6 Frequency and Regression Analysis

R.J. Oosterbaan[1]

6.1 Introduction

Frequency analysis, regression analysis, and screening of time series are the most common statistical methods of analyzing hydrologic data.

Frequency analysis is used to predict how often certain values of a variable phenomenon may occur and to assess the reliability of the prediction. It is a tool for determining design rainfalls and design discharges for drainage works and drainage structures, especially in relation to their required hydraulic capacity.

Regression analysis is used to detect a relation between the values of two or more variables, of which at least one is subject to random variation, and to test whether such a relation, either assumed or calculated, is statistically significant. It is a tool for detecting relations between hydrologic parameters in different places, between the parameters of a hydrologic model, between hydraulic parameters and soil parameters, between crop growth and watertable depth, and so on.

Screening of time series is used to check the consistency of time-dependent data, i.e. data that have been collected over a period of time. This precaution is necessary to avoid making incorrect hydrologic predictions (e.g. about the amount of annual rainfall or the rate of peak runoff).

6.2 Frequency Analysis

6.2.1 Introduction

Designers of drainage works and drainage structures commonly use one of two methods to determine the design discharge. These are:
- Select a design discharge from a time series of measured or calculated discharges that show a large variation;
- Select a design rainfall from a time series of variable rainfalls and calculate the corresponding discharge via a rainfall-runoff transformation.

Frequency analysis is an aid in determining the design discharge and design rainfall. In addition, it can be used to calculate the frequency of other hydrologic (or even non-hydrologic) events.

Because high discharges and rainfalls are comparatively infrequent, the selection of the design discharge can be based on the low frequency with which these high values are permitted to be exceeded. This frequency of exceedance, or the design frequency, is the risk that the designer is willing to accept. Of course, the smaller the risk, the more costly are the drainage works and structures, and the less often their full capacity

[1] International Institute for Land Reclamation and Improvement

will be reached. Accordingly, the design frequency should be realistic – neither too high nor too low.

The following methods of frequency analysis are discussed in this chapter:
- Counting of the number of occurrences in certain intervals (interval method, Section 6.2.2);
- Ranking of the data in ascending or descending order (ranking method, Section 6.2.3);
- Application of theoretical frequency distributions (Section 6.4).

Recurrence predictions and the determination of return periods on the basis of a frequency analysis of hydrologic events are explained in Section 6.2.4.

A frequency – or recurrence – prediction calculated by any of the above methods is subject to statistical error because the prediction is made on the basis of a limited data series. So, there is a chance that the predicted value will be too high or too low. Therefore, it is necessary to calculate confidence intervals for each prediction. A method for constructing confidence intervals is given in Section 6.2.5.

Frequency predictions can be disturbed by two kinds of influences: periodicity and a time trend. Therefore, screening of time series of data for stationarity, i.e. time stability, is important. Although screening should be done before any other frequency analysis, it is explained here at the end of the chapter, in Section 6.6.

6.2.2 Frequency Analysis by Intervals

The interval method is as follows:
- Select a number (k) of intervals (with serial number i, lower limit a_i, upper limit b_i) of a width suitable to the data series and the purpose of the analysis;
- Count the number (m_i) of data (x) in each interval;
- Divide m_i by the total number (n) of data in order to obtain the frequency (F) of data (x) in the i-th interval

$$F_i = F(a_i < x \leq b_i) = m_i/n \tag{6.1}$$

The frequency thus obtained is called the frequency of occurrence in a certain interval. In literature, m_i is often termed the frequency, and F_i is then the relative frequency. But, in this chapter, the term frequency has been kept to refer to F_i.

The above procedure was applied to the daily rainfalls given in Table 6.1. The results are shown in Table 6.2, in Columns (1), (2), (3), (4), and (5). The data are the same data found in the previous edition of this book.

Column (5) gives the frequency distribution of the intervals. The bulk of the rainfall values is either 0 or some value in the 0-25 mm interval. Greater values, which are more relevant for the design capacity of drainage canals, were recorded on only a few days.

From the definition of frequency (Equation 6.1), it follows that the sum of all frequencies equals unity

$$\sum_{i=1}^{k} F_i = \sum_{i=1}^{k} m_i/n = n/n = 1 \tag{6.2}$$

Table 6.1 Daily rainfall in mm for the month of November in 19 consecutive years

Year	Day 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1948	-	-	-	3	3	45	15	-	1	5	-	-	4	6	-	-
1949	-	-	2	10	9	4	10	-	-	-	-	-	-	-	-	-
1950	11	3	-	2	13	-	8	26	12	1	5	6	-	-	-	-
1951	29	-	6	99	4	3	-	-	-	-	-	-	5	3	1	-
1952	111	8	8	21	1	-	11	11	26	1	-	-	-	-	-	1
1953	-	-	-	-	1	14	4	33	3	-	12	-	11	15	3	25
1954	-	-	-	-	-	-	-	1	-	1	5	1	7	-	1	36
1955	-	-	-	10	-	23	3	-	49	12	57	2	-	1	-	-
1956	-	-	2	9	-	-	-	6	3	-	-	2	-	-	9	16
1957	4	-	41	46	-	-	-	-	-	-	23	7	1	18	8	2
1958	92	3	-	2	-	-	-	9	6	5	-	13	-	-	1	17
1959	-	65	19	-	35	3	27	10	-	13	32	1	16	2	-	-
1960	-	-	-	-	9	10	-	-	-	-	-	-	-	-	-	-
1961	41	158	1	-	10	-	6	11	-	-	1	-	7	1	13	12
1962	-	9	-	10	-	4	-	-	2	5	-	13	16	24	2	7
1963	74	7	-	2	4	-	10	-	-	-	-	42	11	-	-	1
1964	7	23	4	-	1	13	29	40	13	14	1	4	-	-	-	-
1965	11	13	-	2	-	-	-	3	-	8	56	3	44	5	-	-
1966	11	-	-	-	-	3	-	-	2	54	65	16	-	-	-	-

Year	Day 17	18	19	20	21	22	23	24	25	26	27	28	29	30	Total	Max
1948	40	12	-	-	-	-	-	-	-	-	-	-	-	-	134	45
1949	-	-	-	-	-	-	-	-	-	-	-	-	-	-	35	10
1950	-	-	-	-	-	-	7	4	-	22	5	31	67	-	223	67
1951	-	1	-	-	-	-	-	-	-	19	5	3	21	46	245	99
1952	-	-	-	5	7	4	8	2	53	3	6	1	3	21	312	111
1953	21	-	-	2	11	-	2	18	38	-	5	4	7	6	235	38
1954	4	-	-	-	4	-	-	1	-	-	3	-	-	-	64	36
1955	18	3	-	-	-	-	11	4	-	1	1	23	15	2	235	57
1956	14	9	-	-	-	-	-	-	-	-	-	-	-	30	100	30
1957	4	-	-	6	38	3	14	2	-	-	-	-	1	13	231	46
1958	22	3	1	20	-	20	7	14	1	1	22	1	22	12	294	92
1959	-	-	-	-	-	-	1	-	-	-	-	7	12	-	243	65
1960	7	3	3	-	-	-	-	-	28	24	22	-	-	8	114	28
1961	11	-	-	2	4	-	-	-	-	-	-	-	-	-	278	158
1962	-	-	-	-	200	94	4	-	5	1	12	14	-	-	422	200
1963	4	1	14	-	4	8	-	-	-	20	5	-	30	5	242	74
1964	-	3	2	-	-	-	-	-	20	4	37	15	6	4	240	40
1965	-	-	4	4	2	3	-	-	-	11	-	-	0	-	169	56
1966	19	14	-	-	9	-	-	-	-	-	-	-	1	7	201	65

In hydrology, we are often interested in the frequency with which data exceed a certain, usually high design value. We can obtain the frequency of exceedance $F(x > a_i)$ of the lower limit a_i of a depth interval i by counting the number M_i of all rainfall values x exceeding a_i, and by dividing this number by the total number of rainfall data. This is shown in Table 6.2, Column (6). In equation form, this appears as

$$F(x > a_i) = M_i/n \tag{6.3}$$

Table 6.2 Frequency analysis of daily rainfall, based on intervals, derived from Table 6.1 (Column numbers are in brackets)

Serial number	Depth interval (mm)		Number of observations with $a_i < x \leq b_i$	Frequency $F(a_i < x \leq b_i)$	Exceedance frequency $F(x > a_i)$	Cumulative frequency $F(x \leq a_i) = 1 - F(x > a_i)$	Return period	
							T_r (days) n/M_i	T_r (years) $\frac{n/30}{M_i}$
i	Lower limit a_i excl.	Upper limit b_i incl.	m_i	m_i/n (Eq. 6.1)	M_i/n (Eq. 6.3)	= 1-(6) (Eq. 6.5)	= 1/(6) (Eq. 6.9)	= (8)/30 (Eq. 6.10)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	< 0	0	285	0.500	1.000	0.000	1	0.033
2	0	25	246	0.432	0.500	0.500	2	0.067
3	25	50	25	0.0439	0.0684	0.932	15	0.49
4	50	75	8	0.0140	0.0246	0.975	41	1.4
5	75	100	3	0.00526	0.0105	0.989	95	3.2
6	100	125	1	0.00175	0.00526	0.995	190	6.3
7	125	150	0	0.00000	0.00351	0.996	285	9.5
8	150	175	1	0.00175	0.00351	0.996	285	9.5
9	175	200	1	0.00175	0.00175	0.998	570	19

$k = 9$ $\qquad$ $n = \Sigma m_i = 570$

Frequency distributions are often presented as the frequency of non-exceedance and not as the frequency of occurrence or of exceedance. The frequency of non-exceedance is also referred to as the cumulative frequency. We can obtain the frequency of non-exceedance $F(x \leq a_i)$ of the lower limit a_i by calculating the sum of the frequencies over the intervals below a_i.

Because the sum of the frequencies over all intervals equals unity, it follows that

$$F(x > a_i) + F(x \leq a_i) = 1 \qquad (6.4)$$

The cumulative frequency (shown in Column (7) of Table 6.2) can, therefore, be derived directly from the frequency of exceedance as

$$F(x \leq a_i) = 1 - F(x > a_i) = 1 - M_i/n \qquad (6.5)$$

Columns (8) and (9) of Table 6.2 show return periods. The calculation of these periods will be discussed later, in Section 6.2.4.

Censored Frequency Distributions

Instead of using all available data to make a frequency distribution, we can use only certain selected data. For example, if we are interested only in higher rainfall rates, for making drainage design calculations, it is possible to make a frequency distribution only of the rainfalls that exceed a certain value. Conversely, if we are interested in water shortages, it is also possible to make a frequency distribution of only the rainfalls that are below a certain limit. These distributions are called censored frequency distributions.

In Table 6.3, a censored frequency distribution is presented of the daily rainfalls, from Table 6.1, greater than 25 mm. It was calculated without intervals $i = 1$ and $i = 2$ of Table 6.2.

The remaining frequencies presented in Table 6.3 differ from those in Table 6.2 in that they are conditional frequencies (the condition in this case being that the rainfall is higher than 25 mm). To convert conditional frequencies to unconditional frequencies, the following relation is used

$$F = (1 - F^*)F' \qquad (6.6)$$

where

F = unconditional frequency (as in Table 6.2)
F' = conditional frequency (as in Table 6.3)
F^* = frequency of occurrence of the excluded events (as in Table 6.2)

As an example, we find in Column (7) of Table 6.3 that $F'(x \leq 50) = 0.641$. Further, the cumulative frequency of the excluded data equals $F^*(x \leq 25) = 0.932$ (see Column (7) of Table 6.2). Hence, the unconditional frequency obtained from Equation 6.6 is

$$F(x \leq 50) = (1 - 0.932) \times 0.641 = 0.0439$$

This is exactly the value found in Column (5) of Table 6.2.

Table 6.3 Censored frequency distribution of daily rainfalls higher than 25 mm, based on intervals, derived from Table 6.1 (column numbers are in brackets)

Serial number	Depth interval (mm)		Number of observations with $a_i < x \le b_i$	Conditional frequency $F'(a_i < x \le b_i)$	Conditional frequency $F'(x > a_i)$	Conditional frequency $F'(x \le a_i) = 1 - F(x > a_i)$	Conditional return period T'_r (days) n/M_i	T'_r years $\frac{n/30}{M_i}$
i	Lower limit a_i excl.	Upper limit b_i incl.	m_i	m_i/n	M_i/n	= 1-(6)	= 1/(6)	= (8)/30
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	25	50	25	0.641	1.000	0.000	1.0	0.033
2	50	75	8	0.205	0.359	0.641	2.8	0.093
3	75	100	3	0.0769	0.154	0.846	6.5	0.22
4	100	125	1	0.0256	0.0769	0.923	13.0	0.43
5	125	150	0	0.0000	0.0513	0.949	19.5	0.65
6	150	175	1	0.0256	0.0513	0.949	19.5	0.65
7	175	200	1	0.0256	0.0256	0.974	39.0	1.3

$k = 7$ $\qquad$ $n = \Sigma m_i = 39$

6.2.3 Frequency Analysis by Ranking of Data

Data for frequency analysis can be ranked in either ascending or descending order. For a ranking in descending order, the suggested procedure is as follows:

- Rank the total number of data (n) in descending order according to their value (x), the highest value first and the lowest value last;
- Assign a serial number (r) to each value x (x_r, r = 1,2,3,...,n), the highest value being x_1 and the lowest being x_n;
- Divide the rank (r) by the total number of observations plus 1 to obtain the frequency of exceedance

$$F(x > x_r) = \frac{r}{n+1} \tag{6.7}$$

- Calculate the frequency of non-exceedance

$$F(x \leq x_r) = 1 - F(x > x_r) = 1 - \frac{r}{n+1} \tag{6.8}$$

If the ranking order is ascending instead of descending, we can obtain similar relations by interchanging $F(x > x_r)$ and $F(x \leq x_r)$.

An advantage of using the denominator n + 1 instead of n (which was used in Section 6.2.2) is that the results for ascending or descending ranking orders will be identical.

Table 6.4 shows how the ranking procedure was applied to the monthly rainfalls of Table 6.1. Table 6.5 shows how it was applied to the monthly maximum 1-day rainfalls of Table 6.1. Both tables show the calculation of return periods (Column 7), which will be discussed below in Section 6.2.4. Both will be used again, in Section 6.4, to illustrate the application of theoretical frequency distributions.

The estimates of the frequencies obtained from Equations 6.7 and 6.8 are not unbiased. But then, neither are the other estimators found in literature. For values of x close to the average value ($\overline{x}$), it makes little difference which estimator is used, and the bias is small. For extreme values, however, the difference, and the bias, can be relatively large. The reliability of the predictions of extreme values is discussed in Section 6.2.5.

6.2.4 Recurrence Predictions and Return Periods

An observed frequency distribution can be regarded as a sample taken from a frequency distribution with an infinitely long observation series (the 'population'). If this sample is representative of the population, we can then expect future observation periods to reveal frequency distributions similar to the observed distribution. The expectation of similarity ('representativeness') is what makes it possible to use the observed frequency distribution to calculate recurrence estimates.

Representativeness implies the absence of a time trend. The detection of possible time trends is discussed in Section 6.6.

It is a basic law of statistics that if conclusions about the population are based on a sample, their reliability will increase as the size of the sample increases. The smaller

Table 6.4 Frequency distributions based on ranking of the monthly rainfalls of Table 6.1

Rank	Rainfall (descending)		Year	$F(x > x_r)$	$F(x \le x_r)$	T_r (years)
r	x_r	x_r^2 *		$r/(n+1)$	$1-r/(n+1)$	$(n+1)/r$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	422	178084	1962	0.05	0.95	20
2	312	97344	1952	0.10	0.90	10
3	294	86436	1958	0.15	0.85	6.7
4	278	77284	1961	0.20	0.80	5.0
5	245	60025	1951	0.25	0.75	4.0
6	243	59049	1959	0.30	0.70	3.3
7	242	58564	1964	0.35	0.65	2.9
8	240	57600	1963	0.40	0.60	2.5
9	235	55225	1953	0.45	0.55	2.2
10	235	55225	1955	0.50	0.50	2.0
11	231	53361	1957	0.55	0.45	1.82
12	223	49729	1950	0.60	0.40	1.67
13	201	40401	1966	0.65	0.35	1.54
14	169	28561	1965	0.70	0.30	1.43
15	134	17956	1948	0.75	0.25	1.33
16	114	12996	1960	0.80	0.20	1.25
17	100	10000	1956	0.85	0.15	1.18
18	64	4096	1954	0.90	0.10	1.11
19	35	1225	1949	0.95	0.05	1.05

$n = 19 \quad \sum_{r=1}^{n} x_r = 4017 \quad \sum_{r=1}^{n} x_r^2 = 1003161$

* Tabulated for parametric distribution-fitting (see Section 6.4)

the frequency of occurrence of an event, the larger the sample will have to be in order to make a prediction with a specified accuracy. For example, the observed frequency of dry days given in Table 6.2 (0.5, or 50%) will deviate only slightly from the frequency observed during a later period of at least equal length. The frequency of daily rainfalls of 75-100 mm (0.005, or 0.5%), however, can be easily doubled (or halved) in the next period of record.

A quantitative evaluation of the reliability of frequency predictions follows in the next section.

Recurrence estimates are often made in terms of return periods (T), T being the number of new data that have to be collected, on average, to find a certain rainfall value. The return period is calculated as $T = 1/F$, where F can be any of the frequencies discussed in Equations 6.1, 6.3, 6.5, and 6.6. For example, in Table 6.2, the frequency F of 1-day November rainfalls in the interval of 25-50 mm equals 0.04386, or 4.386%. Thus the return period is $T = 1/F = 1/0.04386 = 23$ November days.

In hydrology, it is very common to work with frequencies of exceedance of the variable x over a reference value x_r. The corresponding return period is then

$$T_r = \frac{1}{F(x > x_r)} \tag{6.9}$$

Table 6.5 Frequency distributions based on ranking of the maximum 1-day rainfalls per month of Table 6.1

Rank	Rainfall (descending)		Year	$F(x > x_r)$	$F(x \leq x_r)$	T_r (years)
r	x_r	x_r^2 *		r/(n+1)	1-r/(n+1)	(n+1)/r
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	200	40000	1962	0.05	0.95	20
2	158	24964	1961	0.10	0.90	10
3	111	12321	1952	0.15	0.85	6.7
4	99	9801	1951	0.20	0.80	5.0
5	92	8464	1958	0.25	0.75	4.0
6	74	5476	1963	0.30	0.70	3.3
7	67	4489	1950	0.35	0.65	2.9
8	65	4225	1966	0.40	0.60	2.5
9	65	4225	1959	0.45	0.55	2.2
10	57	3249	1955	0.50	0.50	2.0
11	56	3136	1965	0.55	0.45	1.82
12	46	2116	1957	0.60	0.40	1.67
13	45	2025	1948	0.65	0.35	1.54
14	40	1600	1964	0.70	0.30	1.43
15	38	1444	1953	0.75	0.25	1.33
16	36	1296	1954	0.80	0.20	1.25
17	30	900	1956	0.85	0.15	1.18
18	28	784	1960	0.90	0.10	1.11
19	10	100	1949	0.95	0.05	1.05

$n = 19 \quad \sum_{r=1}^{n} x_r = 1317 \quad \sum_{r=1}^{n} x_r^2 = 130615$

* Tabulated for parametric distribution-fitting (see Section 6.4)

For example, in Table 6.2 the frequency of exceedance of 1-day rainfalls of $x_r = 100$ mm in November is $F(x > 100) = 0.00526$, or 0.526%. Thus the return period is

$$T_{100} = \frac{1}{F(x > 100)} = \frac{1}{0.00526} = 190 \text{ (November days)}$$

In design, T is often expressed in years

$$T(\text{Years}) = \frac{T}{\text{number of independent observations per year}} \tag{6.10}$$

As the higher daily rainfalls can generally be considered independent of each other, and as there are 30 November days in one year, it follows from the previous example that

$$T_{100}(\text{years}) = \frac{T_{100}\text{ (November days)}}{30} = \frac{190}{30} = 6.33 \text{ years}$$

This means that, on average, there will be a November day with rainfall exceeding 100 mm once in 6.33 years.

If a censored frequency distribution is used (as it was in Table 6.3), it will also be necessary to use the factor 1-F* (as shown in Equation 6.6) to adjust Equation 6.10 This produces

$$T(\text{Years}) = \frac{T' / (1-F^*)}{\text{number of independent observations per year}} \qquad (6.11)$$

where T′ is the conditional return period ($T' = 1/F'$).

In Figure 6.1, the rainfalls of Tables 6.2, 6.4, and 6.5 have been plotted against their respective return periods. Smooth curves have been drawn to fit the respective points as well as possible. These curves can be considered representative of average future frequencies. The advantages of the smoothing procedure used are that it enables interpolation and that, to a certain extent, it levels off random variation. Its disadvantage is that it may suggest an accuracy of prediction that does not exist. It is therefore useful to add confidence intervals for each of the curves in order to judge the extent of the curve's reliability. (This will be discussed in the following section.)

From Figure 6.1, it can be concluded that, if T_r is greater than 5, it makes no significant difference if the frequency analysis is done on the basis of intervals of all 1-day rainfalls or on the basis of maximum 1-day rainfalls only. This makes it possible to restrict the analysis to maximum rainfalls, which simplifies the calculations and produces virtually the same results.

The frequency analysis discussed here is usually adequate to solve problems related to agriculture. If there are approximately 20 years of information available, predictions for 10-year return periods, made with the methods described in this section, will be reasonably reliable, but predictions for return periods of 20 years or more will be less reliable.

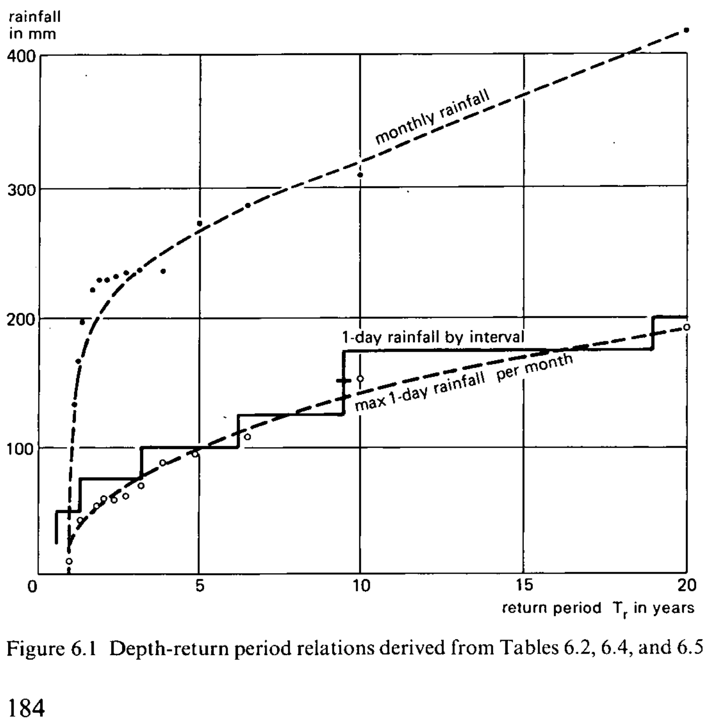

Figure 6.1 Depth-return period relations derived from Tables 6.2, 6.4, and 6.5

6.2.5 Confidence Analysis

Figure 6.2 shows nine cumulative frequency distributions that were obtained with the ranking method. They are based on different samples, each consisting of 50 observations taken randomly from 1000 values. The values obey a fixed distribution (the base line). It is clear that each sample reveals a different distribution, sometimes close to the base line, sometimes away from it. Some of the lines are even curved, although the base line is straight.

Figure 6.2 also shows that, to give an impression of the error in the prediction of future frequencies, frequency estimates based on one sample of limited size should be accompanied by confidence statements. Such an impression can be obtained from Figure 6.3, which is based on the binomial distribution. The figure illustrates the principle of the nomograph. Using N = 50 years of observation, we can see that the 90% confidence interval of a predicted 5-year return period is 3.2 to 9 years. These values are obtained by the following procedure:

- Enter the graph on the vertical axis with a return period of T_r = 5, (point A), and move horizontally to intersect the baseline, with N = ∞, at point B;

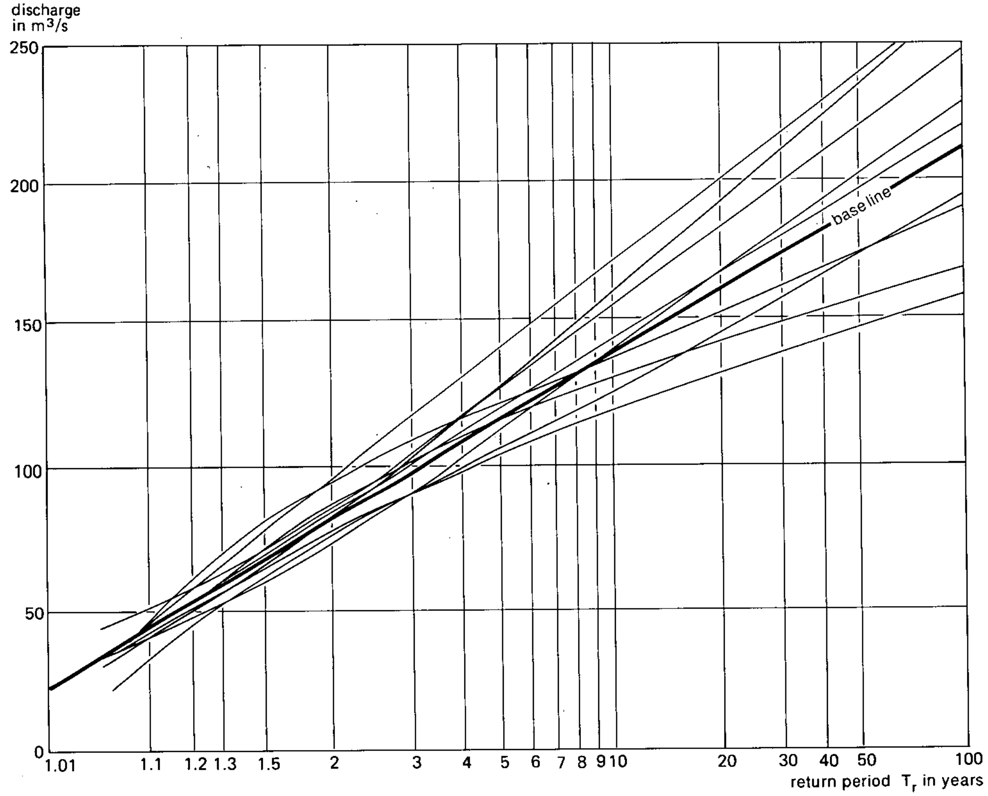

Figure 6.2 Frequency curves for different 50-year sample periods derived from the same base distribution (after Benson 1960)

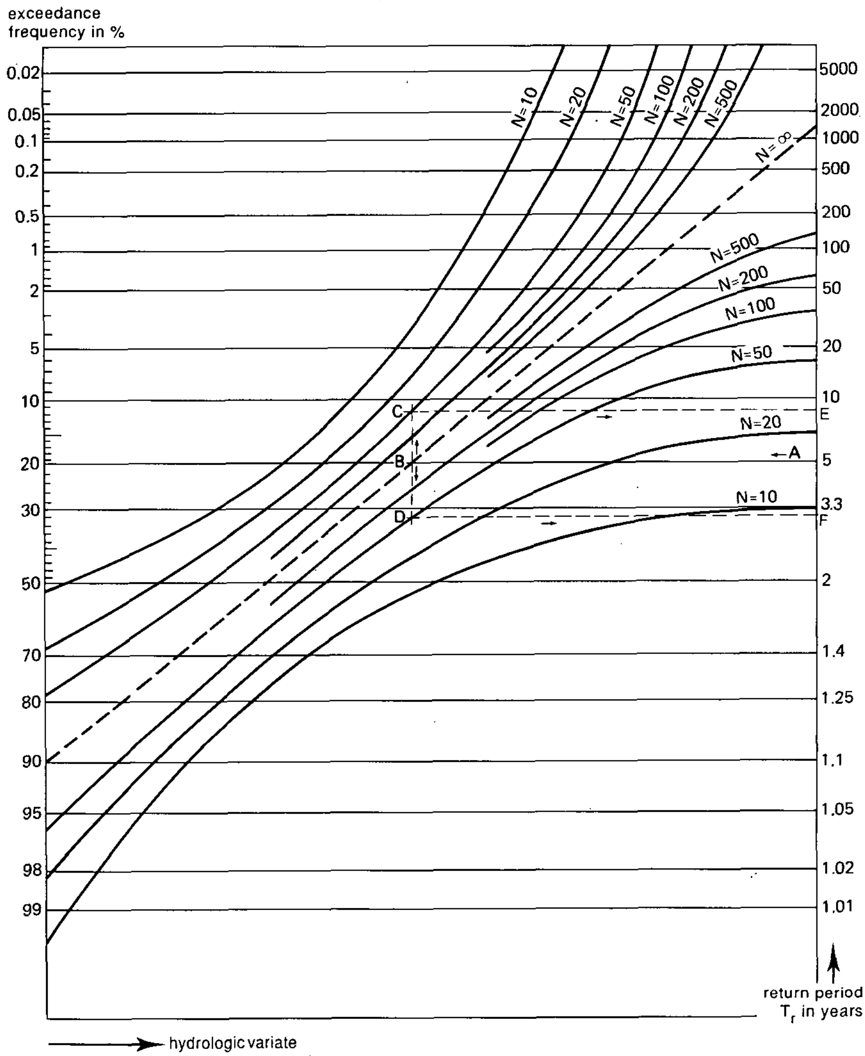

Figure 6.3 90% Confidence belts of frequencies for different values of sample size N

- Move vertically from the intersection point (B) and intersect the curves for N = 50 to obtain points C and D;
- Move back horizontally from points C and D to the axis with the return periods and read points E and F;
- The interval from E to F is the 90% confidence interval of A, hence it can be predicted with 90% confidence that T_r is between 3.2 and 9 years. Nomographs for confidence intervals other than 90% can be found in literature (e.g. in Oosterbaan 1988).

By repeating the above procedure for other values of T_r, we obtain a confidence belt.

In theory, confidence belts are somewhat wider than those shown in the graph. The reason for this is that mean values and standard deviations of the applied binomial distributions have to be estimated from a data series of limited length. Hence, the true means and standard deviations can be either smaller or larger than the estimated ones. In practice, however, the exact determination of confidence belts is not a primary concern because the error made in estimating them is small compared to their width.

The confidence belts in Figure 6.3 show the predicted intervals for the frequencies that can be expected during a very long future period with no systematic changes in hydrologic conditions. For shorter future periods, the confidence intervals are wider than indicated in the graphs. The same is true when hydrologic conditions change.

In literature, there are examples of how to use a probability distribution of the hydrologic event itself to construct confidence belts (Oosterbaan 1988). There are advantages, however, to use a probability distribution of the frequency to do this. This method can also be used to assess confidence intervals of the hydrologic event, which we shall discuss in Section 6.4.

6.3 Frequency-Duration Analysis

6.3.1 Introduction

Hydrologic phenomena are continuous, and their change in time is gradual. Because they are not discrete in time, like the yield data from a crop, for example, they are sometimes recorded continuously. But before continuous records can be analyzed for certain durations, they must be made discrete, i.e. they must be sliced into predetermined time units. An advantage of continuous records is that these slices of time can be made so small that it becomes possible to follow a variable phenomenon closely. Because many data are obtained in this way, discretization is usually done by computer.

Hydrologic phenomena (e.g rainfall) are recorded more often at regular time intervals (e.g. daily) than continuously. For phenomena like daily rainfall totals, it is difficult to draw conclusions about durations shorter than the observation interval. Longer durations can be analyzed if the data from the shorter intervals are added. This technique is explained in the following section.

The processing of continuous records is not discussed, but the principles are almost the same as those used in the processing of measurements at regular intervals, the main distinction being the greater choice of combinations of durations if continuous records are available.

Although the examples that follow refer to rainfall data, they are equally applicable to other hydrologic phenomena.

6.3.2 Duration Analysis

Rainfall is often measured in mm collected during a certain interval of time (e.g. a day). For durations longer than two or more of these intervals, measured rainfalls can be combined into three types of totals:

- Successive totals;
- Moving totals;
- Maximum totals.

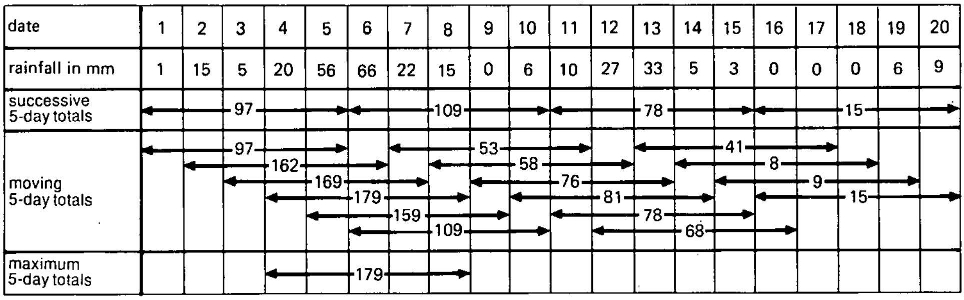

Figure 6.4 Illustration of various methods for the composition of 5-day totals

Examples of these combining methods are given in Figure 6.4 for 5-day totals that are made up of 1-day rainfalls.

To form successive 5-day totals, break up the considered period or season of measurement into consecutive groups of 5 days and calculate the total rainfall for each group. Successive totals have the drawback of sometimes splitting periods of high rainfall into two parts of lesser rainfall, thus leading us to underestimate the frequency of high rainfall.

To form moving 5-day totals, add the rainfall from each day in the considered period to the totals from the following 4 days. Because of the overlap, each daily rainfall will be represented 5 times. So even though, for example, in November there are 26 moving 5-day totals, we have only 6 non-overlapping totals (the same as the number of successive 5-day totals) to calculate the return period. The advantage of the moving totals is that, because they include all possible 5-day rainfalls, we cannot underestimate the high 5-day totals. The drawbacks are that the data are not independent and that a great part of the information may be of little interest.

To avoid these drawbacks, censored data series are often used. In these series, the less important data are omitted (e.g. low rainfalls – at least when the design capacity of drainage canals is being considered), and only exceedance series or maximum series are selected. Thus we can choose, for example, a maximum series consisting of the highest 5-day moving totals found for each month or year and then use the interval or ranking procedure to make a straightforward frequency distribution or return-period analysis for them. We must keep in mind, however, that the second highest rainfall in a certain month or year may exceed the maximum rainfall recorded in some other months or years and that, consequently, the rainfalls estimated from maximum series with return periods of less than approximately 5 years will be underestimated in comparison with those obtained from complete or exceedance series. It is, therefore, a good idea not to work exclusively with maximum series when making calculations for agriculture.

6.3.3 Depth-Duration-Frequency Relations

Having analyzed data both for frequency and for duration, we arrive at depth-duration-frequency relations. These relations are valid only for the point where the observations were made, and not for larger areas. Figure 6.5A shows that rainfall-return period relations for short durations are steeper for point rainfalls than for area rainfalls, but that, for longer durations, the difference is less. Figure 6.5B illustrates qualitatively the effect of area on the relation between duration-frequency curves. It shows how rainfall increases with area when the return period is short ($T_r < 2$), whereas, for long return periods ($T_r > 2$), the opposite is true. It also shows that larger areas have less variation in rainfall than smaller areas, but that the mean rainfall is the same. Note that, in both figures, the return period of the mean value is $T_r \approx 2$. This means an exceedance frequency of $F(x > x_r) = 1/T_r \approx 0.5$, which corresponds to the median value. So it is assumed that the mean and the median are about equal.

Instead of working with rainfall totals of a certain duration, we can work with the average rainfall intensity, i.e. the total divided by the duration (Figure 6.6).

Procedure and Example

The data in Table 6.1 are from a tropical rice-growing area. November, when the rice seedlings have just been transplanted, is a critical month: an abrupt rise of more than 75 mm (the maximum permissible storage increase) of the standing water in the paddy fields due to heavy rains would be harmful to the seedlings. A system of ditches is to be designed to transport the water drained from the fields.

To find the design discharge of the ditches, we first use a frequency-duration analysis to determine the frequency distributions of, for example, 1-, 2-, 3-, and 5-day rainfall totals. From this analysis, we select and plot these totals with return periods of 5, 10, and 20 years (Figure 6.7).

To find the required design discharge in relation to the return period (accepted risk of inadequate drainage), we draw tangent lines from the 75-mm point on the depth axis to the various duration curves. The slope of the tangent line indicates the design discharge. If we shift the tangent line so that it passes through the zero point of the coordinate axes, we can see that, for a 5-year return period, the drainage capacity should be 25 mm/d. We can see that the maximum rise would then equal the permissible rise (75 mm), and that it would take about 5 days to drain off all the water from this rainstorm. If we take the design return period as 10 years, the discharge capacity

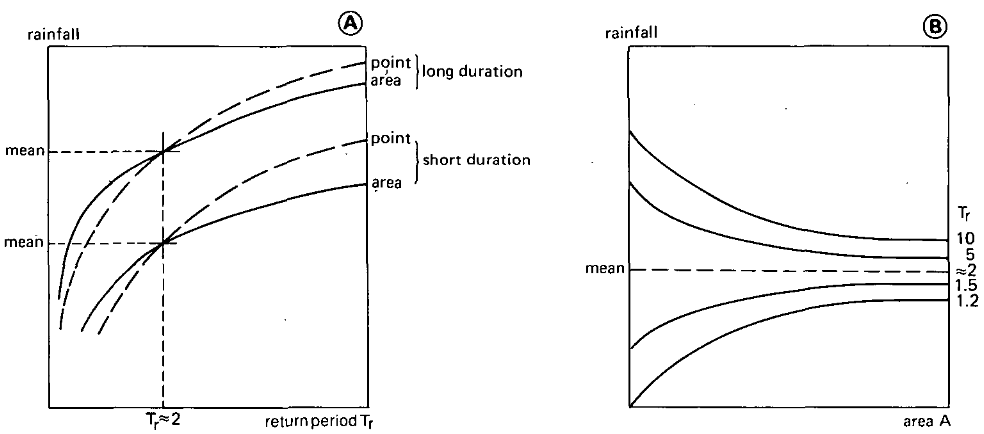

Figure 6.5 The influence of area size on frequency-duration relations of rainfall. A: Flattening effect of the duration on area rainfalls as compared with point rainfalls. B: Flattening effect of the size of the area on area rainfalls of various return periods as compared with point rainfalls

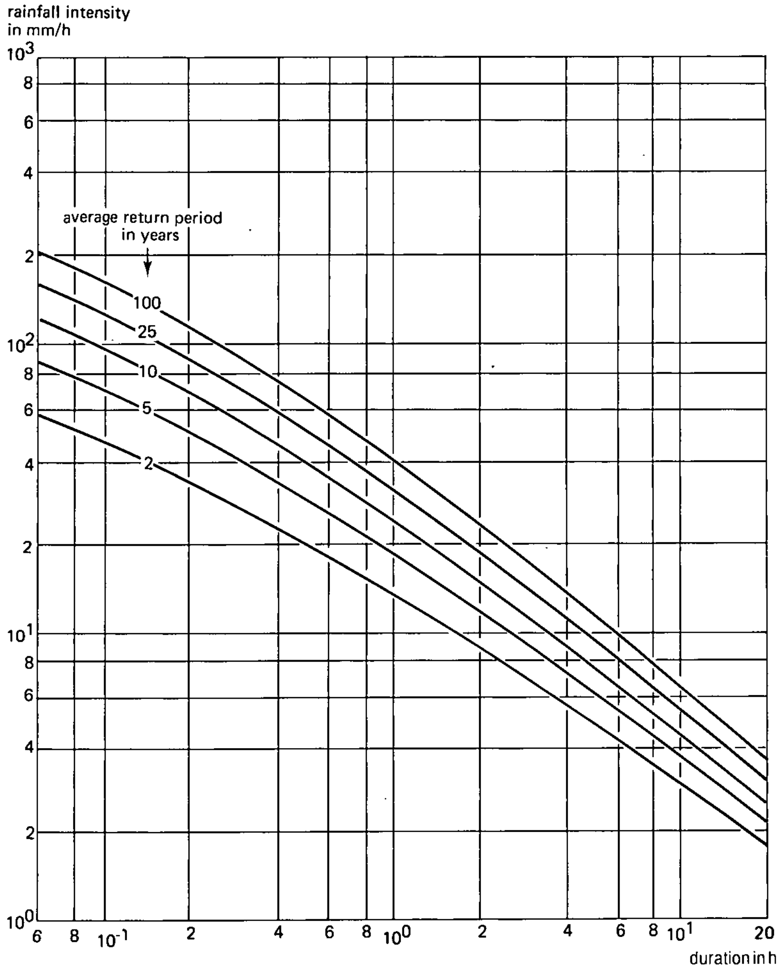

Figure 6.6 Intensity-duration-frequency relations of rainfall (Fresno, California, U.S.A. 1903-1941)

would be 50 mm/d, and for 20 years it would be 150 mm/d. We can also see that the critical durations, indicated by the tangent points, become shorter as the return period increases (to about 1.4, 0.9, and 0.4 days). In other words: as the return period increases, the design rainfall increases, the maximum permissible storage becomes relatively smaller, and we have to reckon with more intensive rains of shorter duration.

Because the return periods used in the above example are subject to considerable statistical error, it will be necessary to perform a confidence analysis.

So far, we have analyzed only durations of a few days in a certain month. Often, however, it is necessary to expand the analysis to include longer durations and all months of the year. Figure 6.8 illustrates an example of this that is useful for water-resources planning.

In addition to deriving depth-duration-frequency relations of rainfall, we can use these same principles to derive discharge-duration-frequency relations of river flows.

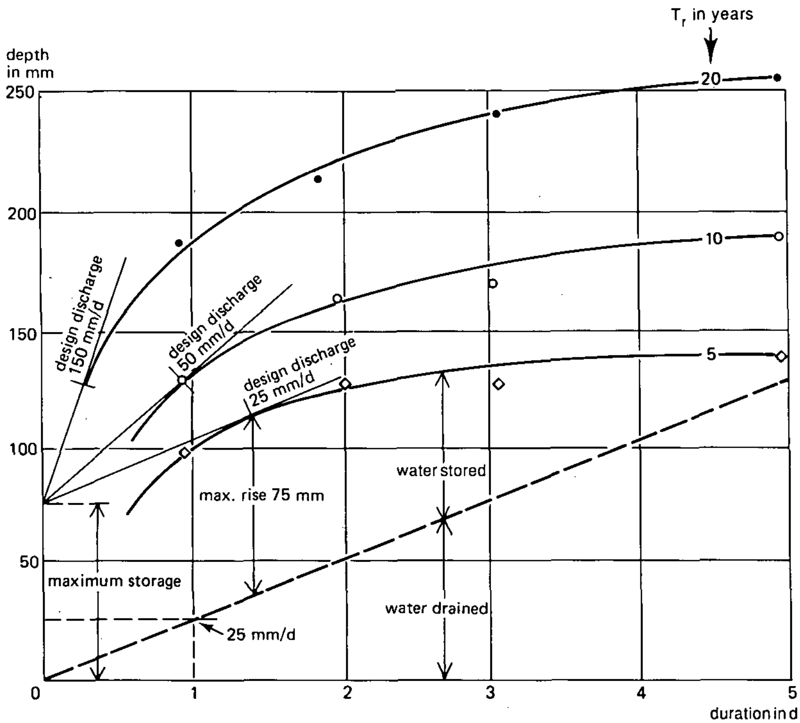

Figure 6.7 Depth-duration-frequency relation derived from Table 6.1 and applied to determine the design discharge of a surface drainage system

6.4 Theoretical Frequency Distributions

6.4.1 Introduction

To arrive at a mathematical formula for a frequency distribution, we can try to fit a theoretical frequency distribution (given by a mathematical expression) to the data series. If the theoretical distribution fits the data reasonably well, it can be used to convert the confidence limits of frequencies or return periods into the confidence limits of the hydrologic phenomenon studied (Section 6.4.6). Further, the fitted distribution can be used not only to interpolate, but also to extrapolate, i.e. to find return periods of extreme values that were not apparent during the relatively short period of observation. We should, however, be very cautious with such extrapolations because:

- Observed frequencies are subject to random variation and so, consequently, the same is true of the fitted theoretical distribution;
- The error will increase as the phenomenon becomes more extreme or exceptional;
- Many different theoretical distributions can be made to fit the observed distribution well, but they can lead to different predictions for extrapolated values.

Of the many existing theoretical frequency distributions, only three have been selected for discussion in this chapter. They are:

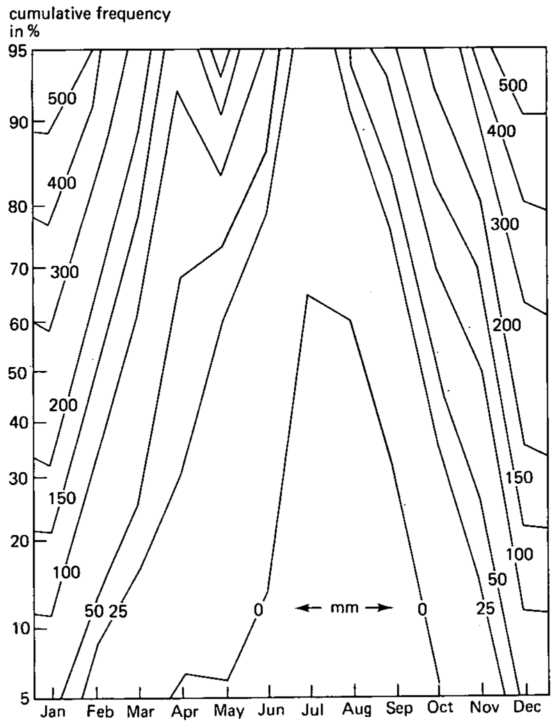

Figure 6.8 Frequencies of monthly rainfalls (Antalya, Turkey, 15 years of observations)

- The normal distribution, which is widely applicable and which forms the basis of many frequency analyses;
- The Gumbel distribution, which is very often used to analyze the frequency of maximum series;
- The exponential distribution, which is very simple and which can often be used instead of the Gumbel distribution.

The majority of hydrologic frequency curves can be described adequately with these few theoretical frequency distributions. The choice of the most appropriate theoretical distribution is a matter of judgement.

6.4.2 Principles of Distribution Fitting

There are two methods of fitting theoretical distributions to the data. They are:
- The plotting, graphic, or regression method. Plot the results obtained from the ranking method on probability paper of a type that corresponds to the selected theoretical distribution and construct the best-fitting line;
- The parametric method. Determine the parameters of the theoretical distribution (e.g. the mean and standard deviation) from the data.

Examples of distribution fitting are given in the following sections. The emphasis will be on the parametric method.

It has been observed that hydrologic data averaged over a long duration (e.g. average yearly discharges) often conform to the normal distribution. Similarly, the maxima inside long-time records (e.g. the maximum 1-day discharge per year) often conform to the Gumbel or to the exponential distribution. According to probability theory, this conformation becomes better as the records from which the maximum is chosen become longer, long records being the best guarantee of a reliable distribution fitting.

Determining the Parameters

For theoretical frequency distributions, the following parameters (characteristics of the distribution) are used:

- μ, the mean value of the distribution;
- σ, the standard deviation of the distribution, which is a measure for the dispersion of the data.

These parameters are estimated from a data series with Estimate (μ) = $\bar{x}$ and Estimate (σ) = s, where $\bar{x}$ and s are determined from

$$\bar{x} = \frac{1}{n} \Sigma x_i \tag{6.12}$$

$$s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 = \frac{1}{n-1} \left\{ \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n} \right\} \tag{6.13}$$

where x_i is the value of the i-th observation of phenomenon x, and n is the total number of observations. Hence, i = 1, 2, 3, .., n.

Once $\bar{x}$ and s are known, estimated frequencies can be calculated from the theoretical distributions for each value of x. The estimated parameters, like the frequency, are subject to random error, which becomes smaller as n increases.

In this chapter, the parametric method is preferred over the plotting method because the estimates of $\bar{x}$ and s (Equations 6.12 and 6.13) are unbiased, whereas the advance estimates of frequencies, which are needed for the plotting method, are probably not unbiased. Moreover, the parametric method is simpler and more straightforward.

The plotting method introduces an artificially high correlation between the data and the frequencies because of the ranking procedure, and the relatively small deviations of the plotting positions from the fitted distribution are no measure of reliability (Section 6.4.6).

6.4.3 The Normal Distribution

The normal frequency distribution, also known as the Gauss or the De Moivre distribution, cannot be expressed directly as a frequency of occurrence. Hence, it is expressed as a frequency density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \tag{6.14}$$

where

$f(x)$ = the normal frequency density function of x
x = the normal variate ($-\infty < x < \infty$)
μ = the mean of the distribution
σ = the standard deviation of the distribution.

A frequency of occurrence in a certain interval a-b can be found from

$$F(a < x < b) = \int_a^b f(x)\,dx \qquad (6.15)$$

so that the cumulative (or non-exceedance) frequency of x_r equals

$$F(x < x_r) = \int_{-\infty}^{x_r} f(x)\,dx$$

and the exceedance frequency of x_r equals

$$F(x > x_r) = \int_{x_r}^{\infty} f(x)\,dx$$

To solve this, it is necessary to use tables of the standard normal distribution, as analytic integration is not possible.

Because $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (compare with Equation 6.2), it follows that

$$F(x > x_r) = 1 - F(x \leq x_r) \qquad (6.16)$$

which is comparable to Equation 6.4.

Figure 6.9A illustrates a normal frequency density function. We can see that the density function is symmetric about μ. The mode u, i.e the value of x where the function is maximum, coincides with the mean μ. The frequency of both the exceedance and the non-exceedance of μ and u equals 0.5, or 50%. Therefore, the median g, i.e. the

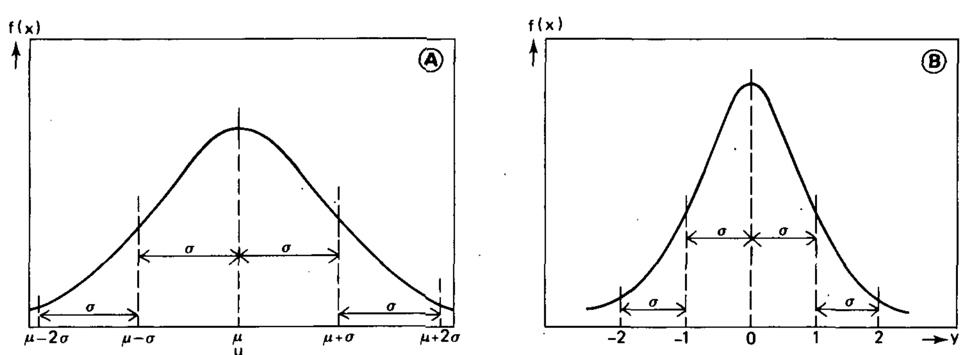

Figure 6.9 Normal distributions and some common properties. A: Normal frequency density function. B: Standard normal frequency density function

Table 6.6 Frequencies of exceedance f of the standard normal variate y (positive y values only)

y	f	y	f	y	f
0.0	.50	1.0	.16	2.0	.023
0.1	.46	1.1	.14	2.1	.018
0.2	.42	1.2	.12	2.2	.014
0.3	.38	1.3	.097	2.3	.011
0.4	.34	1.4	.081	2.4	.008
0.5	.31	1.5	.067	2.5	.006
0.6	.27	1.6	.055	2.6	.005
0.7	.24	1.7	.045	2.7	.004
0.8	.21	1.8	.036	2.8	.003
0.9	.18	1.9	.029	2.9	.002

value of x that indicates exactly 50% exceedance and non-exceedance, also coincides with the mean.

If $\mu = 0$ and $\sigma = 1$, the distribution is called a standard normal distribution (Figure 6.9B). Further, using the variate y instead of x to indicate that the distribution is a standard normal distribution, we see that the density function (Equation 6.14) changes to

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) \qquad (6.17)$$

Tables of frequencies f(y) can be found in statistical handbooks (e.g. Snedecor and Cochran 1986). If we use either of the transformations $x = \mu + \sigma y$ or $y = (x - \mu)/\sigma$ or, if we use the estimated values $\bar{x}$ for μ and s for σ (Equation 6.12 and 6.13)

$$x = \bar{x} + s.y \quad \text{or} \quad y = (x - \bar{x})/s \qquad (6.18)$$

we can use the tabulated standard distribution (e.g. Table 6.6) to find any other normal distribution.

The central limit theory states that, whatever the distribution of x, in a sample of size n the arithmetic mean ($\bar{x}_n$) of x will approach a normal distribution as n increases.

An annual rainfall, being the sum of 365 daily rainfalls x_i, equals $365\bar{x}_i$. Because n is large (365), annual rainfalls are usually normally distributed. The general effect of the duration on the shape of the frequency distribution is illustrated in Figure 6.10.

If there is a sample series available that is assumed to have a normal distribution, we can estimate μ and σ using Equations 6.12 and 6.13. The standard error $\sigma_{\bar{x}}$ of the arithmetic mean $\bar{x}$ of the sample is smaller than the standard deviation σ_x of the individual values of the distribution. So, for independent data $x_1, ..., x_n$, we obtain

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \qquad (6.19)$$

Hence, the estimated value $s_{\bar{x}}$ of $\sigma_{\bar{x}}$ equals

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \qquad (6.20)$$

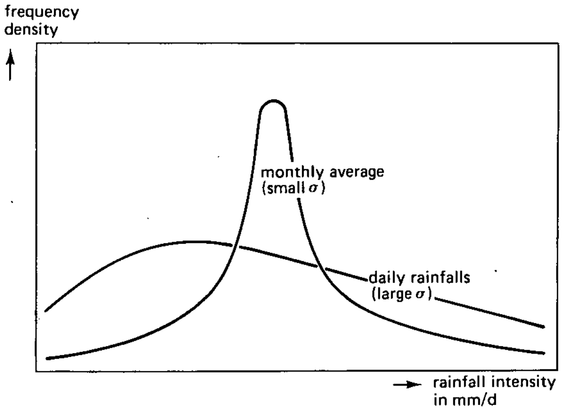

Figure 6.10 Monthly average rainfall intensities (in mm/d) have a narrower, more peaked, and more symmetrical frequency distribution than daily rainfalls

For example, the average monthly rainfall intensity, expressed in mm/d, has a standard deviation $\sqrt{30}$ times smaller than that of the individual daily rainfalls. In other words, the average monthly rainfall intensity has a frequency distribution that is narrower, but more highly peaked, than the average daily rainfall intensity (Figure 6.10).

If the distribution is skewed, i.e. asymmetrical, we can often work with the root-normal or with the log-normal distribution (B in Figure 6.11), simply by using $z = \sqrt{x}$ or $z = \log x$ and then by applying the principles of the normal distribution to z instead of to x. If, however, there are many observations with zero values (of which no logarithm can be taken), we should use a censored normal distribution without the small values of x (A in Figure 6.11).

Procedure and Example

For an idea of how to use the normal distribution, let us look at Figure 6.12, where the monthly totals of Table 6.4 have been plotted on normal probability paper. The probability axis has been constructed to present the cumulative normal distribution as a straight line. The parameters have been estimated from Table 6.4, according to

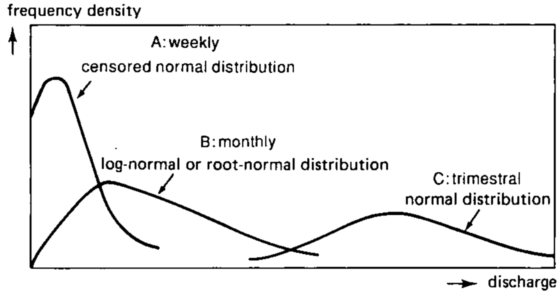

Figure 6.11 Frequency distributions of total discharge of different durations: weekly, monthly, and trimestral

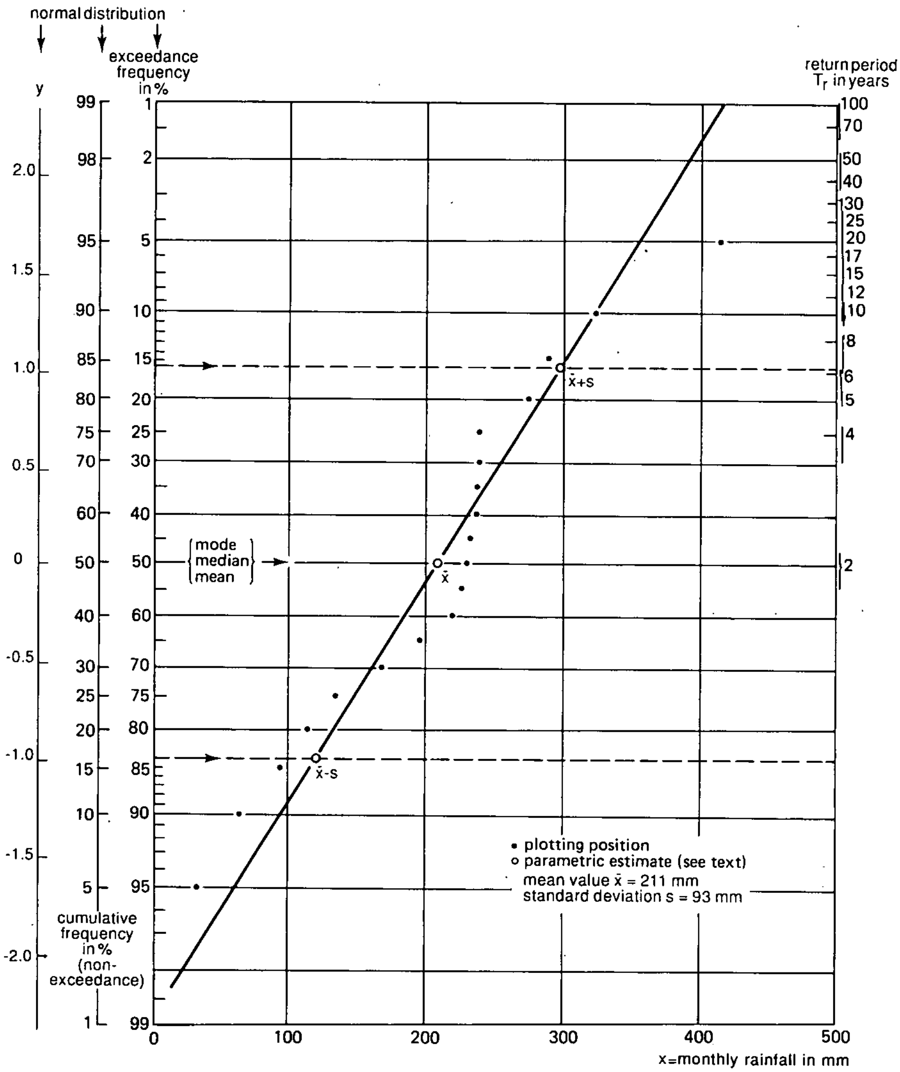

Figure 6.12 Monthly rainfalls, plotted on normal probability paper with a fitted line, based on the parametric method (derived from Table 6.4)

Equations 6.12 and 6.13, as

$$\bar{x} = \frac{\Sigma x}{n} = \frac{4017}{19} = 211$$

$$s = \sqrt{\frac{\Sigma x^2 - n(\bar{x})^2}{n-1}} = \sqrt{\frac{1003161 - 19 \times (211)^2}{18}} = 93$$

The value $x = \bar{x} + s = 211 + 93 = 304$ has a corresponding y value equal to 1 (Equation 6.18). Table 6.6 shows that this y value corresponds to a frequency of exceedance of f = 0.16, or 16%, from which it follows that the cumulative (non-exceedance) frequency is 0.84, or 84% (Equation 6.16). Because the normal distribution is symmetrical, we find that the value $x = \bar{x} - s = 211 - 93 = 118$ should correspond to a frequency of non-exceedance of 1 – 0.84 = 0.16, or 16%, and a y value equal to –1.

Accordingly, in Figure 6.12, the values x = 304 and x = 118 are plotted against the 84% and 16% non-exceedance (cumulative) frequencies. The mean value $\bar{x} = 211$ (for which y = 0, as in Equation 6.18) can be plotted against the 50% cumulative frequency (Table 6.6). A straight line can be drawn through the above three points.

We can conclude that the estimated return period of the observed monthly rainfall total of 422 mm is approximately 100 years instead of the 20 years we find in Table 6.4. There is, however, a 10% chance that the return period of this rainfall is smaller than 7 years or greater than 5000 years (Figure 6.3). This will be discussed further in Section 6.4.6.

The Log-Normal Distribution

An example of the application of the log-normal distribution is given in Figure 6.13. The data used here are derived from Table 6.5, which shows monthly maximum one-day rainfalls.

Because we can expect the maximum 1-day rainfalls to follow a skewed distribution, we are using the log-values (z = log x) of the rainfall instead of the real values (x), the assumption being that this transformation will make the frequency distribution symmetrical.

The procedure for normal distribution fitting is now exactly the same as before. So with the data from Table 6.5, we can calculate that $\bar{z} = 1.75$ and that s = 0.29, meaning that, if we plot the value $z = \bar{z} + s = 2.04$ against the 16% (y = 1) exceedance frequency and the value $z = \bar{z} - s = 1.46$ against the 84% (y = –1) exceedance frequency, we can draw a straight line through these points, as shown in the figure.

The figure also shows that a rainfall of 200 mm, for which z = log 200 = 2.30, has a return period of about 30 years, whereas in Table 6.5 this return period is about 20 years.

In addition to the log-normal distribution, the figure shows a confidence belt that was constructed according to the principles of confidence analysis. From this belt, we can see that a rainfall of 100 mm (point A) has a 90% confidence interval, ranging from 70 mm (point B) to 180 mm (point C). The return period of this rainfall (5 years) has a confidence interval that ranges from 2.5 years (point D) to almost 15 years (point E).

We shall interpret the data in Figure 6.13 further in Section 6.4.6.

6.4.4 The Gumbel Distribution

The Gumbel distribution (Gumbel 1954), also called the Fisher-Tippett Type I distribution of extreme values, can be written as a cumulative frequency distribution

$$F(x_N < x_r) = \exp\{-\exp(-y)\} \qquad (6.21)$$

where

x_N = the maximum x from a sample of size N

x_r = a reference value of x_N

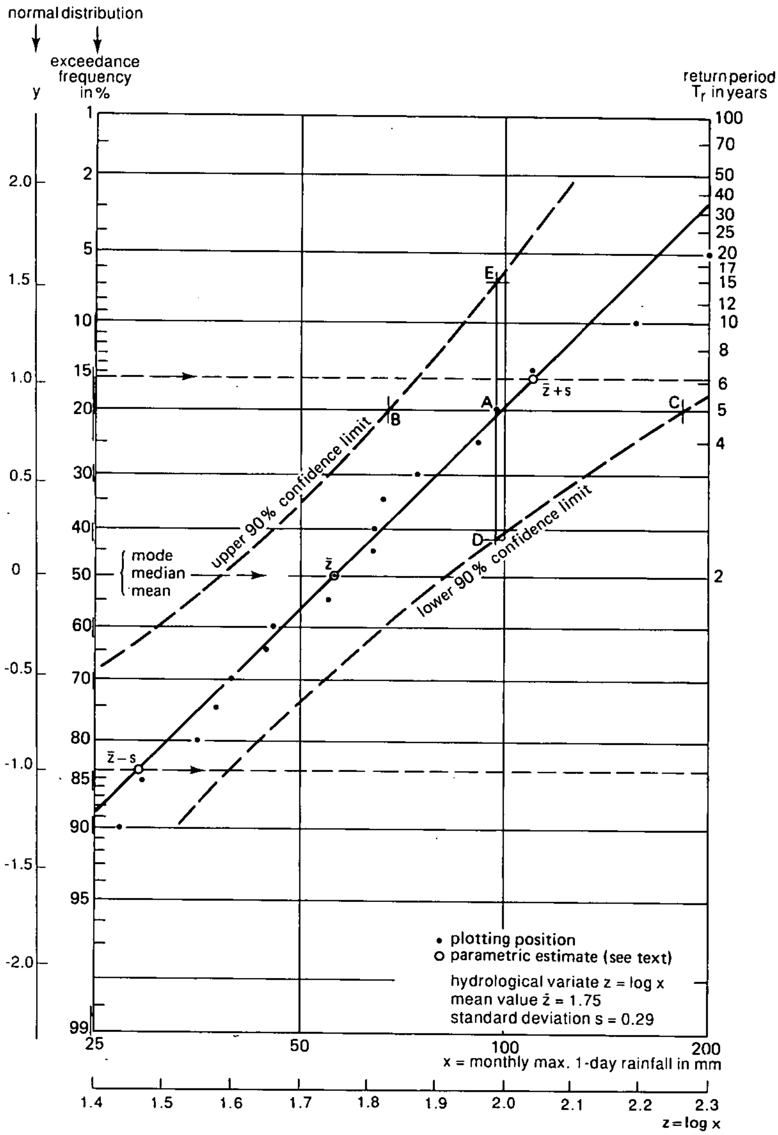

Figure 6.13 The log values of the data of Table 6.5, plotted on normal probability paper with a fitted line, based on the parametric method

y $= \alpha(x_r - u)$, the reduced Gumbel variate

u $= \mu - c/\alpha$, the mode of the Gumbel distribution

μ = mean of the Gumbel distribution

c = Euler's constant = 0.577

$$\alpha = \frac{\pi}{\sigma\sqrt{6}}$$

σ = the standard deviation of the Gumbel distribution.

By estimating μ and σ (Equations 6.12 and 6.13), we can determine the entire frequency distribution.

The Gumbel distribution is skewed to the right, with $u < \mu$ and the median g in between. The introduction of $x_r = u$ into the equation for the Gumbel distribution yields

$$F(x_N < u) = e^{-1} = 0.37 \tag{6.22}$$

Therefore the probability of non-exceedance of the mode u equals 0.37, or 37%, and the probability of exceedance is $1 - 0.37 = 0.63$, or 63%.

The cumulative probability distribution of the maximum value in a sample of size N, drawn from an exponential distribution will asymptotically approach the Gumbel distribution as N increases. Hydrologists assume that this asymptotic approach occurs when $N > 10$, and so they frequently use the Gumbel distribution to find annual or monthly maxima of floods or to find rainfalls of short duration (less than 1/10 of a year or of a month).

To determine the Gumbel distribution, we need several (n) samples of size N (total $n \times N$ data) from which to select the n maxima. In this way, annual, monthly, or seasonal maximum series can be composed for various durations (each duration containing at least $N = 10$ independent data from which to choose the maximum).

Taking natural logarithms twice, we can write the Gumbel distribution as

$$y = \alpha(x_r - u) = -\ln\{-\ln F(x_N < x_r)\} \tag{6.23}$$

Gumbel probability paper is constructed to allow plotting of cumulative frequencies on a $-\ln(-\ln)$ scale, which yields a linear relationship with x_N. A straight line of best fit can thus be drawn or calculated by regression analysis.

Procedure and Example

For this example, the monthly maximum 1-day rainfalls presented in Table 6.5 are used. Figure 6.14 shows the cumulative frequencies plotted on Gumbel probability paper and a straight line calculated from Equation 6.23. As estimates of μ and σ, we get

$$\bar{x} = \Sigma x/n = 1317/19 = 69 \text{ (Equation 6.12)}$$

$$s^2 = \frac{1}{n-1}(\Sigma x^2 - n\bar{x}^2) = \frac{1}{18}(130615 - 19 \times 69^2) = 2231 \text{ (Equation 6.13)}$$

$$s = \sqrt{2231} = 47$$

so that, according to Equation 6.21

$$\alpha = \pi/s\sqrt{6} = \pi/47\sqrt{6} = 0.027$$

$$u = \bar{x} - c/\alpha = 69 - 0.577/0.027 = 48$$

Substitution of the above estimates into the equation $y = \alpha(x_r - u)$ gives $y = 0.027(x_r - 48)$. This is the expression of a straight line on Gumbel probability paper (Equation 6.23). Determination of two arbitrary points gives

$$y = 0 \rightarrow x_r = u = 48 \text{ mm, and } F(x_N < 48) = 0.37$$

$$y = 3 \rightarrow x_r = 159 \text{ mm, and } F(x_N < 159) = 0.95$$

The plotting of these two points produces the straight line that characterizes the Gumbel distribution (Figure 6.14).

6.4.5 The Exponential Distribution

The exponential distribution can be written as an exceedance frequency distribution

$$F(x > x_r) = \exp\{-\lambda(x_r - a)\} \tag{6.24}$$

where

x_r = a reference value of x
a = the minimum value of x_r
$\lambda = 1/(\mu - a)$
μ = the mean of the distribution.

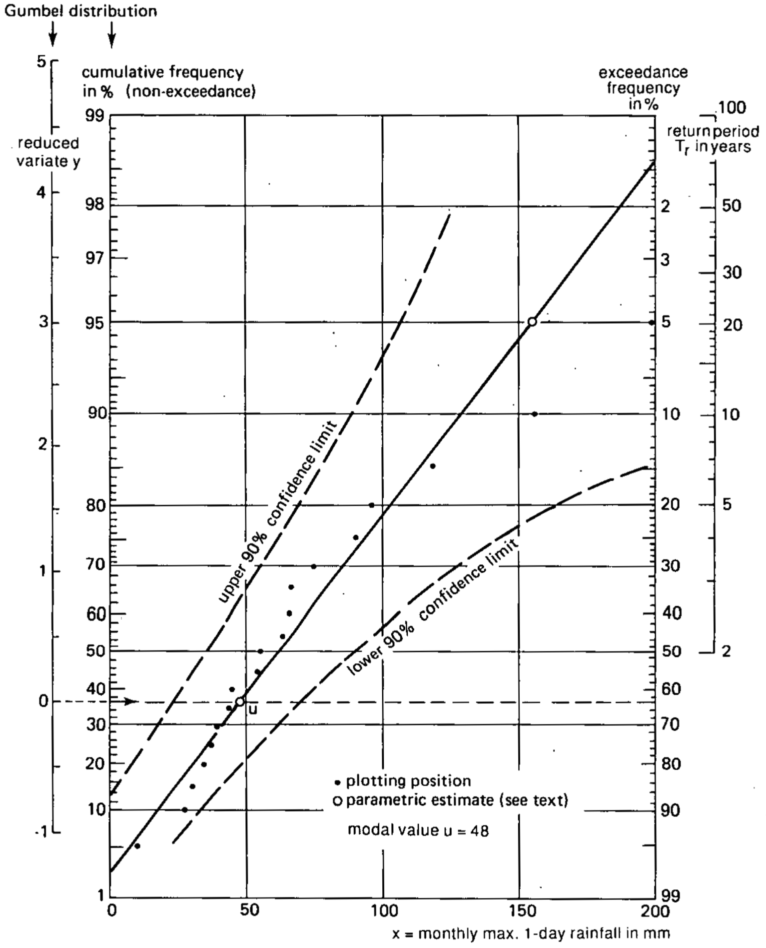

Figure 6.14 The data of Table 6.5, plotted on Gumbel probability paper with a fitted line, based on the parametric method. The 90% confidence limits are shown

Theoretically, the value of the standard deviation equals $\sigma = \mu = a + 1/\lambda$. Hence we need to know either σ or μ. Contrary to the normal distribution and the Gumbel distribution, both of which have two parameters, the exponential distribution has only one parameter if a is known.

For $x_r = \mu$, we find from Equation 6.24 that $F(x > \mu) = e^{-1} = 0.37$. Hence, the mean and the median are not equal, and the distribution is skewed.

The exponential distribution can be used for maxima selected from certain series of data, just as we saw for the Gumbel distribution in the previous section, or it can be used for selected values that surpass a certain minimum value (censored series).

Equation 6.24 can also be written as

$$\lambda (x_r - a) = -\ln \{F(x > x_r)\} \tag{6.25}$$

meaning that a plot of x_r or $(x_r - a)$ versus $-\ln \{F(x > x_r)\}$ will yield a straight line. For $x_r = \mu$, we find from Equation 6.25 that $-\ln F(x > \mu) = 1$.

Procedure and Example

Let us apply an exponential distribution to the maximum monthly 1-day rainfalls given in Table 6.5. The estimate of μ is

$$\bar{x} = \Sigma x/n = 1317/19 = 69 \text{ mm (Equation 6.12)}$$

Using a = 10 (the lowest maximum rainfall recorded), we find that $\lambda = 1/(\mu\text{-}a) = 1/(69\text{-}10) = 0.017$. The exponential distribution for the data of Table 6.5 is now expressed as

$$F(x > x_r) = \exp \{-0.017 (x_r - 10)\}$$

Using $x_r = 150$ mm and $x_r = 75$ mm, we find that $F(x > 150) = 0.09$ and that $F(x > 75) = 0.33$, so that $\ln F(x > 150) = -2.4$ and $\ln F(x > 75) = -1.1$. On the basis of the linearity shown in Equation 6.25, these points can be plotted and connected by a straight line (Figure 6.15).

Note that the baseline used by Benson (Figure 6.2) stems from an exponential distribution. The line can be described by the equation $x = \alpha \ln T_r + a$, where α and a are constants. This equation can also be written as λ (x-a) = $\ln T_r$ where $\lambda = 1/\alpha$. Because, according to Equation 6.9, $T_r = 1/F(x > x_r)$, and, therefore, $\ln T_r = -\ln\{F(x > x_r)\}$, the baseline can also be expressed as $\lambda(x\text{-}a) = -\ln\{Fx > x_r)\}$, which is identical to the expression of the exponential distribution given by Equation 6.25.

The Log-Exponential Distribution

Figure 6.16 shows a depth-return period relation of 1-day rainfalls. These rainfalls were derived from the a_i values of Table 6.2 and plotted on double-logarithmic paper. The line represents a log-exponential distribution, for which the rainfall x was transformed into $z = \log x$ in the same way the log-normal distribution was transformed previously.

The straight line can be expressed as

$$\ln T_r = \alpha + \lambda \log x_r$$

where α and λ are constants, and x_r is a certain value of the rainfall x. With $T_r = 1/F(x > x_r)$, this equation changes to

$$-\ln\{F(x > x_r)\} = \lambda(\log x_r + \alpha/\lambda)$$

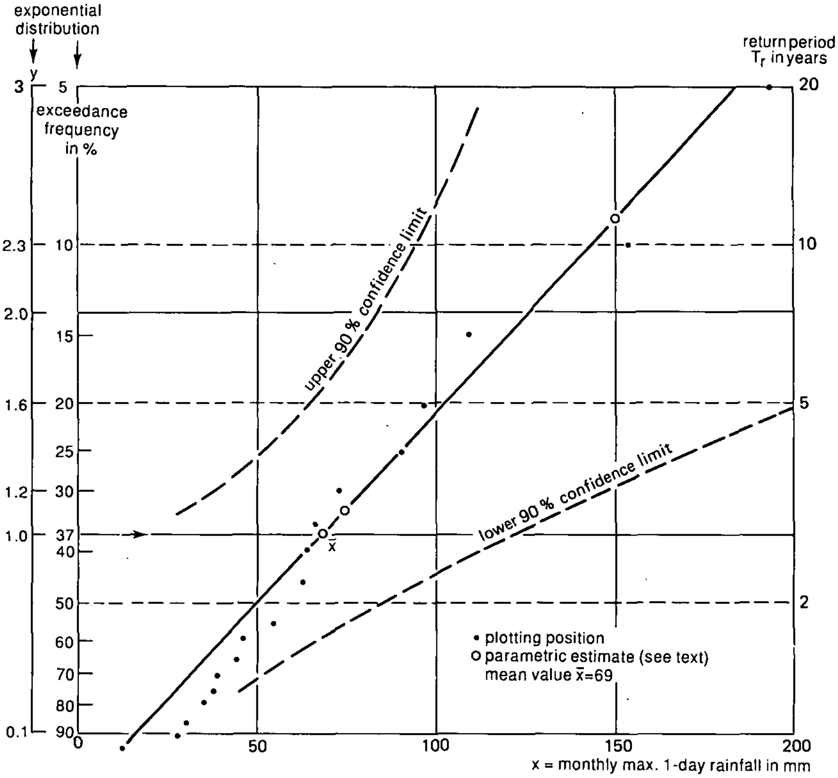

Figure 6.15 The data of Table 6.5, plotted on linear graph paper against the natural logarithm of the exceedance frequency to obtain an exponential frequency distribution. The fitted line, based on the parametric method, and the 90% confidence limits are shown

If we compare this with Equation 6.25, we see that $a = -\alpha/\lambda$, and that the only difference remaining is the presence of $\log x_r$ instead of x_r.

This means that, if the data conform, the log-exponential distribution can be used instead of the exponential distribution. The best fit to the data will determine which distribution to use.

6.4.6 A Comparison of the Distributions

The monthly maximum 1-day rainfalls from Table 6.5 were used to derive the log-normal, the Gumbel, and the exponential frequency distributions, along with their confidence intervals (Figures 6.13, 6.14, and 6.15). We can see that the data, all of which were plotted with the ranking method, do not lie on the straight lines calculated with the parametric estimates of the frequency distributions. Nevertheless, they are fully within the confidence belts. Hence, in this case at least, it is difficult to say whether there is a significant difference between the ranking procedure and the parametric method.

The figures show that the relatively small scatter of the plotting positions around

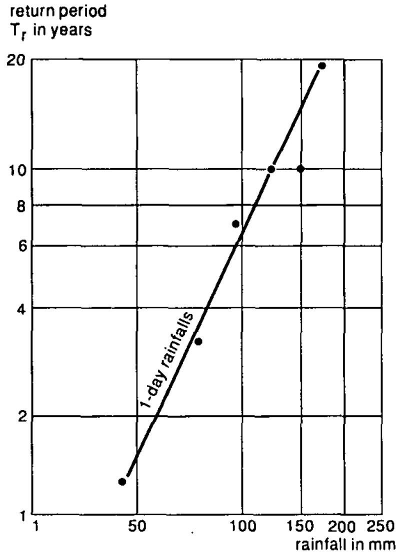

Figure 6.16 Depth-return period relation of 1-day rainfalls derived from Table 6.2, plotted on double-logarithmic paper

the straight line is no measure of the width of the confidence belt. This is owing to the artificially high correlation that the ranking method introduces between the data and the frequencies.

Table 6.7 shows the different return periods of the 200 mm rainfalls estimated with the different theoretical distributions, including the 90% confidence interval (Figures 6.13 and 6.14, and 6.15).

The table shows that the different distributions give different return periods for the same rainfall. Owing to the limited number of available data (19), the confidence intervals are very wide, and the predicted return periods are all well inside all the confidence intervals. Hence, the differences in return period are not significant, and one distribution is no better or worse than the other.

We can prepare a table of confidence limits not only for the return period of a certain rainfall (Table 6.7), but also for the rainfall with a certain return period (Table 6.8). This can be done, however, only after a graphical or a theoretical relation between rainfall and frequency has been established.

Table 6.7 Estimates of the return periods (in years) and the confidence intervals of the 200-mm rainfall of Table 6.5, as calculated according to 4 different methods

Estimation method	Return period (T_r)	90% confidence interval of T_r	
		Lower limit	Upper limit
Ranking method	20	5	400
Log-normal distribution	30	6	500
Exponential distribution	25	5.5	400
Gumbel distribution	60	7	5000