$

or $s_y = 1.63$ ft.

If the houses are built with their floors one foot above the calculated water surface elevation, they will be flooded if the actual depth is greater than 7.37 + 1.00 = 8.37 ft. If the water surface elevation y is normally distributed, then the probability that they will be flooded is evaluated by converting y to the standard normal variable z by subtracting the mean value of y (7.37 ft) from both sides of the inequality and dividing by the standard error (1.63 ft):

$$P(y > 8.37) = P\left(\frac{y - 7.37}{1.63} > \frac{8.37 - 7.37}{1.63}\right)$$

$$= P\left(\frac{y - 7.37}{1.63} > 0.613\right)$$

$$= P(z > 0.613)$$

$$= 1 - F_z(0.613)$$

where F_z is the standard normal distribution function. Using Table 11.2.1 or the method employed in Example 11.2.1, the result is $F_z(0.613) = 0.73$, so $P(y > 8.37) = 1 - 0.73 = 0.27$. There is approximately a 27 percent chance that the houses will be flooded during the design event due to uncertainties in calculating the water level for that event.

This example has treated only parameter uncertainty in the calculations. The true probability that the houses will be flooded is greater than that calculated here, because the critical flood may exceed the design magnitude (due to natural uncertainty).

It is clear from Example 13.3.1 that reasonable amounts of uncertainty in the estimation of Q and n can produce significant uncertainty in flow depth. A 15-percent error in estimating $n = 0.035$ is an error of $0.035 \times 0.15 = 0.005$. This would be indicated from a measurement of 0.035 ± 0.005, which is about as accurate as an experienced hydrologist can get from observation of an existing channel. A 30-percent error in estimating Q is $5000 \times 0.30 = 1500$ cfs. An estimate of $Q = 5000 \pm 1500$ cfs may also reflect the correct order of uncertainty, especially if the design return period is large (e.g., $T = 100$ years).

The use of the channel shape function $(2/3R)(dR/dy) + (1/A)(dA/dy)$ in (13.3.12) depends on knowledge of dR/dy and dA/dy, which may be difficult to obtain for irregularly shaped channels. Also, the assumption that y depends on Q alone may not be valid. In such cases, Eq. (13.3.6) can be used to obtain s_y, treating y as a function of Q and n, and a computer program simulating flow in the channel can be used to estimate the required partial derivatives $\partial y/\partial Q$ and $\partial y/\partial n$ by rerunning the program for various values of Q and n and reading off the computed values of flow depth or water surface elevation. Figure 13.3.1 shows the results of such a procedure for the channel and conditions given in Example 13.3.1. The gradients $\partial y/\partial Q$ and $\partial y/\partial n$ are approximately linear for this example; this validates the use of only first-order terms in the analysis of uncertainty (if the lines were significantly curved, analysis would require keeping the second-order terms in the Taylor-series expansion).

Example 13.3.2. For the same conditions as in Example 13.3.1 ($B = 50$ ft, $Q = 5000$ cfs, $S_o = 0.01$, $n = 0.035$), the variation of flow rate with flow depth at the base case level has been found from Fig. (13.3.1) to be $\partial Q/\partial y = 1028$ cfs/ft, and the variation of n with flow depth, $\partial n/\partial y = 0.0072$ ft^{-1}. If $CV_Q = 0.30$ and $CV_n = 0.15$, calculate the standard error of y.

Solution. From Eq. (13.3.6),

$$s_y^2 = \left(\frac{\partial y}{\partial Q}\right)^2 s_Q^2 + \left(\frac{\partial y}{\partial n}\right)^2 s_n^2$$

In this case, $s_Q = 5000 \times 0.30 = 1500$, $s_n = 0.035 \times 0.15 = 0.0053$; also, $\partial y/\partial Q = 1/1028$, $\partial y/\partial n = 1/0.0072$. Thus,

$$s_y^2 = \left(\frac{1}{1028}\right)^2 \times (1500)^2 + \left(\frac{1}{0.0072}\right)^2 \times (0.0053)^2$$

or $s_y = 1.63$ ft as computed in Example 13.3.1.

First-Order Analysis of Manning's Equation: Discharge as the Dependent Variable

Another application of Manning's equation is the calculation of the discharge or capacity C of a stream channel or other conveyance structure for a given depth, roughness coefficient n, bottom slope, and cross-sectional geometry. Manning's equation (13.3.7) can be expressed using $R = A/P$ as

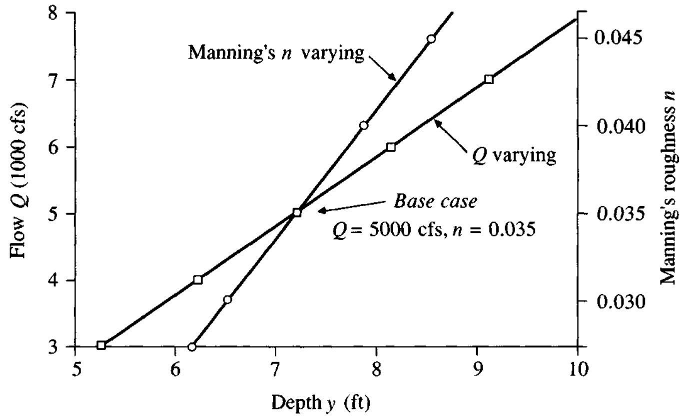

FIGURE 13.3.1
Variation of the flow depth with flow rate and with Manning's n. Rectangular channel with width 50 ft, bed slope 0.01. Uniform flow assumed. (Example 13.3.2).

$$C = Q = \frac{1.49}{n} S_f^{1/2} A^{5/3} P^{-2/3} \tag{13.3.13}$$

in which P is the wetted perimeter. Performing first-order analysis on (13.3.13), the coefficient of variation of the capacity can be expressed as

$$\mathrm{CV}_Q^2 = \mathrm{CV}_n^2 + \frac{1}{4}\mathrm{CV}_{S_f}^2 \tag{13.3.14}$$

assuming $\mathrm{CV}_A \approx 0$ and $\mathrm{CV}_P \approx 0$.

Manning's equation for a channel and flood plain (overbank) can also be expressed as (Chow, 1959)

$$Q = 1.49\left(\frac{1}{n_c} A_c^{5/3} P_c^{-2/3} + \frac{2}{n_b} A_b^{5/3} P_b^{-2/3}\right) S_f^{1/2} \tag{13.3.15}$$

in which n_c and n_b are the roughness coefficients for the channel and the flood-plain, respectively and A_c, P_c, A_b, and P_b are the cross-sectional areas and the wetted perimeters of the channel and the overbank flow. Equation (13.3.15) assumes that the cross-sectional shape of the channel and the flood plain are both symmetrical about the channel center line. This equation can be used to evaluate levee capacity (the flow rate the levee can carry without overtopping). The levee capacity can be considered a random variable related to the independent random variables n_c, n_b, and S_f. Applying first-order analysis, the coefficient of variation of the capacity is (Lee and Mays, 1986)

$$\mathrm{CV}_Q^2 = \frac{1}{4}\mathrm{CV}_{S_f}^2 + \frac{1}{\Psi^2}\mathrm{CV}_{n_c}^2 + \left(\frac{\Psi - 1}{\Psi}\right)^2 \mathrm{CV}_{n_b}^2 \tag{13.3.16}$$

where CV_{A_c}, CV_{P_c}, CV_{A_b}, and CV_{P_b} have been assumed negligible, and

$$\Psi = 1 + 2\left(\frac{n_c}{n_b}\right)\left(\frac{A_b}{A_c}\right)^{5/3}\left(\frac{P_c}{P_b}\right)^{2/3} \tag{13.3.17}$$

In studies of flood data on the Ohio River, Lee and Mays (1986) concluded that uncertainties in the roughness coefficients and the friction slope account for 95 percent of the uncertainties in computing the capacity. They presented a method for determining the uncertainty in the friction slope using the observed flood hydrograph of the river.

13.4 COMPOSITE RISK ANALYSIS

The previous sections have introduced the concepts of inherent uncertainty due to the natural variability of hydrologic phenomena, and model and parameter uncertainty arising from the way the phenomena are analyzed. Composite risk analysis is a method of accounting for the risks resulting from the various sources of uncertainty to produce an overall risk assessment for a particular design. The concepts of loading and capacity are central to this analysis.

The *loading*, or *demand*, placed on a system is the measure of the impact of external events. The demand for water supply is determined by the people who use the water. The magnitude of a flash flood depends on the characteristics of the storm producing it and on the condition of the watershed at the time of the storm. The *capacity*, or *resistance*, is the measure of the ability of the system to withstand the loading or meet the demand.

If loading is denoted by L and capacity by C, then the risk of failure $\bar{R}$ is given by the probability that L exceeds C, or

$$\begin{aligned} \bar{R} &= P\left(\frac{C}{L} < 1\right) \\ &= P(C - L < 0) \end{aligned} \tag{13.4.1}$$

The risk depends upon the probability distributions of L and C. Suppose that the probability density function of L is $f(L)$. This function could be, for example, an Extreme Value or log–Pearson Type III probability density function for extreme values, as described earlier. Given $f(L)$, the chance that the loading will exceed a fixed and known capacity C^* is (see Fig. 13.4.1)

$$P(L > C^*) = \int_{C^*}^{\infty} f(L)\,dL \tag{13.4.2}$$

The true capacity is not known exactly, but may be considered to have probability density function $g(C)$, which could be the normal or lognormal distribution arising from the first-order analysis of uncertainty in the system capacity. For example, if Manning's equation has been used to determine the capacity of a hydraulic structure, the uncertainty in C can be evaluated by first-order analysis as described above. The probability that the capacity lies within a small range

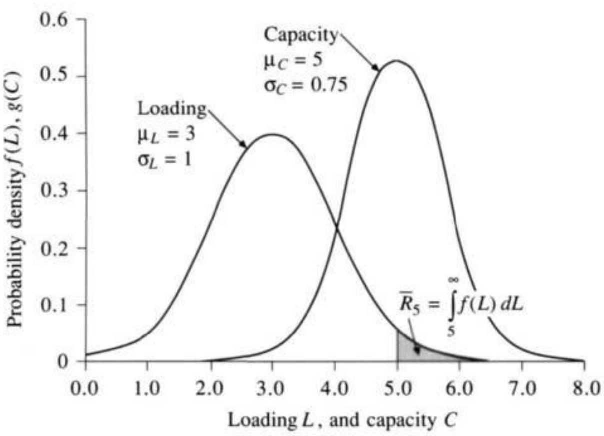

FIGURE 13.4.1
Composite risk analysis. Area shaded is the risk $\overline{R}_5$ of the loading exceeding a fixed capacity of 5 units. The risk that the loading will exceed the capacity when the capacity is random is given by $\overline{R} = \int_{-\infty}^{\infty} [\int_C^{\infty} f(L)\,dL]\, g(C)\,dC$. The loading and capacity shown are both normally distributed (Example 13.4.1).

dC around a value C is $g(C)\,dC$. Assuming that L and C are independent random variables, the *composite risk* is evaluated by calculating the probability that loading will exceed capacity at each value in the range of feasible capacities, and integrating to obtain

$$\overline{R} = \int_{-\infty}^{\infty} \left[\int_{C}^{\infty} f(L)\,dL \right] g(C)\,dC \tag{13.4.3}$$

The *reliability* of a system is defined to be *the probability that a system will perform its required function for a specified period of time under stated conditions* (Harr, 1987). Reliability R is the complement of risk, or the probability that the loading will not exceed the capacity:

$$\begin{aligned} R &= P(L \le C) \\ &= 1 - \overline{R} \end{aligned} \tag{13.4.4}$$

or

$$R = \int_{-\infty}^{\infty} \left[\int_{0}^{C} f(L)\,dL \right] g(C)\,dC \tag{13.4.5}$$

Example 13.4.1. During the coming year, a city's estimated water demand is three units, with a standard deviation of one unit. Calculate (*a*) the risk of demand exceeding supply if the city's water supply system has an estimated capacity of 5 units; (*b*) the risk of failure if the estimate of the capacity has a standard error of 0.75 units. Assume that loading and capacity are both normally distributed.

Solution. (*a*) The loading is normally distributed with $\mu_L = 3$ and $\sigma_L = 1$. Its probability function, from Eq. (11.2.5), is

$$f(L) = \frac{1}{\sqrt{2\pi}\sigma_L} e^{-(L-\mu_L)^2/2\sigma_L^2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-(L-3)^2/2}$$

The risk $\bar{R}$ is evaluated using (13.4.2) with $C^* = 5$:

$$\bar{R} = \int_{C^*}^{\infty} f(L)\,dL$$

$$= \int_{5}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(L-3)^2/2} dL$$

or

$$\bar{R} = 1 - \int_{-\infty}^{5} \frac{1}{\sqrt{2\pi}} e^{-(L-3)^2/2} dL$$

The integral is evalutated by converting the variable of integration to the standard normal variable: $u = (L - \mu_L)/\sigma_L = (L - 3)/1 = L - 3$, so $dL = du$, and $L = 5$ becomes $u = 5 - 3 = 2$; $L = -\infty$ becomes $u = -\infty$, and then

$$\bar{R} = 1 - \int_{-\infty}^{2} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$= 1 - F_z(2)$$

where F_z is the standard normal distribution function. From Table 11.2.1, $F_z(2) = 0.977$, and

$$\bar{R} = 1 - 0.977$$

$$= 0.023$$

The chance that demand will exceed supply for a fixed capacity of 5 is approximately 2 percent.

(*b*) The capacity now has a normal distribution with $\mu_C = 5$ and $\sigma_C = 0.75$. Hence, its probability density is

$$g(C) = \frac{1}{\sqrt{2\pi}\sigma_C} e^{-(C-\mu_C)^2/2\sigma_C^2}$$

$$= \frac{1}{\sqrt{2\pi}(0.75)} e^{-(C-5)^2/2\times(0.75)^2}$$

$$= \frac{1.333}{\sqrt{2\pi}} e^{-(C-5)^2/1.125}$$

and the risk of failure is given by Eq. (13.4.3), with $f(L)$ as before:

$$\bar{R} = \int_{-\infty}^{\infty} \left[\int_{C}^{\infty} f(L)\,dL \right] g(C)\,dC$$

$$= \int_{-\infty}^{\infty} \left[\int_{C}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(L-3)^2/2} dL \right] \frac{1.333}{\sqrt{2\pi}} e^{-(C-5)^2/1.125} dC$$

The integral is evaluated by computer using numerical integration to yield $\bar{R} = 0.052$. Thus, the chance that the city's water demand will exceed its supply during the coming year, assuming the capacity to be normally distributed with mean 5 and standard deviation 0.75 is approximately 5 percent; compare this with the result of 2 percent when the capacity was considered fixed at 5 units.

It is clear from Example 13.4.1 that calculation of the composite risk of failure can be a complicated exercise requiring the use of a computer to perform the necessary integration. This is especially true when more realistic distributions for the loading and capacity are chosen, such as the Extreme Value or log–Pearson Type III distributions for loading, and the lognormal distribution for capacity. Yen and co-workers at the University of Illinois (Yen, 1970; Tang and Yen, 1972; Yen, et al., 1976) and Mays and co-workers at the University of Texas at Austin (Tung and Mays, 1980; Lee and Mays, 1986) have made detailed risk analysis studies for various kinds of open-channel and pipe-flow design problems.

The composite risk analysis described here is a *static* analysis, which means that it estimates the risk of failure under the single worst case loading on the system during its design life. A more complex *dynamic* risk analysis considers the possibility of a number of extreme loadings during the design life, any one of which could cause a failure; the total risk of failure includes the chance of multiple failures during the design life (Tung and Mays, 1980; Lee and Mays 1983).

13.5 RISK ANALYSIS OF SAFETY MARGINS AND SAFETY FACTORS

Safety Margin

The safety margin was defined in Eq. (13.2.5) as the difference between the project capacity and the value calculated for the design loading $SM = C - L$. From (13.4.1), the risk of failure $\bar{R}$ is

$$\begin{aligned}\bar{R} &= P(C - L < 0) \\ &= P(SM < 0)\end{aligned} \tag{13.5.1}$$

If C and L are independent random variables, then the mean value of SM is given by

$$\mu_{SM} = \mu_C - \mu_L \tag{13.5.2}$$

and its variance by

$$\sigma_{SM}^2 = \sigma_C^2 + \sigma_L^2 \tag{13.5.3}$$

so the standard deviation, or standard error of estimate, of the safety margin is

$$\sigma_{SM} = \left(\sigma_C^2 + \sigma_L^2\right)^{1/2} \tag{13.5.4}$$

If the safety margin is normally distributed, then $(SM - \mu_{SM})/\sigma_{SM}$ is a standard normal variate z. By subtracting μ_{SM} from both sides of the inequality in (13.5.1) and dividing both sides by σ_{SM}, it can be seen that

$$\begin{aligned} \bar{R} &= P\left(\frac{SM - \mu_{SM}}{\sigma_{SM}} < \frac{-\mu_{SM}}{\sigma_{SM}}\right) \\ &= P\left(z < -\frac{\mu_{SM}}{\sigma_{SM}}\right) \qquad (13.5.5) \\ &= F_z\left(-\frac{\mu_{SM}}{\sigma_{SM}}\right) \end{aligned}$$

where F_z is the standard normal distribution function.

Example 13.5.1. Calculate the risk of failure of the water supply system in Example 13.4.1, assuming that the safety margin is normally distributed, and that $\mu_C = 5$ units, $\sigma_C = 0.75$ units, $\mu_L = 3$ units, and $\sigma_L = 1$ unit.

Solution. From Eq. (13.5.2), $\mu_{SM} = \mu_C - \mu_L = 5 - 3 = 2$. From (13.5.4), $\sigma_{SM} = (\sigma_C^2 + \sigma_L^2)^{1/2} = (1^2 + 0.75^2)^{1/2} = 1.250$. Using (13.5.5),

$$\begin{aligned} \bar{R} &= F_z\left(-\frac{\mu_{SM}}{\sigma_{SM}}\right) \\ &= F_z\left(-\frac{2}{1.250}\right) \\ &= F_z(-1.60) \end{aligned}$$

which is evaluated using Table 11.2.1 to yield $\bar{R} = 0.055$, which is very close to the value obtained in Example 13.4.1 by numerical integration (an inherently approximate procedure). The risk of failure under the stated conditions is $\bar{R} = 0.055$, or 5.5%.

Note that this method of analysis assumes that the safety margin is normally distributed but does not specify what the distributions of loading and capacity must be. Ang (1973) indicates that, provided $\bar{R} > 0.001$, $\bar{R}$ is not greatly influenced by the choice of distributions for L and C, and the assumption of a normal distribution for SM is satisfactory. For lower risk than this (e.g., $\bar{R} = 0.00001$), the shapes of the tails of the distributions for L and C become critical, and in this case, the full composite risk analysis described in Sec. 13.4 should be used to evaluate the risk of failure.

Safety Factor

The safety factor SF is given by the ratio C/L and the risk of failure can be expressed as $P(SF < 1)$. By taking logarithms of both sides of this inequality

$$\begin{aligned}\bar{R} &= P(\text{SF} < 1)\\ &= P\big(\ln(\text{SF}) < 0\big)\\ &= P\left(\ln\frac{C}{L} < 0\right)\end{aligned} \tag{13.5.6}$$

If the capacity and loading are independent and lognormally distributed, then the risk can be expressed (Huang, 1986)

$$\bar{R} = F_z\left(\frac{-\ln\left[\dfrac{\mu_C}{\mu_L}\left(\dfrac{1+\text{CV}_L^2}{1+\text{CV}_C^2}\right)^{1/2}\right]}{\left\{\ln\left[(1+\text{CV}_C^2)(1+\text{CV}_L^2)\right]\right\}^{1/2}}\right) \tag{13.5.7}$$

Example 13.5.2. Solve Example 13.5.1 assuming capacity and loading are both lognormally distributed.

Solution. From Example 13.5.1, $\mu_C = 5$ and $\sigma_C = 0.75$, and hence $\text{CV}_C = 0.75/5 = 0.15$. Likewise, $\mu_L = 3$ and $\sigma_L = 1$, so $\text{CV}_L = 1/3 = 0.333$. Hence, by Eq. (13.5.7), the risk is

$$\bar{R} = F_z\left(\frac{-\ln\left\{\dfrac{5}{3}\left[\dfrac{1+(0.333)^2}{1+(0.15)^2}\right]^{1/2}\right\}}{\left\{\ln\left[(1+(0.15)^2)(1+(0.333)^2)\right]\right\}^{1/2}}\right)$$

$$= F_z(-1.5463) = 0.061$$

The risk of failure under the above assumptions, then, is 6.1 percent. For the same problem (Example 13.5.1) assuming that the safety margin was normally distributed, the risk was found to be 5.5 percent; the risk level has not changed greatly with use of the lognormal instead of the normal distribution.

Risk–Safety Factor–Return Period Relationship

A common design practice is to choose a return period and determine the corresponding loading L as the design capacity of a hydraulic structure. The safety factor is inherently built into the choice of the return period. Alternatively, the loading value can be multiplied by a safety factor SF; then the structure is designed for capacity $C = \text{SF} \times L$. As discussed in this chapter, there are various kinds of uncertainty associated both with L and with the capacity C of the structure as designed. By composite risk analysis, a risk of failure can be calculated for the selected return period and safety factor. The result of such a calculation is shown

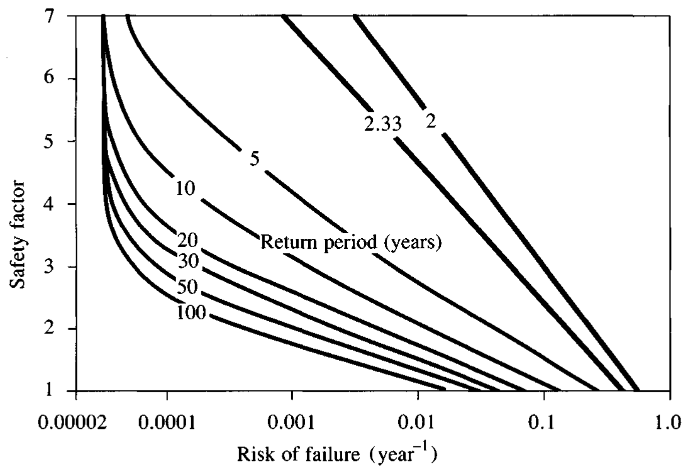

FIGURE 13.5.1
The risk–safety factor–return period relationship for culvert design on the Glade River near Reston, Virginia. The probability distribution for loading used to develop this figure was the Extreme Value Type I distribution of annual maximum floods. A lognormal distribution for the culvert capacity was developed using first-order analysis of uncertainty. The risk level for given return period and safety factor was determined using composite risk analysis. (*Source*: Tung and Mays, 1980.)

in Fig. 13.5.1, which shows a risk chart applying to culvert design on the Glade River near Reston, Virginia. The risk values in the chart represent annual probabilities of failure. For example, if the return period is 100 years and the safety factor 1.0, the risk of failure is 0.015 or 1.5 percent in any given year, while if the safety factor is increased to 2, the risk of failure is reduced to $\bar{R} = 0.006$, or 0.6 percent in any given year.

Current hydrologic design practice copes with the inherent uncertainty of hydrologic phenomena by the selection of the design return period, and with model and parameter uncertainty by the assignment of arbitrary safety factors or safety margins. The risks and uncertainties can be evaluated more systematically using the procedures provided by first-order analysis of uncertainty and composite risk analysis as presented here. However, it must be borne in mind that just as any function of random variables is itself a random variable, the estimates of risk and reliability provided by these methods also have uncertainty associated with them, and their true values can never be determined exactly.

REFERENCES

Ang, A. H.-S., Structural risk analysis and reliability-based design, *J. Structural Div., Am. Soc. Civ. Eng.*, vol. 99, no. ST9, pp. 1891–1910, 1973.

Ang, A. H.-S., and W. H. Tang, *Probability Concepts in Engineering Planning and Design*, vol. I, Basic Principles, and vol. II, Decision, Risk and Reliability, Wiley, New York, 1975 and 1984, respectively.

Chow, V. T., *Open-channel Hydraulics*, McGraw-Hill, New York, 1959.
Corry, M. L., J. S. Jones, and P. L. Thompson, The design of encroachments on flood plains using risk analysis, hydraulic engineering circular no. 17, Federal Highway Administration, U. S. Department of Transportation, 1980.
Harr, M. E., *Reliability-based Design in Civil Engineering*, McGraw-Hill, New York, 1987.
Huang, K.-Z., Reliability analysis on hydraulic design of open channel, in *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B. C. Yen, Water Resources Publications, Littleton, Colo., p. 60, 1986.
Hirsch, R. M., Synthetic hydrology and water supply reliability, *Water Resour. Res.*, vol. 15, no. 6, pp. 1603–1615, December 1979.
Kapur, K. C., and L. R. Lamberson, *Reliability in Engineering Design*, Wiley, New York, 1977.
Lee, H.-L., and L. W. Mays, Improved risk and reliability model for hydraulic structures, *Water Resour. Res.*, vol. 19, no. 6, pp. 1415–1422, 1983.
Lee, H.-L., and L. W. Mays, Hydraulic uncertainties in flood levee capacity, *J. Hyd. Div., Am. Soc. Civ. Eng.*, vol. 112, no. 10, pp. 928–934, 1986.
Milhous, R. T., and W. J. Grenney, The quantification and reservation of instream flows, *Prog. Wat. Tech.*, vol. 13, pp. 129–154, 1980.
National Academy of Sciences, *Safety of Existing Dams: Evaluation and Improvement*, National Academy Press, Washington, D. C., 1983.
Salas, J. D., et al., *Applied modelling of hydrologic time series*, Water Resources Publications, Littleton, Colo., 1980.
Tang, W. H., and B. C. Yen, Hydrologic and hydraulic design under uncertainties, *Proceedings*, International Symposium on Uncertainties in Hydrologic and Water Resources Systems, Tucson, Ariz., vol. 2, pp. 868–882, 1972.
Tung, Y.-K., and L. W. Mays, Risk analysis for hydraulic design, *J. Hyd. Div., Am. Soc. Civ. Eng.*, vol. 106, no. HY5, pp. 893–913, 1980.
U. S. Army Corps of Engineers Hydrologic Engineering Center, computer program catalog, Davis, Calif., August 1986.
World Meteorological Organization, *Guide to Hydrological Practices*, vol. II, Analysis, forecasting, and other applications, WMO no. 168, 4th ed., Geneva, Switzerland, 1983.
Yen, B. C., Risks in hydrologic design of engineering projects, *J. Hyd. Div., Am. Soc. Civ. Eng.*, vol. 96, no. HY4, proc. paper 7229, pp. 959–966, April 1970.
Yen, B. C., H. G. Wenzel, L. W. Mays, and W. H. Tang, Advanced methodologies for design of storm sewer systems, research report no. 112, Water Resources Center, University of Illinois at Urbana-Champaign, August 1976.
Yen, B. C., ed., *Stochastic and Risk Analysis in Hydraulic Engineering*, Water Resources Publications, Littleton, Colo., 1986.

PROBLEMS

13.2.1 The critical drought of record as determined from 30 years of hydrologic data is considered to have lasted for 3 years. If a water supply design is based on this drought and the design life is 50 years, what is the chance that a worse drought will occur during the design life?

13.2.2 In Prob. 13.2.1, what is the chance that a worse drought will occur during the first 10 years of the design life? The first 20 years?

13.2.3 What is the chance that the largest flood observed in 50 years of record will be exceeded during the next 10 years? The next 20 years?

13.2.4 If a structure has a design life of 15 years, calculate the required design return period if the acceptable risk of failure is 20 percent (*a*) in any year, (*b*) over the design life.

13.2.5 A flood plain regulation prevents construction within the 25-year flood plain. What is the risk that a structure built just on the edge of this flood plain will be flooded during the next 10 years? By how much would this risk be reduced if construction were limited to the edge of the 100-year flood plain?

13.2.6 A house has a 30-year design life. What is the chance it will be flooded during its design life if it is located on the edge of the 25-year flood plain? The 100-year flood plain?

13.2.7 Determine the optimum scale of development (return period) for the flood-control measure considered in Example 13.2.3 if the annual capital costs given in Table 13.2.1 are doubled. Use the same damage costs as in Table 13.2.1.

13.2.8 Determine the optimum scale of development (return period) for the flood-control measure considered in Example 13.2.3, if the damage costs are doubled. Annual capital costs remain the same as in Table 13.2.1.

13.2.9 Determine the optimum scale of development (return period) for the flood control measure considered in Example 13.2.3, if the damage costs and the annual capital costs are both doubled.

13.3.1 A rectangular channel is 200 feet wide, has bed slope 0.5 percent, an estimated Manning's n of 0.040, and a design discharge of 10,000 cfs. Calculate the design flow depth. If the coefficient of variation of the design discharge is 0.20 and of Manning's n is 0.15, calculate the standard error of estimate of the flow depth. What is the probability that the actual water level will be more than 1 foot deeper than the expected value? Within what range can the water level for the design event be expected in 70 percent of events?

13.3.2 In Prob. 13.3.1, calculate $\partial y/\partial Q$ and $\partial y/\partial n$ for the conditions given ($Q = 10,000$ cfs and $n = 0.040$) and solve the problem using these derivatives.

13.3.3 Solve Prob. 13.3.1 if the channel is trapezoidal with bottom width 150 ft and side slopes 1 vert. = 3 hor.

13.3.4 Flow in a natural stream channel has been modeled by a computer program and found to have a depth of 15 ft for a flow rate of 8000 cfs and Manning's n value of 0.045. Rerunning the program shows that changing the design discharge by 1000 cfs changes the water surface elevation by 0.8 ft, and changing Manning's n by 0.005 changes the water surface elevation by 0.6 ft. If the design discharge is assumed to be accurate to $\pm$ 30 percent and Manning's n to $\pm$ 10 percent, calculate the corresponding error in the flow depth (or water surface elevation).

13.3.5 Suppose for the conditions given in Example 13.3.1, solved in the text, that the channel wall height adopted is 8.4 ft, that is, the calculated depth of 7.4 ft plus a 1.0 ft freeboard, or safety margin. What safety factor SF is implied by this choice? What would the safety factor be if the true Manning's roughness were 0.045 instead of the 0.035 assumed? Is this a safe design?

13.3.6 Using the first-order analysis of uncertainty for Manning's equation, show that the coefficient of variation of the discharge Q is given by $CV_Q^2 = CV_n^2 + (1/4)CV_{Sf}^2$. What assumptions about the variables in Manning's equation are implied by this equation for CV_Q?

13.3.7 In some instances, flood plain studies are made using channel cross sections determined from topographic maps instead of ground surveys. Extend the first-order analysis of uncertainty for water level in Sec. 13.3 to include uncertainty in the cross-sectional area A and wetted perimeter P. If these variables can be determined with coefficients of variation of 20 percent from topographic maps, calculate the additional risk that the houses in Example 13.3.1 will flood during the design event, resulting from the use of channel cross sections from topographic maps, instead of ground surveys, to delineate the flood plain.

13.4.1 A hydrologic design has a loading with mean value 10 units and standard deviation 2 units. Calculate the risk of failure if the capacity is 12 units. Assume normal distribution for the loading.

13.4.2 Solve Prob. 13.4.1 if the loading is lognormally distributed.

13.4.3 In Prob. 13.4.1, assume that the capacity is normally distributed with mean 12 units and standard deviation 1 unit. Recompute the risk of failure, assuming that the loading is also normally distributed.

13.4.4 About half the total water supply for southern California is provided by long-distance water transfers from northern California and from the Colorado River. The annual demand for these transfers was estimated to be 1.48 MAF (million acre-feet) in 1980, and is projected to rise linearly to 1.77 MAF in 1990. Study of observed annual demands from 1980 to 1985 indicates that the coefficient of variation of observed annual demands around those expected is approximately 0.1 (this variability is due to year-to-year variations in weather and other factors). Estimate the annual demand level that has a 70 percent chance of being equaled or exceeded in 1986 and in 1990. Calculate the chance that observed demands will exceed 2.0 MAF/year in 1986, and in 1990. Assume that the annual demands are normally distributed.

13.4.5 In Prob. 13.4.4, calculate the chance that a limit of 2.0 MAF in water transfers will be exceeded at least once from 1986 to 1990. Assume annual demands are independent from one year to the next.

13.5.1 If capacity and loading are both lognormally distributed, show that risk can be calculated by Eq. (13.5.7):

$$\bar{R} = F_z\left(\frac{-\ln\left[\dfrac{\mu_C}{\mu_L}\left(\dfrac{1+\mathrm{CV}_L^2}{1+\mathrm{CV}_C^2}\right)^{1/2}\right]}{\left\{\ln\left[(1+\mathrm{CV}_C^2)(1+\mathrm{CV}_L^2)\right]\right\}^{1/2}} \right)$$

where F_z denotes the standard normal distribution function.

13.5.2 If capacity and loading are both lognormally distributed, show that risk can be approximated by

$$\bar{R} = F_z\left[\frac{\ln\left(\mu_L/\mu_C\right)}{\left(\mathrm{CV}_L^2 + \mathrm{CV}_C^2\right)^{1/2}}\right]$$

where F_z denotes the standard normal distribution function.

13.5.3 Calculate the risk of failure of an open channel, assuming that the safety margin is normally distributed: Manning's equation is used to compute the capacity, and a first-order analysis is used to determine the coefficient of variation of the capacity C. The mean loading is 5000 cfs and the coefficient of variation of loading is

0.2. The slope of the channel is 0.01 with a coefficient of variation $CV_{S_f} = 0.10$. The Manning's roughness factor is 0.035 and has a coefficient of variation of $CV_n = 0.15$. The channel cross-section is rectangular with width 50 ft and wall height 9 ft. Failure is assumed to occur if the walls are overtopped.

13.5.4 Rework Prob. 13.5.3 to compute the risk of failure, assuming the capacity and loading to be lognormally distributed.

13.5.5 Use the risk analysis of safety margins method to determine the probability that the houses will be flooded in Example 13.3.1 in the text.