The Abdus Salam
International Centre
for Theoretical Physics

2444-10

College on Soil Physics - 30th Anniversary (1983-2013)

25 February - 1 March, 2013

Frequency analysis of rainfall data

RAES Dirk
Katholieke Universiteit Leuven
Department of Earth and Environmental Sciences
Celestijnenlaan 200E, BE-3001 Leuven
BELGIUM

Frequency analysis of rainfall data

Draft version - May 2004

Dirk RAES
K.U.Leuven
Inter-University Programme in Water Resources Engineering (IUPWARE)
Leuven, Belgium.

This note outlines a statistically technique to estimate rainfall depths that can be expected for selected probabilities or return periods. With the help of a frequency analysis on historical rainfall data, the magnitude of the rainfall depths can be estimated. The estimates are required for the design and management of irrigation and drainage projects. The technique is also useful for analysing time series of other agro-meteorological or hydrological variables.

The text of this note is based on reference statistic textbooks (Haan, 1986; Snedecor and Cochran, 1980) and notes from the World Meteorological Organization (WMO, 1981, 1983, 1990). To demonstrate the frequency analysis on rainfall data, multiple examples are worked out on time series of rainfall data extracted from the FAOCLIM databank (FAO, 2000).

Frequency analysis requires considerable computations and careful plotting. Efficiency can be gained by using software such as RAINBOW. This software has been specially designed to carry out frequency analyses and to test the homogeneity of data sets. RAINBOW is freely available and an installation file and reference manual can be downloaded from the web (go to www.iupware.be, and select downloads and next software).

Table of contents

Frequency analysis of rainfall data

1. Rainfall variability in time

The total rainfall received in a given period at a location is highly variable from one year to another. The variability depends on the type of climate and the length of the considered period. In general it can be stated that the drier the climate, the higher the variability of rainfall in time. The same hold for the length of the period: the shorter the period the higher the annual variability of rainfall in that period.

In Figure 1 the total annual rainfall observed in Tunis (Tunisia) is plotted for the 1930 – 1990 period. Annual rainfall varied between 220 mm (year 1943) and 912 mm (year 1953). The average and standard deviation of the total rainfall for this 60-year period are respectively 466 mm and 131 mm. The average gives information on the normal amount of rainfall one can expect in the area. It can be used to obtain an idea of the departure of the annual rainfall from the normal, or to compare climatic regions.

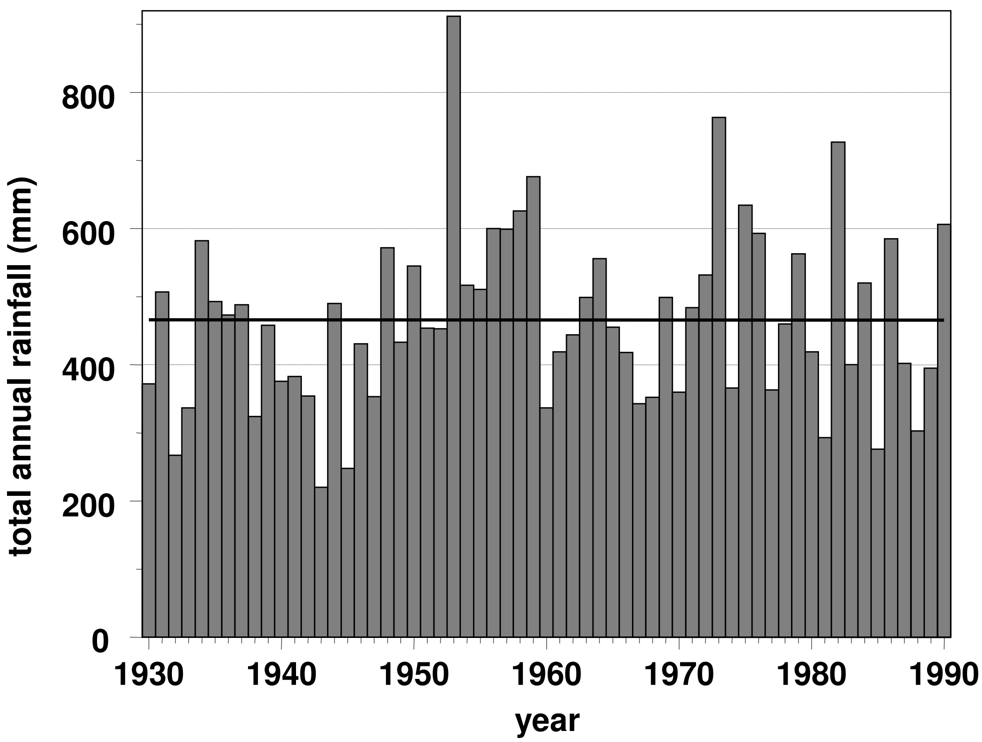


Figure 1. Total annual rainfall recorded in Tunis (Tunisia) for the period 1930-1990, with indication of the average rainfall (horizontal line) (source FAOCLIM, 2000).

Because of the strong variability of rainfall in time, the design and management of irrigation water supply and flood control systems are not based on the long-term average of rainfall records but on particular rainfall depths that can be expected for a specific probability or return period. These rainfall depths can only be obtained by a thoroughly analysis of long time series of historic rainfall data.

Although time series of historic rainfall data are characterized by their average and standard variation, these values cannot be blindly used to estimate design rainfall depths that can be expected with a specific probability or return period. Applying this technique to a data set can produce misleading results since the actual characteristics of the distribution are ignored and it is assumed that they follow a particular distribution. To avoid this type of error, it is essential that the goodness of the assumed distribution be checked before design rainfall depths are estimated.

2. Probability of exceedance and return period

2.1 Rainfall depths expected for specific probability (X_P)

Estimates of rainfall depths (X_P) or intensities that can be expected for a specific probability during a specific reference period (hour, day, week, 10-day, month, year) are required for the management and design of irrigation and drainage projects. In this note the probability refers to the probability of exceedance and it specifies the likelihood that the actual rainfall during that period will be equal to or higher than the estimated rainfall depth X_P. Since the rainfall depth X_P is the amount of rain that can be expected or might be exceeded in a given period for a specific probability, it refers to the minimum amount of rain one can rely on during the reference period, and therefore is often denoted as '*dependable rainfall*' in irrigation sciences.

2.2 Probability of exceedance (P_X)

The probability of exceedance refers to the probability of the occurrence of a rainfall depth greater than some given value X_P. The probability of exceedance (P_X) is expressed as a fraction (on a scale ranging from zero to one) or as a percentage chance with a scale ranging from 0 to 100 percent. If the estimates refer to the rainfall depth that can be expected or might be exceeded in a year during the reference period than it can be expressed as a set number of years out of a total number of years.

If 4.8 mm is the 10-day rainfall depth that can be expected on average in 8 years out of 10 during the first decade of January in Tunis, than is 4.8 mm the 80% dependable annual rainfall for that decade. As the rainfall amount increases, its probability of exceedance decreases. For Tunis, 12.2 mm is the 50% and 23.0 mm is the 20% dependable rainfall for the first decade of January.

2.3 Return period (T_X)

The return period (also called the recurrence interval) T_X is the period expressed in number of years in which the annual observation is expected to return. It is the reciprocal value of the probability when expressed as a fraction:

$$T_X = \frac{1}{P_X}$$

The above 20% dependable rainfall (P_X = 0.20) has a return period of (1/(0.20) =) 5 years. Or on average once every 5 years the rainfall in the 1st decade of January will be larger than 23 mm in Tunis. The 50% dependable rainfall has a return period of 2 years, or on average once every 2 years the rainfall depth is exceeded.

2.4 Probability of exceedance for design purposes

The selection of the probability of exceedance (P_X) or return period (T_X) for design purposes is related to the damage the excess or the shortage of rainfall may cause, the risk one wants to accept and the lifetime of the project. The following examples illustrate the selection of the probability:

- The design of an open shallow drainage system in sloping land in Western Europe is based on a design storm (often a 6-hour storm) that has a return period of 10 to 25 years for agricultural land and 30 to 100 years for urbanized basins (Smedema and Rycroft, 1983). The design norm determines the reliability of the system in terms of flooding. The design storm is the 6-hour rainfall that has a probability of exceedance of 4 to 10 % for agricultural land and only 1 to 3 % for urbanized basins.
- On the other hand, low dependable rainfall depths are used when designing irrigation systems. Rainfall depths that are exceeded in 3 out of 4 years or 4 out of 5 years during the peak period are often selected as design rainfall (Doorenbos and Pruitt, 1977). The corresponding probabilities of exceedance are high and respectively 75% and 80%. A higher level of dependable rainfall, say 9 years out of 10, may even be selected for high value crops. Selecting the 90% dependable rainfall as the design norm will result in a lower risk but in bigger canals or larger pipes.

The design rainfall is thus and economic parameter and its selection essential involves the optimisation of the expected benefits in relation to the costs.

2.5 Probabilities of exceedance for management purposes

Information on the rainfall depth that can be expected in a specific period under various weather conditions is required for management and planning purposes. FAO (Smith, 1992) uses the following rules for the determination of dry, normal and humid weather conditions:

- The weather condition in a period is called dry if the rainfall received during that period will be exceeded 4 out of 5 years, i.e. having a probability of exceedance of 80%;
- The rainfall in a period is normal, if the rainfall received during that period will be exceeded in 1 out of 2 years. The probability of exceedance is equal to 50%;
- The weather condition in a period is called humid if the rainfall received during that period is exceeded 1 out of 5 years, i.e. having a probability of exceedance of 20%.

The rainfall depths for Tunis for each of the 36 decades that can be expected with a 80, 50 and 20% probability of exceedance are plotted in Figure 2. The rainfall depths correspond with the amount of rain that can be expected in each of the decades when the weather conditions are respectively dry, normal and humid.

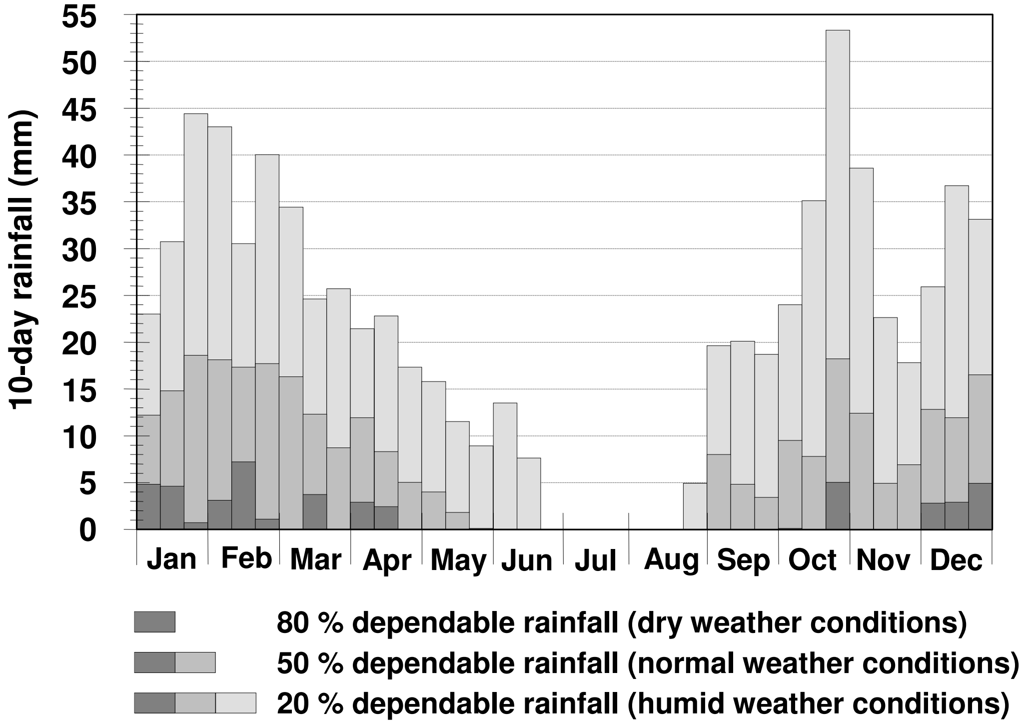


Figure 2. Rainfall depths for Tunis expected in a decade for dry, normal and humid weather conditions

3. Time series of rainfall data

In a frequency analysis, estimates of the probability of occurrence of future events are based on the analysis of historical rainfall records. By assuming that the past and future data sets are stationary and have no apparent trend one may expect that future time series will reveal frequency distributions similar to the observed one. It is obvious that the longer the data series the more similar the frequency distribution will be to the probability distribution. In short series accurate determination for rainfall depths that can be expected for selected return periods is not possible. Estimates of dependable rainfall are less reliable if the corresponding return period exceeds the observation period. As the number of observations increases, the error in determining expected rainfall gradually diminishes. Although the required length of the time series depends on the magnitude of variability of the precipitation climate, a period of 30 years and over normally is thought to be very satisfactory. However, if interest lies in extreme rainfall events, larger number of years will be required.

The basis of all statistical analysis is that the data series must be composed of random variables selected from a single population usually infinite in extent. Therefore, rainfall records from a climatic station should be carefully checked for homogeneity, or in other words, it should be checked if the collected observations are from the same population. If there is an apparent trend in time, the observation period should be restricted to the period where the data is homogeneous and representative for the planned use. A test for homogeneity of time series is presented in Section 5 of this note and an example of a long rainfall record where two, statically significant, distinct periods can be distinguished is presented in Section 6.4 (Detecting shifts in the total annual rainfall).

When hydrological data sets are analysed an extreme event might be present that appear to be from a different population. In probability plots such an outlier will plot far off the line defined by the other points. Such an outlier has to be excluded from the analysis unless some historical information is available about its probability of exceedance.

4. Frequency analysis and probability plotting

From a frequency analysis, the estimates of rainfall depths for selected probabilities or return periods required for the design can be obtained. The analysis consists in collecting historical data over a sufficient number of years and subsequently:

- ranking the data and assigning plotting positions by estimating the probability of exceedance with one or another method;
- selecting a distributional assumption and plotting the data in a probability plot;
- verifying the goodness of the selected distribution. If unsatisfactory another distribution should be selected or the data should be transformed so that the transformed data follow the selected distribution;
- determining rainfall depths (X_P) that can be expected for selected probabilities or return period from the probability plot.

4.1 Estimation of the probability of exceedance

The first step in the frequency analysis is the ranking of the rainfall data. After the rainfall data are ranked, a serial rank number (r) ranging from 1 to n (number of observations) is assigned. Subsequently the probability have to be determined that should be assigned to each of the rainfall depths. If the data are ranked in descending order, the highest value first and the lowest value last, the probability is an estimate of the probability that the corresponding rainfall depth will be exceeded. When data are ranked from the lowest to the highest value, the probability refers to the probability of non-exceedance. Hence the probabilities are estimates of cumulative probabilities. They are formed by summing the probabilities of occurrence of all events greater then (probability of exceedance) or less than (probability of non-exceedance) some given rainfall depth. Since these probabilities are unknown the probabilities of exceedance have to be estimated by one or another method. Several methods are listed in Table 1. The Weibull, Sevruk and Geiger, and the Gringorten methods are theoretically better sound. The probabilities will be the plotting positions of the ranked rainfall data in the probability plot.

Table 1. Methods for estimating probabilities of exceedance or non-exceedance of ranked data, where r is the rank number and n the number of observations.

Method	**Estimate of probability of exceedance or non-exceedance (%)**
California (California State Department, 1923)	$\frac{r}{n}\ 100$
Hazen (Hazen, 1930)	$\frac{(r-0.5)}{n}\ 100$
Weibull (Weibul, 1939)	$\frac{r}{(n+1)}\ 100$
Gringorten (WMO, 1983)	$\frac{(r-0.44)}{(n+0.12)}\ 100$
Sevruk and Geiger (Sevruk and Geiger, 1981)	$\frac{(r-3/8)}{(n+1/4)}100$

To demonstrate the calculation procedure, the annual total rainfall for the 37-year period 1960 - 1996 for Bombay (India) is extracted from the FAOCLIM database (FAO, 2000). The rainfall data are listed in Annex 1 (Table A1.1). Since data for 4 years are missing there are only 33 observations (n = 33) in the time series. The 33 annual total rainfall depths were subsequently ranked from high to low and the corresponding probabilities of exceedance were estimated with various methods (Table 2). It can be noted that all of the relationships give similar values near the centre of the distribution but may vary somewhat in the tails.

Table 2. Probabilities of exceedance of the ranked annual rainfall estimated with various methods.

Ranked rainfall X (mm)	Rank number r	Estimate of probability of exceedance P_X			
		Hazen (%)	Weibull (%)	Gringorten (%)	Sevruk & Geiger (%)
3107	1	1.5	2.9	1.7	1.9
2791	2	4.5	5.9	4.7	4.9
2626	3	7.6	8.8	7.7	7.9
2620	4	10.6	11.8	10.7	10.9
2469	5	13.6	14.7	13.8	13.9
2444	6	16.7	17.6	16.8	16.9
2441	7	19.7	20.6	19.8	19.9
2423	8	22.7	23.5	22.8	22.9
2382	9	25.8	26.5	25.8	25.9
2373	10	28.8	29.4	28.9	28.9
2370	11	31.8	32.4	31.9	32.0
2316	12	34.8	35.3	34.9	35.0
2205	13	37.9	38.2	37.9	38.0
2141	14	40.9	41.2	40.9	41.0
2119	15	43.9	44.1	44.0	44.0
2031	16	47.0	47.1	47.0	47.0
2029	17	50.0	50.0	50.0	50.0
2026	18	53.0	52.9	53.0	53.0
1964	19	56.1	55.9	56.0	56.0
1937	20	59.1	58.8	59.1	59.0
1923	21	62.1	61.8	62.1	62.0
1900	22	65.2	64.7	65.1	65.0
1882	23	68.2	67.6	68.1	68.0
1850	24	71.2	70.6	71.1	71.1
1835	25	74.2	73.5	74.2	74.1
1832	26	77.3	76.5	77.2	77.1
1629	27	80.3	79.4	80.2	80.1
1560	28	83.3	82.4	83.2	83.1
1553	29	86.4	85.3	86.2	86.1
1540	30	89.4	88.2	89.3	89.1
1511	31	92.4	91.2	92.3	92.1
1287	32	95.5	94.1	95.3	95.1
960	33	98.5	97.1	98.3	98.1
Total number of observations: n = 33					
Mean annual rainfall: $\overline{X}$ **= 2063 mm** **Standard deviation: s = 453.2 mm**					

4.2 Probability plot

A probability plot is a plot of the rainfall depths versus their probabilities of exceedance as determined by one or another method (Figure 3). When the data are plotted on arithmetic paper, where both axes have a linear scale, the data are not likely to be on a straight line but to follow a S-shaped curve. This plot is sometimes called a percentage ogive.

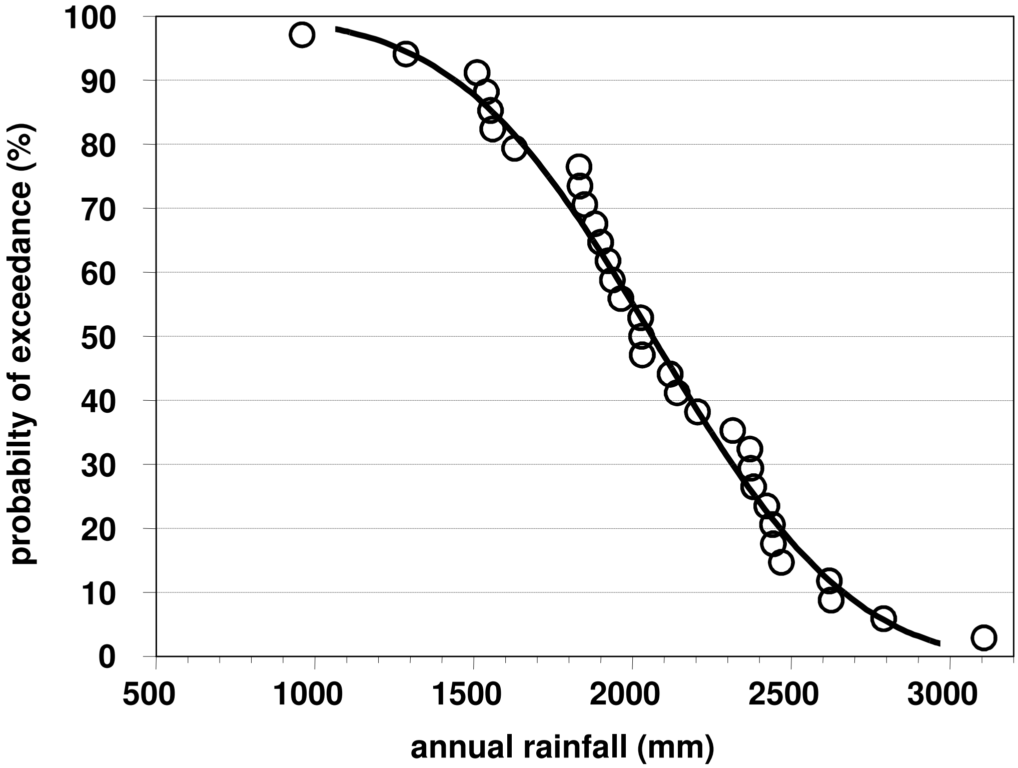


Figure 3. Probability plot of the total annual rainfall for Bombay by using linear scales for both axes.

By selecting a probability distribution, the vertical axis of the probability plot is rescaled so that the data will fall on a straight line if it is distributed as selected (Figure 4). On probability paper the cumulative distribution of the total population will fall on that straight line. This makes the verification of the goodness of selected distribution easier. Figure 4 refers to a normal distribution, but the same is true for other distributions. Only the rescaling of the vertical axis will be different.

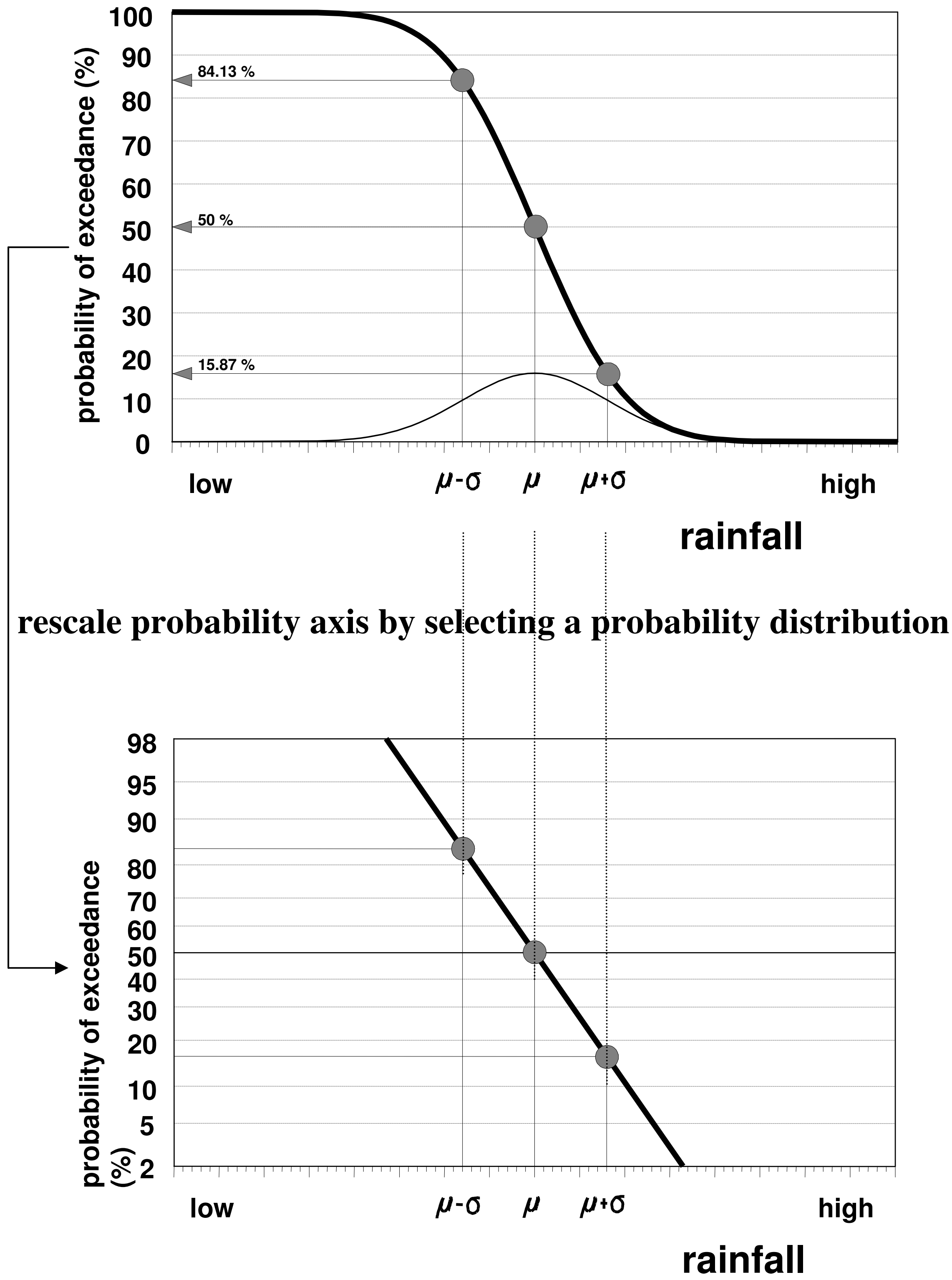


Figure 4. Effect of the rescaling of the vertical axis of a probability plot

If the data in a probability plot fall in a reasonable alignment, it may be assumed that the data can be approximated by the assumed distribution. Since the data is only a sample of the total population it would be rare for a set of data to plot exactly on a line and a decision must be made as to whether or not the deviations from the line are random variations or represent true deviations indicating the data does not follow the given probability distribution. By fitting a line through the points an indication of the goodness of fit is given by the coefficient of determination (R^2) of the fitted line. Owing to sampling variations, the points will depart somewhat from the line even with data that follow perfectly the assumed distribution.

When probability plots are made and a line drawn through the data, the tendency to extrapolate the data to extreme probabilities of exceedance is great. However, if the data do not truly follow the assumed distribution, the error in extrapolation can be quite large. This is the result of the presence of a large number of data points in the central part of the distribution that strongly determines the slope of the line. The few data points in the tail of the distribution hardly affect the slope of the line (see for example Figure 5).

Probability plotting is an excellent graphical technique for testing distributional assumptions. When the points in the probability plot do not fall in a reasonable alignment, the data is most likely not distributed as the selected distribution, especially if the points deviate from the straight line in some systematic matter. The extent and types of departures from the assumed distribution might give information on a more appropriate distribution. One can either rescale the abscissa of the probability plot (See normalizing the data, Section 4.3.3) or select another distribution.

4.3 Normal distribution

Hydrologic random variables, which are the sum of a large number of independent effects, are often normally distributed. The yearly rainfall, being the sum of many individual rainfall events, is a typical example. It has been found that 80% of the stations with long records receive rainfall according to a normal distribution.

4.3.1 Probability plot

By plotting the data on specially designed normal probability paper (Annex 3), the assumption that the data is normal distributed can be easily checked. In normal probability paper the probability scale (i.e. the vertical axis) is stretched at both ends and compressed in the centre. The distortion is such that in the probability plot the points should plot perfectly on a straight line if the data is normal distributed (Figure 4). Or the typical S shaped curve on arithmetic paper is stretched to a straight line on normal probability paper.

In Figure 5, the annual rainfall of Bombay is plotted versus its probability of exceedance (estimated with the Weibull formula) on normal probability paper. Since the data in the probability plot fall in a reasonable alignment, it may be assumed that the annual total rainfall in Bombay is normal distributed.

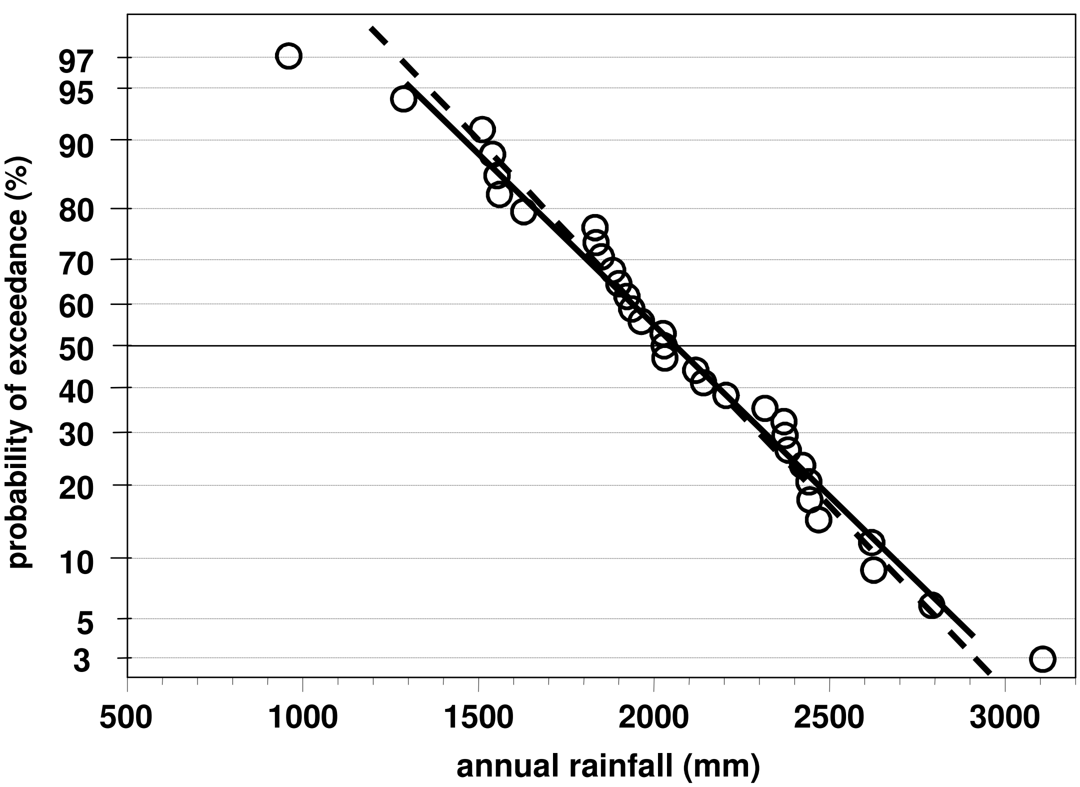


Figure 5. Probability plot on normal probability paper of the total annual rainfall for Bombay with indication of the best fitted line (full line) and the normal line (dotted line)

4.3.2 Estimating rainfall amounts for selected probabilities

Graphical solution:
By fitting a straight line through the points, the rainfall corresponding to various probabilities of exceedance are easily derived from the probability plot. The goodness of fit can be evaluated by the coefficient of determination (R^2). Estimates of the annual total rainfall X_P for Bombay for selected probabilities of exceedance derived from the fitted line ($R^2 = 0.98$) are presented in column 3 of Table 3. The return periods T_X specified in column 2 are the reciprocal of the probabilities of exceedance P_X and hence refer to rainfall depths greater than the estimated depths in column 3. In column 4 probabilities of non-exceedance (i.e. 100 – P_X) are presented. The return periods corresponding with the probabilities of non-exceedance are specified in column 5. They refer to occurrence of rainfall depths less than the estimated depths X_P in column 3.

Table 3. Estimated annual rainfall for Bombay for selected probabilities and return periods derived from the fitted line in the probability plot (Figure 5).

Probability of exceedance P_X (%)	Return Period T_X (years)	Estimated annual rainfall X_P (mm)	Probability of non-exceedance 100-P_X (%)	Return Period (years)
10	10	2686	90	1.11
20	5	2472	80	1.25
30	3.33	2318	70	1.43
40	2.50	2186	60	1.67
50	2	2063	50	2
60	1.67	1940	40	2.50
70	1.43	1808	30	3.33
80	1.25	1654	20	5
90	1.11	1440	10	10

Numerical solution:
When annual rainfall is completely normal distributed the data in a probability plot will fall perfectly on the normal line. On this line the mean rainfall ($\overline{X}$) corresponds with the 50 percent probability of exceedance, the $\overline{X}$ + s (standard deviation) corresponds with 15.87 % and the $\overline{X}$ -s with the 84.13 % probability of exceedance (Figure 4). Since the normal distribution is completely characterized by its average and standard deviation, they can be used to estimate rainfall for selected probabilities or return periods:

$$X_P = \overline{X} \pm k\ s$$

where X_P is the rainfall depth having a specific probability of exceedance, $\overline{X}$ is the sample mean, s the standard deviation and k a frequency factor. The sign and magnitude of the frequency factor vary according to the selected probability of exceedance (Annex 2, Table A2.1).

For the example of the annual total rainfall in Bombay, the sample mean is 2063 mm and the standard deviation 453.2 mm. As can be observed in Figure 5, the fitted line is not exactly the same as the normal line. Hence the estimated rainfall depths for selected probabilities or return period derived from the graphical solution (Table 3) will be slightly different from the rainfall depths obtained by the numerical solution (Table 4). The difference in both lines is caused by the sampling size. If the annual total rainfall in Bangkok is normal distributed, the two lines becomes identical when the sample size gets very large.

Table 4. Estimated annual rainfall for Bombay for selected probabilities and return periods obtained by the numerical solution.

Probability of exceedance P_X (%)	**Estimated annual rainfall (mm)**		**Return Period T_X (years)**
	$\overline{X} \pm k\,s$	X_P	
10	2063 + 1.28 (453.2) =	2643	10
20	2063 + 0.84 (453.2) =	2444	5
30	2063 + 0.53 (453.2) =	2303	3.33
40	2063 + 0.255 (453.2) =	2179	2.50
50	2063 + 0 (453.2) =	2063	2
60	2063 - 0.255 (453.2) =	1947	1.67
70	2063 - 0.53 (453.2) =	1823	1.43
80	2063 - 0.84 (453.2) =	1682	1.25
90	2063 - 1.28 (453.2) =	1483	1.11

4.3.3 Normalizing data (transformation of data)

Since dealing with a normal distribution has several practical advantages, it is common practice to transform data that are not normally distributed so that the resulting normalized data can be presented by the normal curve. The transformation of the data will change the scale of the records (i.e. the abscissa of the probability plot).

For positively skewed data a transformation can be used to reduce higher values by proportionally greater amounts than smaller values. This transformation will rescale the magnitude of the records and the transformed data might be closer to the normal distribution than the original data. This is illustrated in Figure 6, were positively skewed rainfall data was transformed to normality by rescaling the horizontal axis. By plotting the transformed data on a linear scale a normal distribution is obtained.

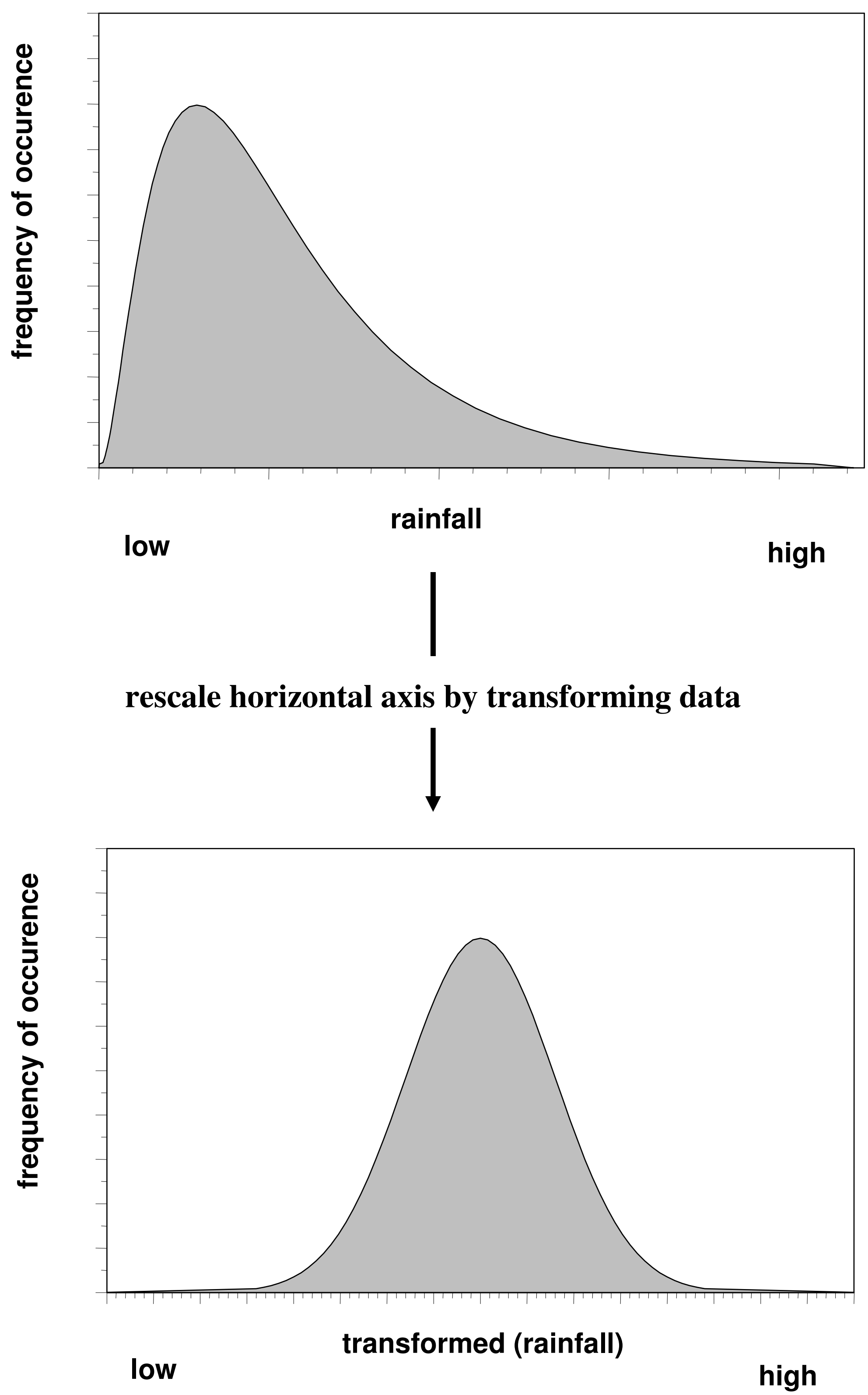

frequency of occurence
rainfall
low
high
rescale horizontal axis by transforming data
frequency of occurence
transformed (rainfall)
low
high

Figure 6. Transformation of positively skewed rainfall data.

Operators to rescale the data are (Figure 7):

- square root (resulting in a fairly moderate transformation)
- cube root
- logarithm (resulting in a substantive transformation)

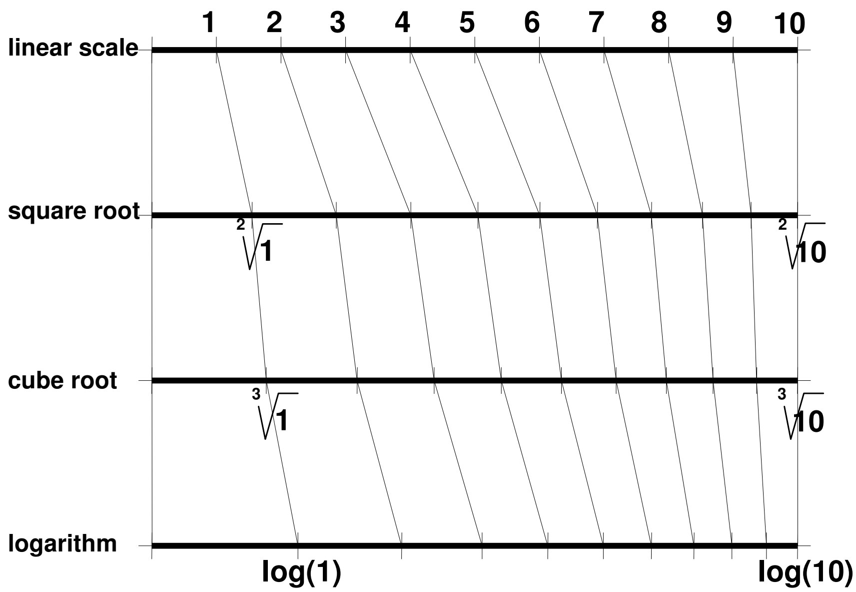


Figure 7. Effect of the transformation of data by several mathematical operators on the scaling of the horizontal axis.

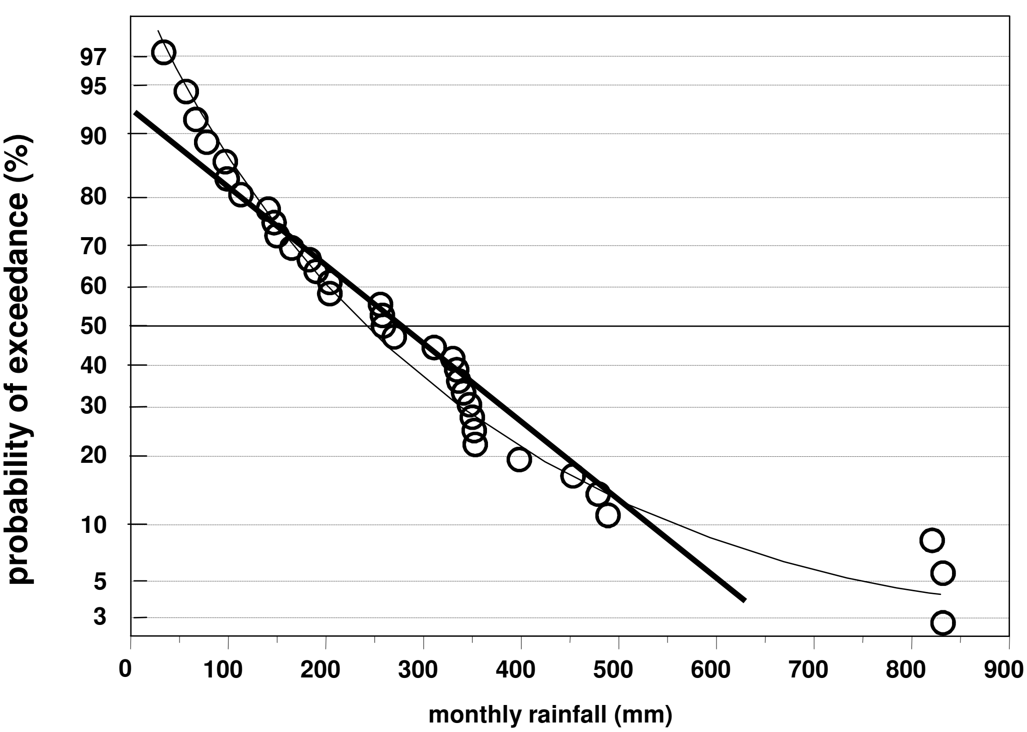


Figure 8. Probability plot of the September rainfall for Bombay.

Rainfall data for short periods (pentade (5-day), week, decade (10-day), month) are often not normally distributed but positively skewed. There is a relative high frequency of low rainfall below the mean and only a relative small numbers of high precipitations. This positive skew is one of the characteristics of the statistical distribution of rainfall data for short periods.

The mean rainfall in September in Bombay for example for the period 1960 – 1994 (Annex 1, Table A1.1) is not normal distributed. The mean monthly value is 295 mm. From the 35 observations, 19 are below and 16 are above the mean. The arithmetic mean is hence not the centre of the distribution. The fact that a straight line cannot be easily fitted through the points in the probability plot (Figure 8), indicates that the monthly rainfall is not normal distributed.

By taking the square root of the monthly rainfall, the data can be normalized. This is illustrated in Figure 9 by plotting the probability of the square root of the monthly rainfall data in a probability plot. The problem of line fitting is solved and the data fall in a reasonable alignment. The Figure reveals that the square root of the September rainfall of Bombay is normal distributed.

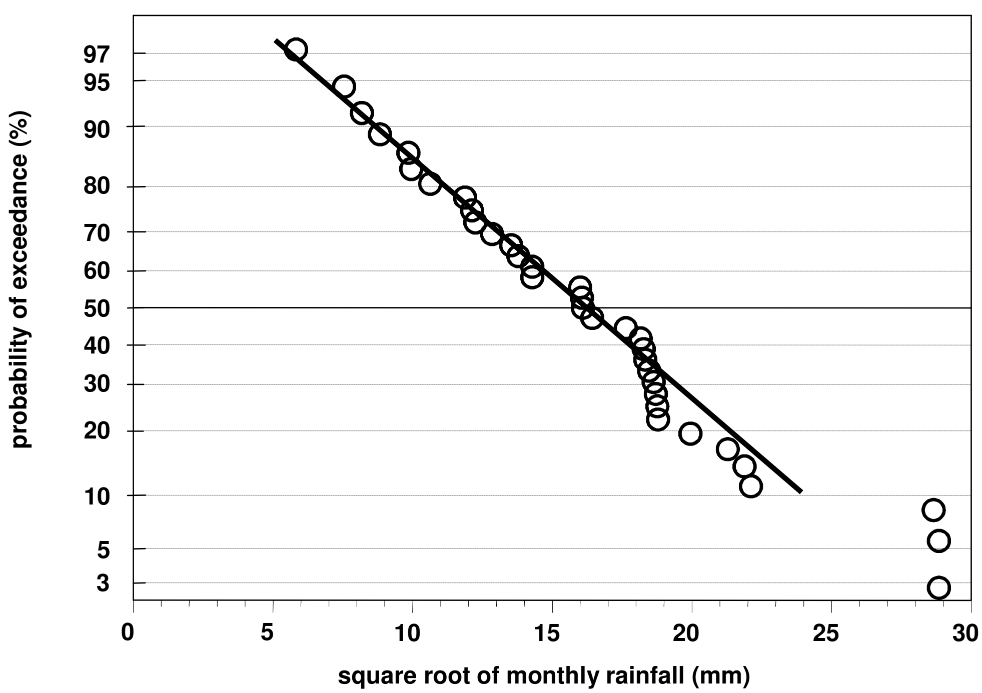


Figure 9. Probability plot of the square root of the September rainfall for Bombay on normal probability paper.

The estimated rainfall corresponding to various probabilities and return periods listed in Table 5 are derived from the fitted line through the data points in the probability plot (Figure 9).

Table 5. Estimated rainfall for September in Bombay for selected probabilities and return periods derived from the probability plot (Figure 9).

Probability of exceedance	Estimated rainfall		Return period
	Square root	Monthly rainfall	
P_X (%)	$\sqrt{X_P}$	X_P (mm)	T_X (years)
10	23.6	576	10
20	21.1	455	5
30	19.2	376	3.33
40	17.7	315	2.50
50	16.2	263	2
60	14.8	216	1.67
70	13.2	170	1.43
80	11.4	123	1.25
90	8.9	71	1.11

4.3.4 Data sets with zero rainfall

For months at the onset or cessation of the rainy season, or for small periods such as weeks or 10-day periods, rainfall data might be zero or near zero in some of the years. As such the rainfall data is bounded on the left by zero or near zero values. If the occurrence of low rainfall is high, the frequency distribution becomes severely skewed.

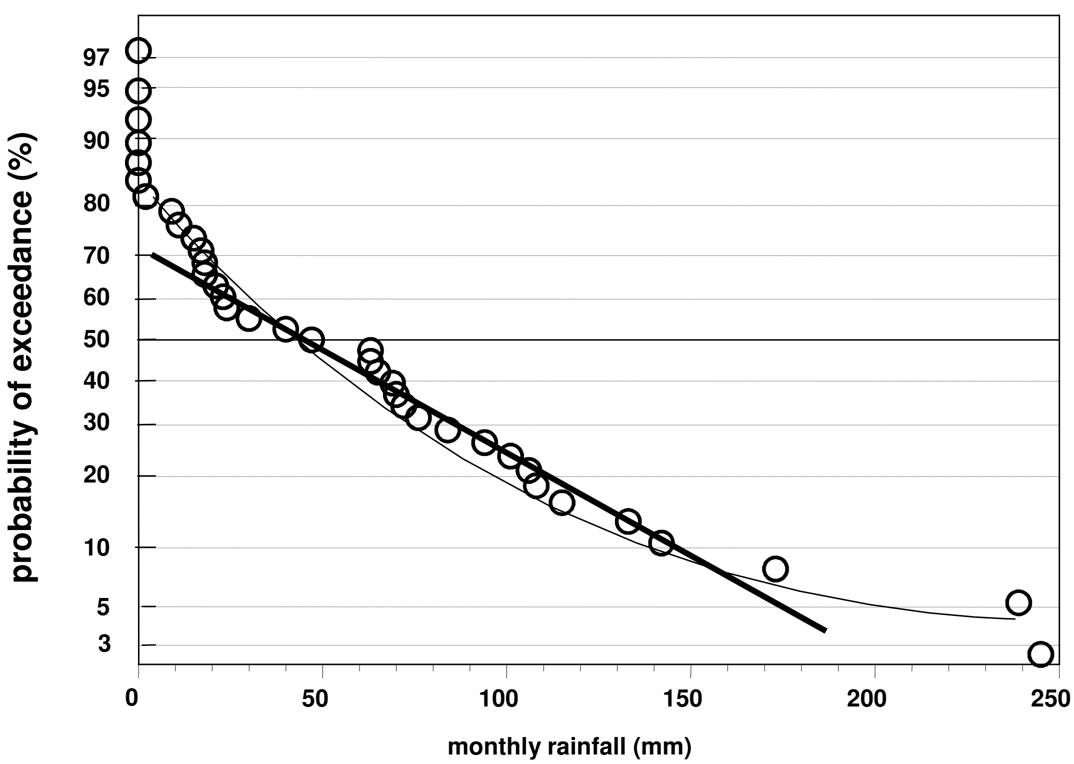


Figure 10. Probability plot of the October rainfall for Bombay on normal probability paper.

October in Bombay is a month in transition between the rainy and dry season. When the rainy season starts early, the rainfall in October might be important, but when the onset is late, the monthly rainfall might be zero or small. In the period 1960 – 1996, the rainfall is zero or small (< 10 mm) in 8 out of the 37 years i.e. in 22% of the years (Annex 1, Table A1.1). The data of the total time series are hence not normal distributed (Figure 10).

A method to analyse time-series with zero or near zero rainfall (the so called nil values) is to separate temporarily the nil values from the non-nil values. By excluding the nil's from the frequency analysis, the frequency distribution becomes less skewed to the left, and the data can be analysed. Figure 11 indicates that the logarithms of the non-nil values are normal distributed. The logarithms of the monthly rainfall plot more or less on a straight line in the probability plot. Although from this graph the probability of various levels of rainfall can be estimated, the results will be seriously biased since years with nil rainfall are ignored.

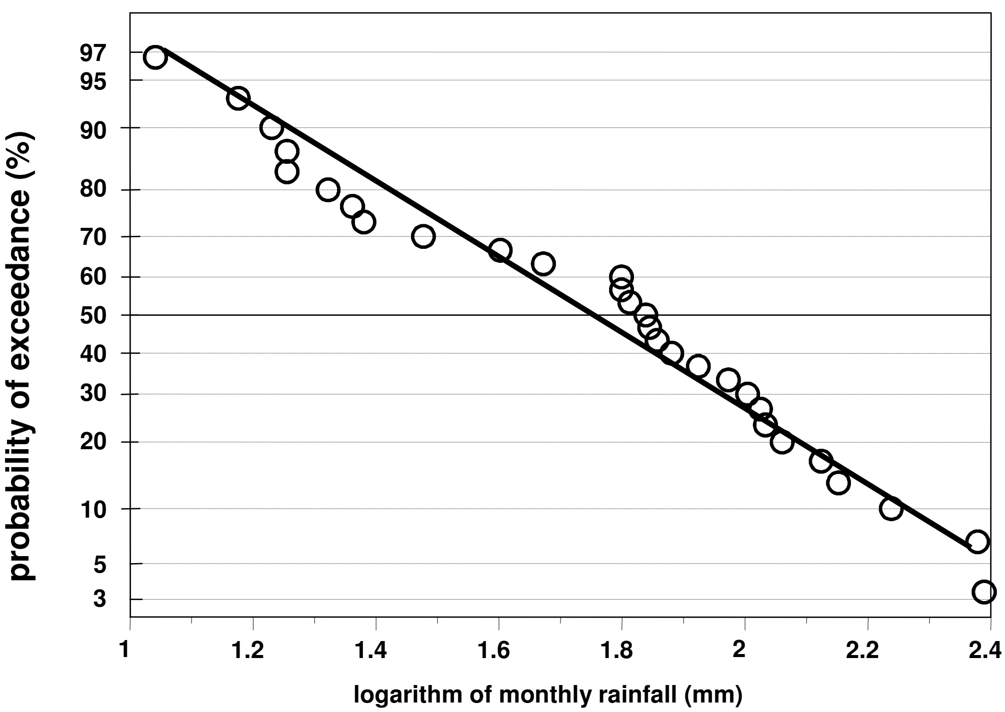


Figure 11. Probability plot of the logarithm of the non-nil October rainfall for Bombay on normal probability paper.

By calculating the global probability, the nil and no-nil rainfall can be combined. The global probability is calculated by

$$G_X(x) = (1 - p)F_X(x)$$

where $G_X(x)$ is the cumulative probability distribution of all X (prob($X \geq x$) for $X \geq$ nil), p is the probability that X is nil, and $F_X(x)$ is the cumulative probability distribution of the non-nil values of X (i.e. Prob($X \geq$ nil) for $X >$ nil).

The global distribution is a mixed distribution with a finite probability that X = nil and a continuous distribution of probability for X > nil. An example for the October rainfall is worked out in Table 6.

Table 6. Estimated rainfall for October in Bombay for selected probabilities and return periods derived from the probability plot of the non-nil values (Figure 11) and the probability (p = 0.22) that rainfall is zero or near zero (nil).

Global Probability G_X (%)				**Estimated Rainfall X_P (mm)**	**Return Period T_X (years)**
	Non-nil values (Figure 11)				
	$F_X = G_X/(1-p)$	**Log(X)**	**X**		
0	0	…	…	**…**	∞
10	12.8	2.215	164	**164**	**10**
20	25.6	2.021	105	**105**	**5**
30	38.3	1.876	75	**75**	**3.33**
40	51.0	1.745	56	**56**	**2.5**
50	63.8	1.613	41	**41**	**2**
60	76.6	1.462	29	**29**	**1.67**
70	89.3	1.253	18	**18**	**1.43**
78	100	…	…	**(nil)**	**1.28**
	Nil values (p = 22%)				
78					**1.28**
80				**≤10**	**1.25**
90				**(nil)**	**1.11**
100					**1**

4.4 Estimates of extreme rainfall (Gumbel distribution)

Estimates of extreme rainfall depths or intensities are required for the design of drainage or sewer systems. Therefore the annual maximum series, which is the set of maximum values observed during a period (minute(s), hour(s) or day) in each year, are analyzed. Several distribution functions can be selected to estimate extreme rainfall during the considered period. The Gumbel distribution which is skewed often gives satisfactory results.

As an example the annual maximum daily rainfall recorded in 24 hours in Uccle (Brussels) over the 50-year period from 1934 to 1983 has been analysed. In Annex 1 Table A1.2 the annual maximum daily rainfall as reported by Demarée (1985) are listed. In Figure 12 the probability plot is given. The cumulative probabilities were calculated by means of the Gringorten method. Values of annual daily maximum rainfall derived from the probability plot estimated by the Gumbel distribution function are listed in Table 7.

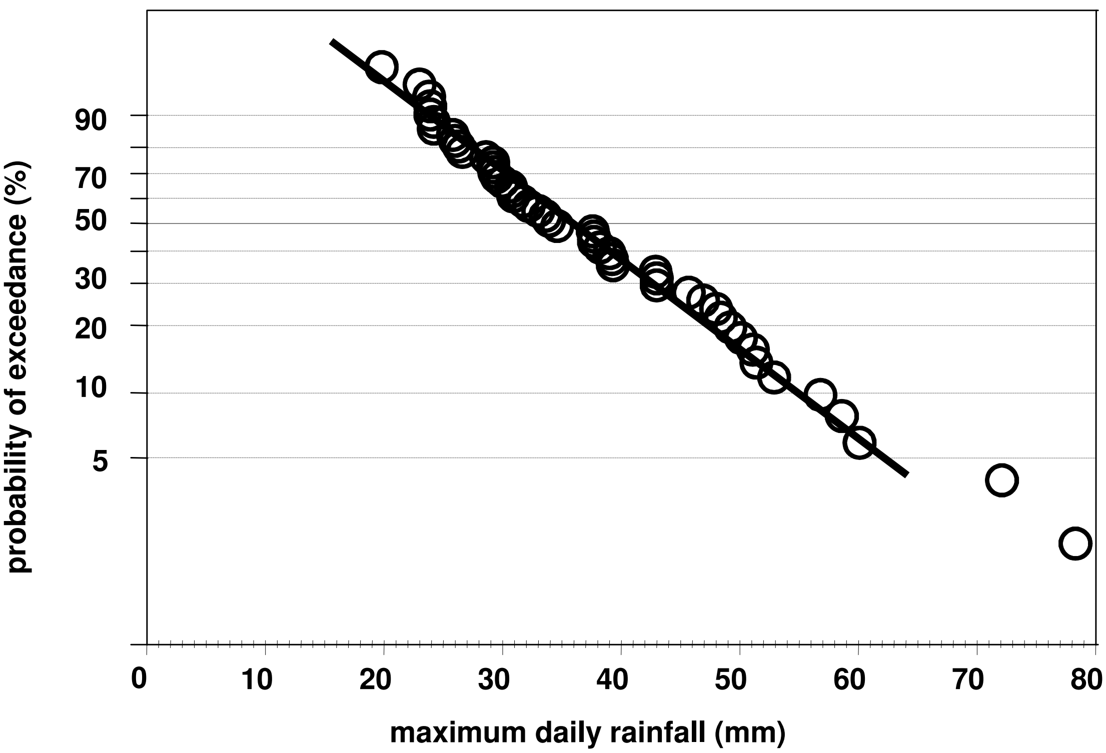


Figure 12. Probability plot of the annual daily maximum rainfall for Brussels on Gumbel probability paper.

Table 7. Estimated annual daily maximum rainfall for Uccle (Brussels, Belgium) for selected probabilities and return periods derived from the probability plot (Figure 12).

Probability of exceedance (%)	Annual daily maximum rainfall (mm)	Return period (years)
5	62.1	20
10	54.9	10
20	47.3	5
50	35.9	2

As for the normal distribution a series of annual maxima can be analyzed also numerically by considering the mean $\overline{X}$ and standard deviation s of the observed series of extremes:

$$X_P = \overline{X} \pm k\ s$$

where X_P is the rainfall depth that can be expected for a specific probability and k is the frequency factor dependent upon the sample size n and the required return period or probability of exceedance. Values for k are presented in Annex 2, Table A2.2.

4.5 Other distributions

5. Homogeneity test of time series

Frequency analysis of rainfall data requires that the data be homogeneous and independent. The restriction of homogeneity assures that the observations are from the same population. One of the test of homogeneity (Buishand, 1982) is based on the cumulative deviations from the mean:

$$S_k = \sum_{i=1}^{k} \left(X_i - \overline{X} \right) \qquad k = 1, \ldots, n$$

where X_i are the records from the series X_1, X_2, …, X_n and $\overline{X}$ the mean. The initial value of $S_{k=0}$ and last value $S_{k=n}$ are equal to zero (Figure 13). When plotting the S_k's (sometimes called a residual mass curve) changes in the mean are easily detected. For a record X_i above normal the $S_{k=i}$ increases, while for a record below normal $S_{k=i}$ decreases. For a homogenous record one may expect that the S_k's fluctuate around zero since there is no systematic pattern in the deviations of the X_i's from their average value $\overline{X}$.

To test the homogeneity of the data set, the cumulative deviations are often rescaled. This is obtained by dividing the S_k's by the sample standard deviation value (s). By evaluating the maximum (Q) or the range (R) of the rescaled cumulative deviations from the mean, the homogeneity of the data of a time series can be tested:

$$Q = \max \left[\frac{S_k}{s} \right]$$

$$R = \max \left(\frac{S_k}{s} \right) - \min \left(\frac{S_k}{s} \right)$$

High values of Q or R are an indication that the data of the time series is not from the same population and that the fluctuations are not purely random. Critical values for the test-statistic which test the significance of the departures from homogeneity are presented in Table 8. The percentage points in this table are based on 19,999 synthetic sequences of Gaussian random numbers.

Table 8. Percentage points of Q/√n and R/√n for rejecting the homogeneity of time series with a 90, 95 and 99 % probability (Buishand, 1982).

Sample Size	Q/√n			R/√n		
n	90%	95%	99%	90%	95%	99%
10	1.05	1.14	1.29	1.21	1.28	1.38
20	1.10	1.22	1.42	1.34	1.43	1.60
30	1.12	1.24	1.46	1.40	1.50	1.70
40	1.13	1.26	1.50	1.42	1.53	1.74
50	1.14	1.27	1.52	1.44	1.55	1.78
100	1.17	1.29	1.55	1.50	1.62	1.86
∞	1.22	1.36	1.63	1.62	1.75	2.00

In Figure 13 the cumulative deviations from the mean of the total annual rainfall data for the (complete) time series 1940 – 1987 for Bombay are plotted. In this graph, the vertical-axis is rescaled and lines presenting various probabilities with which the homogeneity of the data can be rejected are plotted. Since the rescaled S_k's fluctuate around zero and are far off the lines where the homogeneity is rejected, the data of the time series can be considered as homogeneous.

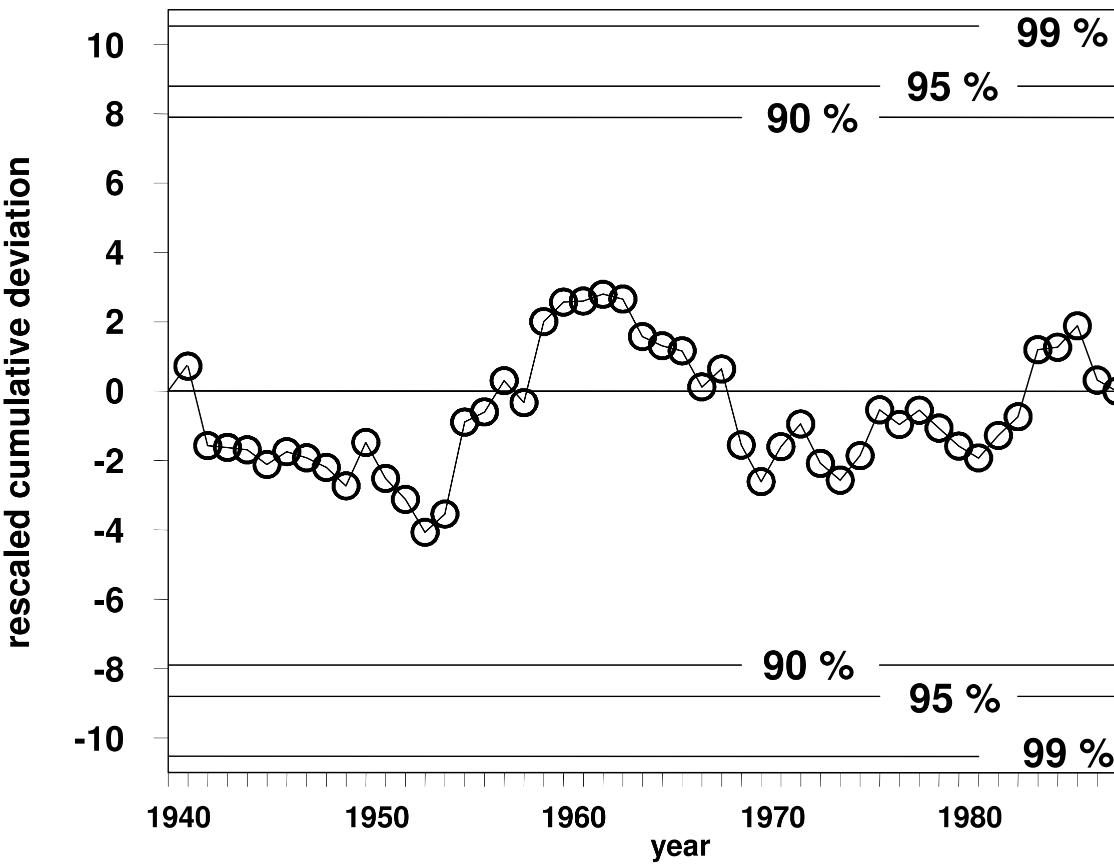


Figure 13. Rescaled cumulative deviations from the mean for the total annual rainfall for Bombay for the time series 1940 – 1987. When the deviation crosses one of the horizontal lines the homogeneity of the data set is rejected with respectively 90, 95 and 99% probability.

6. Practical Applications

6.1 Estimating the likelihood of the occurrence of an event

By means of a frequency analysis the likelihood of the occurrence of an event can be easily estimated. As an example it is checked if the summer in central Belgium was exceptionally dry in 2003. Therefore rainfall data was extracted from the FAOCLIM databank (FAO, 2000) for Uccle (Brussels). The databank contains time series of 164 years (from 1833 to 1996) of monthly data for this weather station. Since the rainfall record for one of the years (1989) is incomplete, that year is excluded from the analysis. The summer rainfall for each of the years is obtained by adding up the monthly rainfall for 4 months: June, July, August and September. The average summer rainfall for the time series of the 163 years is 283.7 mm and its standard deviation is 66.08 mm.

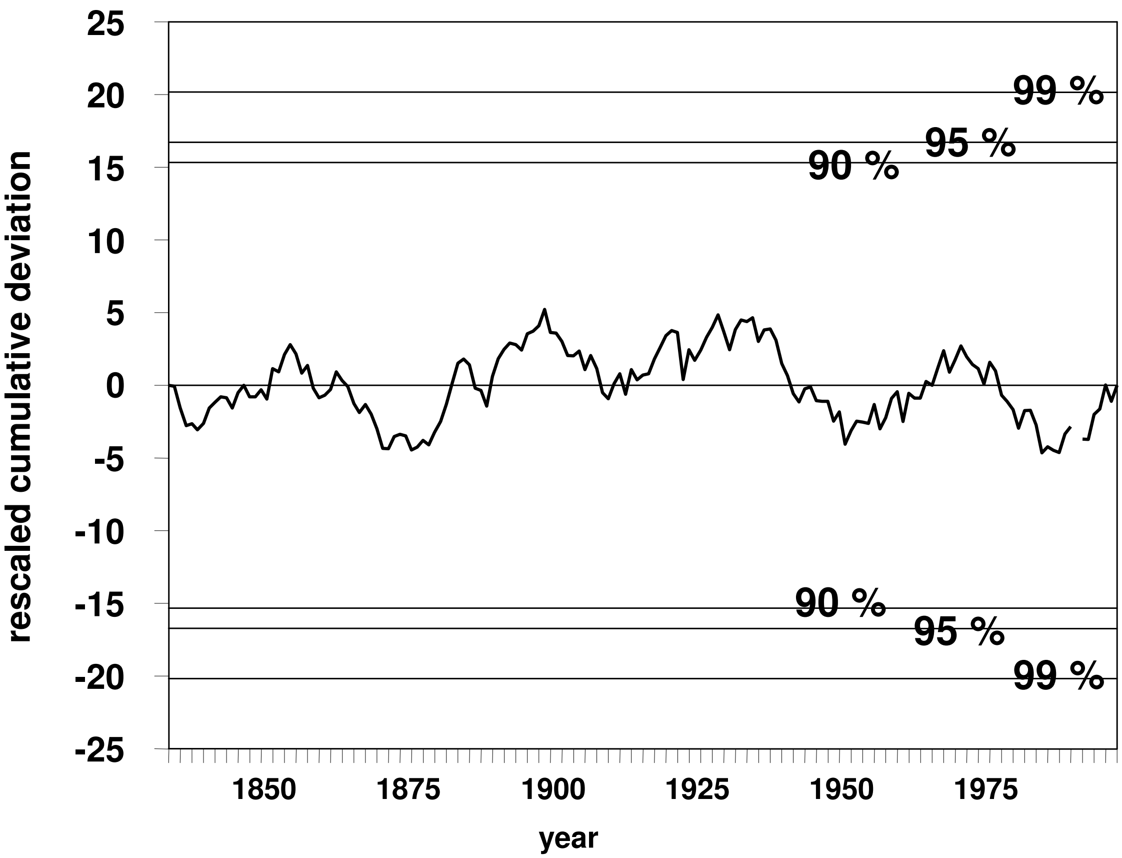


Figure 14. Homogeneity test for the time series 1833-1996 of summer rainfall for Uccle (Brussels, Belgium).

Before analysing the data it was checked if the summer rainfall for the time series is a homogeneous dataset. The homogeneity test for the time series indicated that there is no systematic pattern in the deviations of the summer rainfalls from their mean (Figure 14).

Subsequently it was graphically checked if the data are normal distributed. The probability plot (Figure 15) indicates that the data is normal distributed.

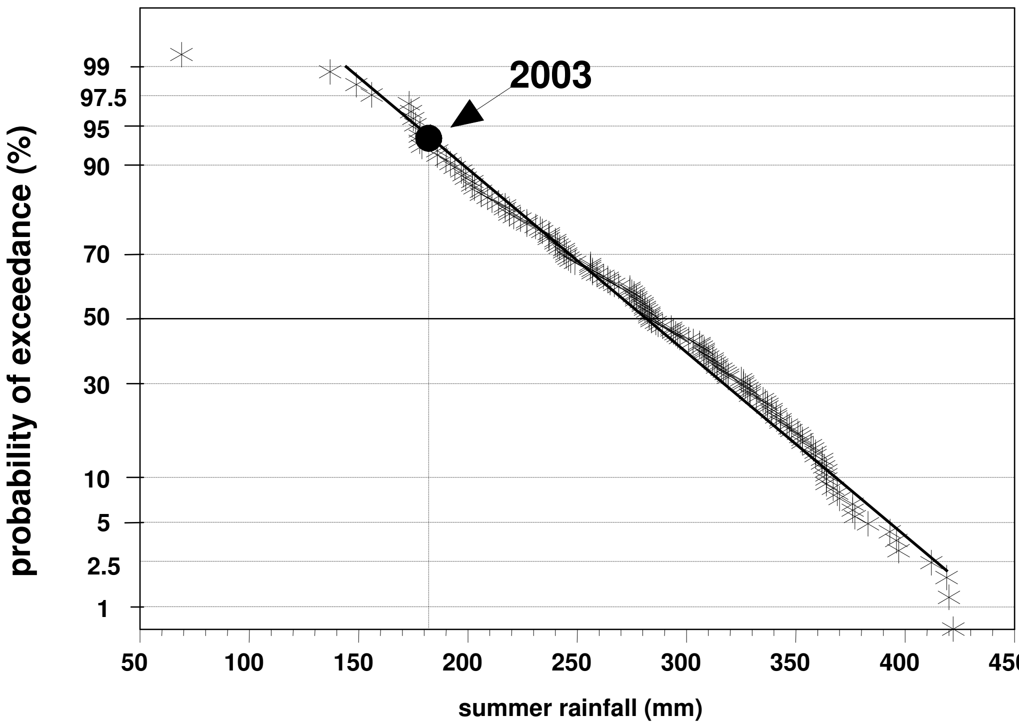


Figure 15. Probability plot of the summer rainfall for Uccle (Brussels, Belgium) on normal probability paper with indication of the rainfall recorded in 2003.

The summer rainfall for 2003 was only 182 mm. By plotting the value on the probability plot, it is evident that the summer in that year was indeed rather dry. The probability that the value is exceeded is 94%, or this means that less rainfall occurs only on average in 6 out of 100 years or the probability of non-exceedance is 6%. The return period for which the summer rainfall is less than the one in 2003 is (1/0.06 =) 16.7 years. According to the classification of meteorological events used by the Royal Meteorological Institute of Belgium, the summer rainfall in 2003 was very abnormal but not yet exceptional (Table 9).

Table 9. Classification of meteorological events used by the Royal Meteorological Institute of Belgium.

Classification	Return period
Normal	
Abnormal	6 year
Very abnormal	10 year
Exceptional	30 year
Very exceptional	100 year

6.2 Classification of the actual weather conditions for guiding irrigation in real time

Irrigation charts (Annex 4) can help farmers in their day-to-day decisions (Raes et al., 2000 and 2002). The chart guides the user in the adjustment of the irrigation interval to the actual weather conditions throughout the growing season. Since the objective of the irrigation chart is to give farmers guidelines for the adjustment of their irrigation calendars to the actual weather conditions, the development of the irrigation calendars requires information on rainfall and evapotranspiration levels that can be expected with various probabilities. The rainfall and evapotranspiration depths are derived statistically from long records of historical 10-day rainfall and reference evapotranspiration (ETo). By combining various probability levels of ETo and rainfall, four weather conditions are distinguished (Table 10) for which an irrigation advice will be formulated with the help of a soil water balance model. The various levels of ETo and Rainfall for Harare (Zimbabwe) considered in the irrigation chart (Annex 4) are plotted in Figure 16.

Table 10. The four distinguished weather conditions in irrigation charts (Annex 4).

Weather condition	Probability level of exceedance	
	ETo	Rain
Hot and dry	20 %	100 % (No rain)
Dry	40 %	80 %
Normal	50 %	50 %
Wet	60 %	20 %

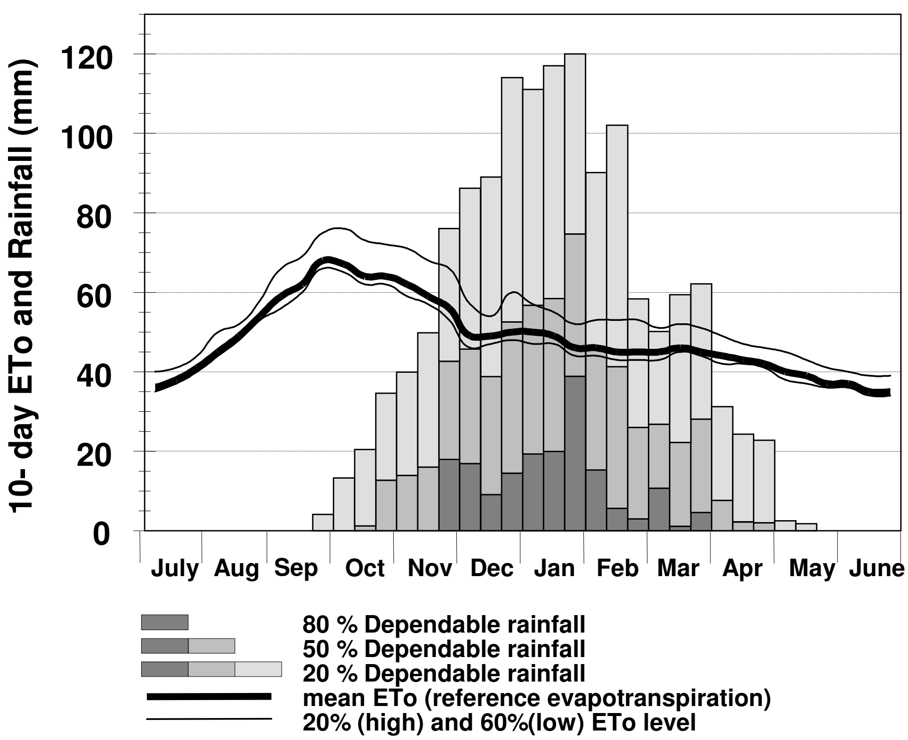


Figure 16. Estimated 10-day Rainfall and ETo levels for selected probabilities of exceedance for Harare (Zimbabwe).

6.3 Intensity-duration-frequency (IDF) curves

Estimates of precipitation intensity for selected return periods are often required for the design of sewer or surface drainage systems. Reliable estimates can only be obtained from a frequency analysis of long time series. Such long time series are available at the Royal Meteorological Institute at Uccle (Brussels, Belgium) which recorded during more than 100 years the rainfall with the help of a recording raingauge (Demarée et al., 1998).

From a 101 years time series (1898-1998) of the 10-minute precipitation depths, the maximum value in each year for various time periods (ranging from 10 minutes to several days) were determined. The annual maximum series were subsequently analysed with the help of the Gumbel distribution. The expected annual maximum rainfall intensities (recorded depth over time) during the considered time periods were determined for selected return periods. The results are quite reliable since the study is based on observation data from a long time series and on homogeneous sets of data recorded by a single, accurate and well-maintained precipitation gauge. The results, the so-called IDF (intensity-duration-frequency) curves are plotted in Figure 17 by using logarithmic scales for both axes. In DDF (depth-duration-frequency) curves the total rainfall depth instead of the intensity is plotted.

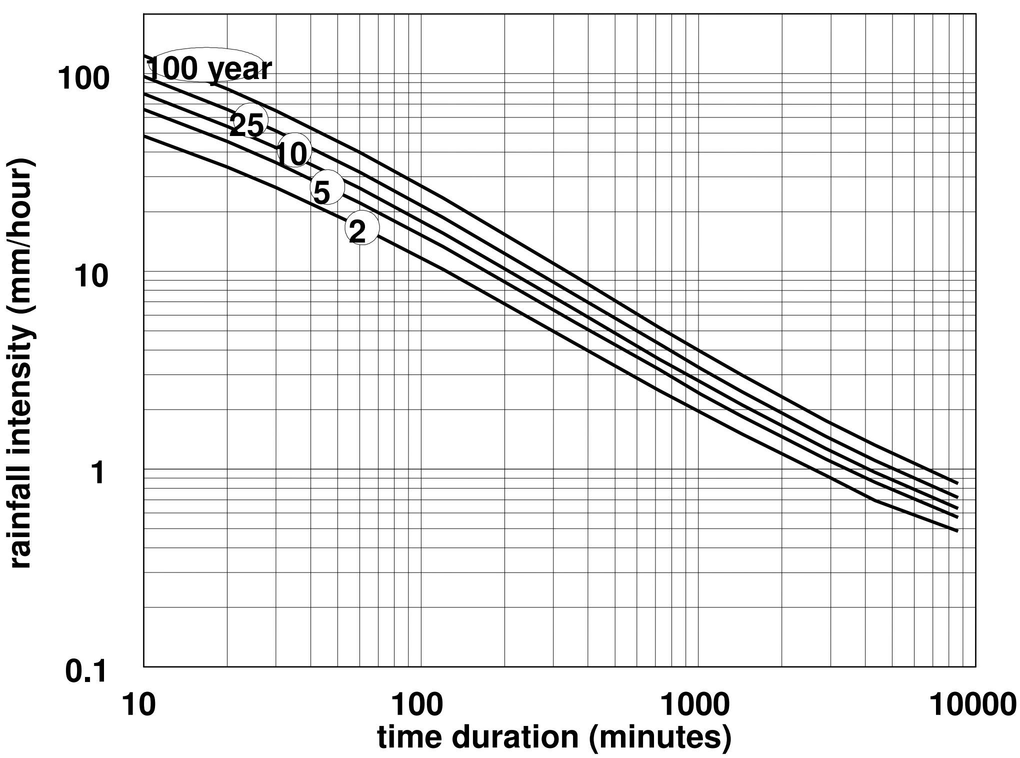


Figure 17. Intensity-Duration-Frequency (IDF) curves for annual rainfall in Uccle (Brussels, Belgium), for selected return periods (source Delbeke, 2001).

6.4 Detecting shifts in the total annual rainfall

Cumulative deviations from the mean are often used to analyse the homogeneity of time series. By plotting the deviations in function of time, strong deviations from the mean can be easily detected. These deviations can be the result of the installation of new recording equipment, altering the position of the equipment or a shift in the climate.

A shift has been detected in the precipitation climate of the Sahelian zone of West Africa and is described by several authors (Demarée, 1987; Vannitsem and Demarée, 1991). As an illustration rainfall data of Saint-Louis (North of Senegal) has been extracted from the FAOCLIM database. In Figure 18, the cumulative deviations from the mean of the total annual rainfall data from 1930 to 1993 are plotted. The homogeneity test clearly reveals the major pattern of the Sahelian precipitation climate. The most striking feature is the clear change of slope in the year 1969. Over the period 1930 – 1969 the total annual rainfall was above normal while over the period 1970 – 1993 the rainfall is below normal rainfall.

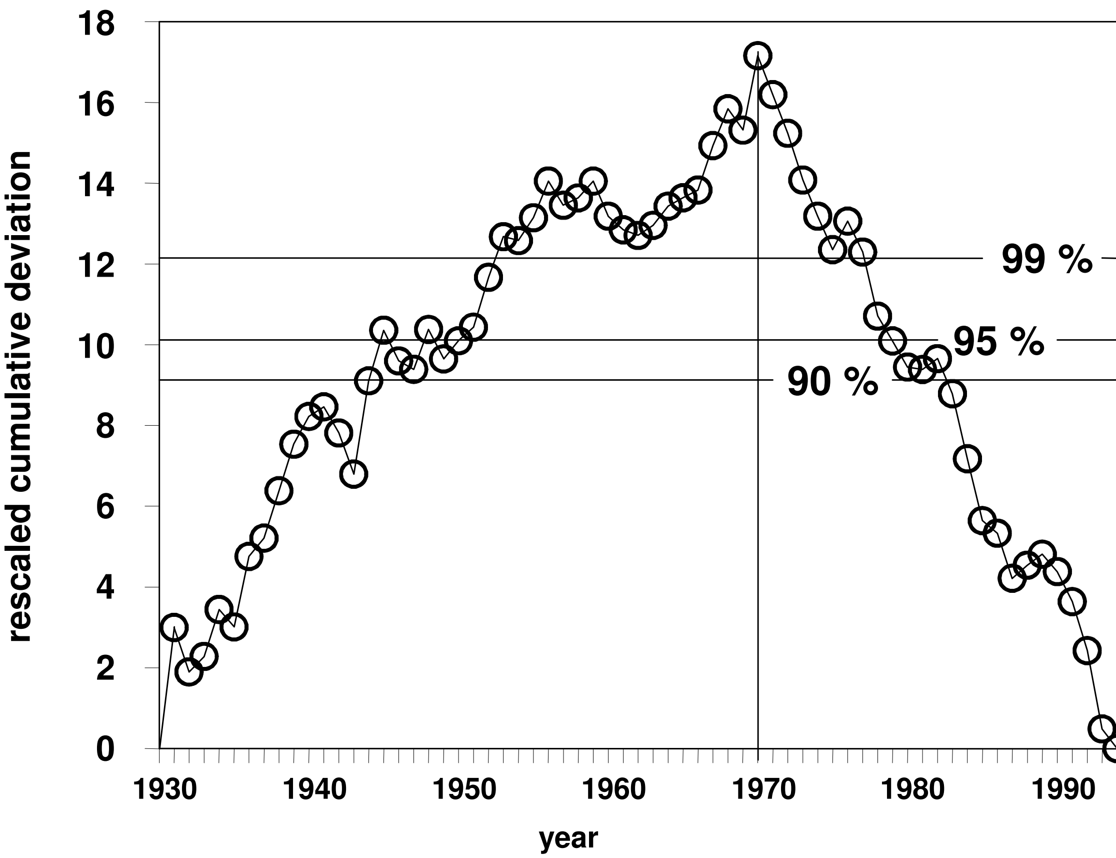


**Figure 18. Rescaled cumulative deviations from the mean for the total annual rainfall for SaintLouis (Senegal) for the time series 1930-1993.
The horizontal lines present the 90, 95 and 99% probability with which the homogeneity of the data is rejected.**

The reference period 1930 – 1993 can be split up into two statistically significant periods different in the mean: 1930 – 1969 with a mean annual rainfall of 355 mm and 1970 – 1993 with a mean of 211 mm (Figure 19). The jump in the mean between the years 1969/1970 separates the two periods.

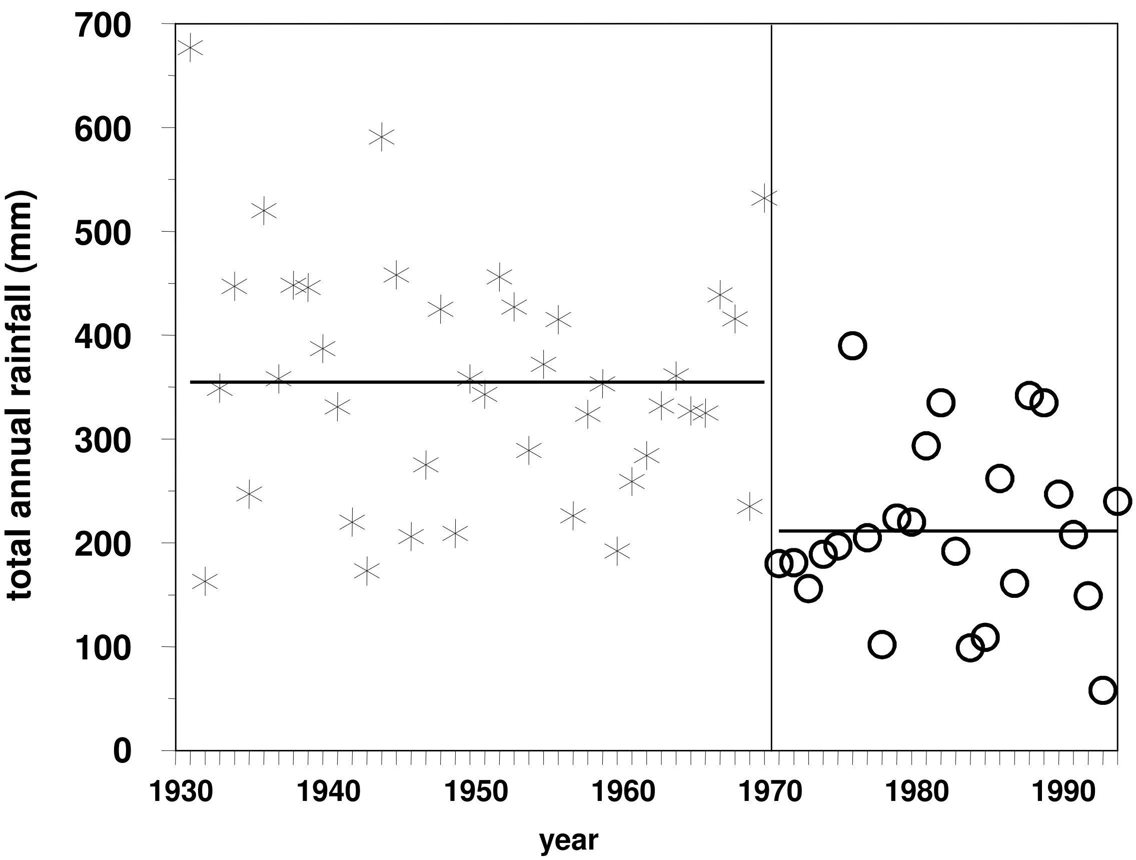


Figure 19. Two statistically significant distinct periods in the precipitation climate of Saint-Louis (Senegal), with indication of the total annual rainfall data for the two time series (asterisks and circles) and their corresponding means (horizontal lines).

7. RAINBOW software

RAINBOW (Raes et al., 1996) is a software tool designed to study agro-meteorological or hydrologic records by means of a frequency analysis and to test the homogeneity of the record. The program is especially suitable for predicting the probability of exceedance of either low or high rainfall events, both of which are important parameters for the design and management of irrigation systems, drainage network, and reservoirs. The software is freely available on the web. To DOWNLOAD go to http://www.iupware.be and select downloads and next software.

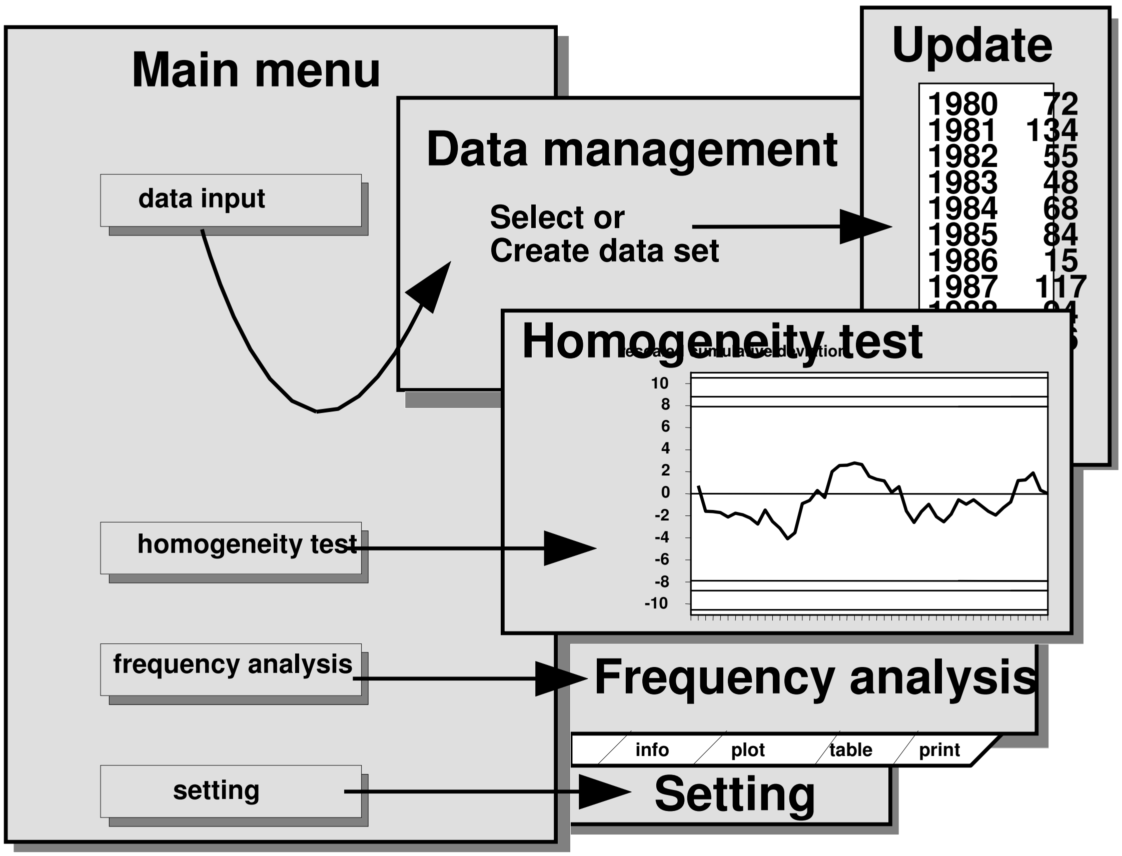


Figure 20. Structure of the RAINBOW program

The hierarchical structure of the program is presented in Figure 20. From the ***Main Menu*** the user has access to:

- the ***Data Management menu*** to select an existing or to create a new data set,
- the ***Homogeneity test menu*** for testing the homogeneity of the data set,
- the ***Frequency analysis menu*** for analysing the data set (Figure 21),
- the ***Setting menu*** where specific program parameters can be set.

The analysis starts with the selection or creation of a data file in the ***Data Management*** menu. In stead of creating files when running RAINBOW, the user can also directly add, update and delete files in the subdirectory as long as the user respects the structure and extension of the files (Annex 5).

By selecting ***Setting*** in the ***Main menu*** the user can

- exclude outliers from the data analysis;
- specify the value for nil when analysing data sets with near zero rainfall;
- split time series in two periods when the homogeneity of the total record is rejected.

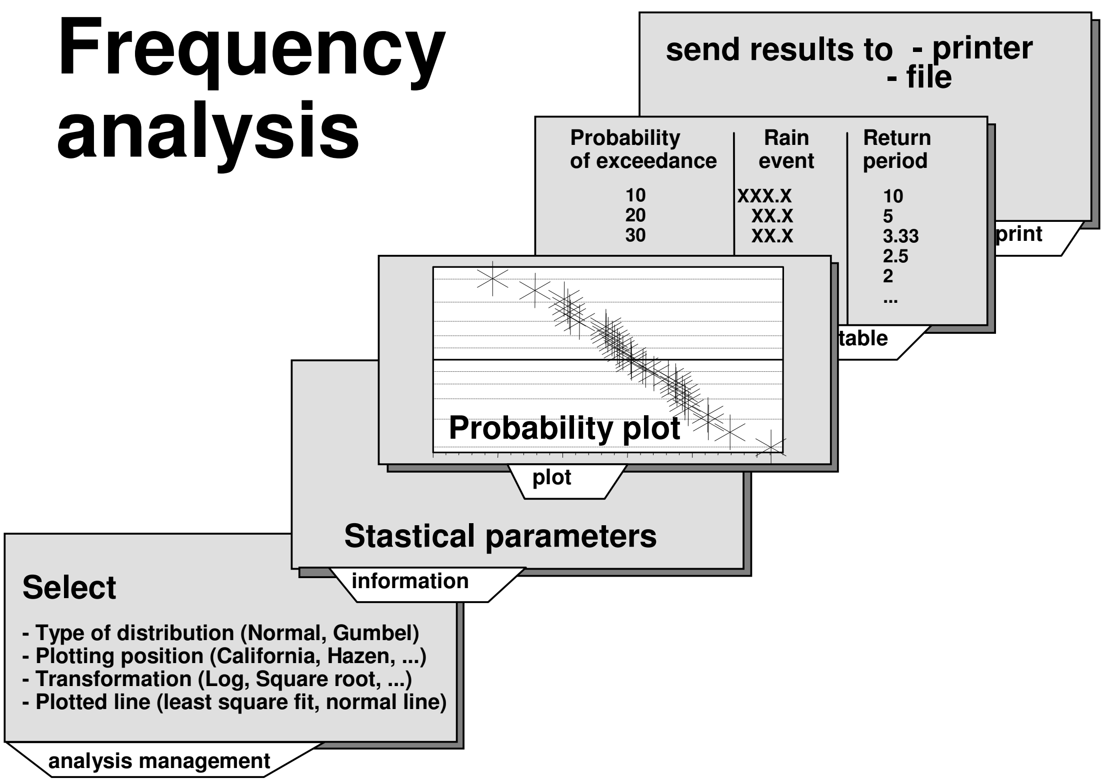


Figure 21. Pages of the *Frequency analysis* menu.

The information available on the different pages of the ***Frequency analysis menu*** is presented in Figure 21. After the selection in the *Analysis Management page* of the type of distribution, the method for estimating the probability of exceedance, and the type of line that will be fitted through the data points in the probability plot, the user can graphically check the validity of the selected distribution. When the points in the probability plot (*Probability plot page*) fall in a reasonable alignment, the user finds estimates for rainfall depths for selected probabilities and return period in a table on the *Probability table page*. When the points do not fall in a reasonable alignment, the data is most likely not distributed as the selected distribution. One can either select another distribution or attempt to normalize the data by selecting a mathematical operator to transform the data (*Analysis management page*). Statistical information on the goodness of fit and on the distribution is listed on the *Information page*.

8. References

Buishand, T.A. 1982. Some methods for testing the homogeneity of rainfall records. Journal of Hydrology, (58): 11 – 27.

California State Department of Public works. 1923. Flow in California streams. Bulletin 5. Chapter 5 (cited by Haan, 1986).

Delbeke, L. 2001. Extreme neerslag in Vlaanderen. Ministerie van de Vlaamse Gemeenschap, Afdeling Water, Brussel, België. 123 pp.

Demarée, G. 1985. Intensity-Duration-Frequency relationship of point precipitation at Uccle – Reference period 1934-1983. Royal Meteorological Institute of Belgium, Publications series A (116). 52 pp.

Demarée, G. 1987. Traitement des données pluviométriques de la station de Kaedi (Mauritanie). Royal Meteorological Institute of Belgium, Publications series C (24). 40 pp.

Demarée, G.R., De Corte, M., Derasse, S., Devorst M., en Trapenard, Ch. 1998. Een kranige honderdjarige: De Hellmann-Fuess pluviograaf van het Koninklijk Meteorologisch Instituut te Ukkel. Water (100): 145 - 149.

Doorenbos, J. and Pruitt, W.O. 1977. Crop water requirements. FAO Irrigation and Drainage Paper N° 24, Rome, Italy. 144 pp.

FAO. 2000. FAOCLIM – World-wide agroclimatic database. Environment and Natural Resources Working Paper No. 5. (CD-ROM). FAO - Agro meteorology group, Rome, Italy.

Haan, C.T. 1986. Statistical methods in Hydrology. Ames, Iowa, USA: 128 -160.

Hazen, A. 1930. Flood flow, a study of frequencies and magnitudes. John Wiley and Sons, INC. New York (as cited by Haan, 1986).

Holvoet, K. 2002. Participative development of irrigation guidelines for a sustainable water use in the Highveld region of Zimbabwe. Master dissertation, K.U.Leuven, Fac. of Agric. and Applied Biol. Sciences, Leuven, Belgium. 82 pp.

Raes, D., Mallants, D. and Song Z. 1996. RAINBOW – a software package for analysing hydrologic data. In Blain W.R. (Ed.) Hydraulic Engineering Software VI. Computational Mechanics Publications, Southampton, Boston: 525-534.

Raes, D., Sahli, A., Van Looij, J., Ben Mechlia, N., and Persoons, E. 2000. Charts for guiding irrigation in real time. Irrigation and Drainage Systems. Volume 14(4): 343-352.

Raes, D., Smith, M., De Nys, E., Holvoet, K. and Makarau, A. 2002. Charts with indicative irrigation intervals for various weather conditions. ICID 18th Congress, 21-28 July 2002, Montreal, Canada (http://www.fao.org/ag/agl/aglw/ias/docs/paper3.pdf).

Sevruk, B. and Geiger, H. 1981. Selection of distribution types for extremes of precipitation. World Meteorological Organisation, Operational Hydrology Report, No. 15, WMO-No. 560, Geneva.

Smedema, L.K. and Rycroft, D.W. 1983. Land drainage: planning and design of agricultural systems. Batsford Academic and Educational Ltd., London, UK. 376 pp.

Smith, M. 1992. CROPWAT – a computer program for irrigation planning and management. FAO Irrigation and Drainage Paper, N° 46., Rome, Italy. 125 pp.

Snedecor, G.W. and Cochran, W.G. 1980. Statistical methods. Iowa State university press, USA. 507 pp.

Vannitsem, S. et Demarée, G. 1991. Détection et modélisation des sécheresses au Sahel – Proposition d'une nouvelle méthodologie. Hydrol. Continent (6): 155 – 171.

Weibull, W. 1939. A statistical study of the strength of material. Ing. Vetenskaps Akad. Handl. (Stockholm) Vol. 151, pp. 15 (as cited by Haan, 1986).

WMO. 1981. Guide to agricultural meteorological practices. World Meteorological Organization, WMO – No. 134. Geneva, Switzerland.

WMO. 1983. Guide to climatological practices. World Meteorological Organization, WMO – No. 100. Geneva, Switzerland.

WMO. 1990. On the statistical analysis of series of observations. World Meteorological Organization, WMO –N° 415. Geneva, Switzerland. 192 pp.

Annex 1 – Rainfall data

Table A1.1 Rainfall observed in Bombay (India) for the 37-year period 1960 – 1996 (Source FAOCLIM, 2000).

Year	September (mm)	October (mm)	Total (mm)
1960	97	245	2119
1961	270	101	2205
1962	341	21	2029
1963	330	18	1540
1964	258	115	1964
1965	78	0	2026
1966	183	18	1560
1967	204	72	2373
1968	150	40	960
1969	336	0	1553
1970	311	63	2626
1971	347	23	2444
1972	67	0	1511
1973	350	65	1850
1974	147	108	2469
1975	453	173	2791
1976	398	0	1882
1977	479	17	2316
1978	256	63	1835
1979	190	9	1832
1980	141	24	1923
1981	821	69	2441
1982	204	0	2382
1983	489	76	3107
1984	334	15	2141
1985	113	142	2423
1986	99	0	1287
1987	34	47	1937
1988	832	239	-
1989	165	30	1900
1990	352	106	-
1991	57	2	2620
1992	259	70	1629
1993	832	84	2370
1994	353	11	2031
1995	-	133	-
1996	-	94	-
Average	**295**	**62**	**2063**
Standard Deviation	**206.0**	**63.6**	**453.2**

Table A1.2 Annual maximum daily rainfall for Uccle (Brussels, Belgium) for the 50-year period 1934-1983 (Source Demarée, 1985).

Year	Rain (mm)	Year	Rain (mm)	Year	Rain (mm)
1931	-	1951	43.0	1971	45.7
1932	-	1952	51.1	1972	28.6
1933	-	1953	39.2	1973	30.0
1934	23.9	1954	31.7	1974	30.7
1935	37.6	1955	23.0	1975	25.8
1936	23.8	1956	34.6	1976	33.0
1937	29.2	1957	37.7	1977	26.3
1938	33.9	1958	42.9	1978	49.2
1939	37.7	1959	30.8	1979	33.6
1940	60.1	1960	48.0	1980	38.2
1941	30.9	1961	32.2	1981	56.8
1942	72.1	1962	58.6	1982	50.1
1943	51.4	1963	78.3	1983	24.2
1944	24.2	1964	43.0		
1945	46.9	1965	48.4		
1946	26.6	1966	39.3		
1947	29.3	1967	23.9		
1948	29.5	1968	26.0		
1949	19.8	1969	52.9		
1950	39.0	1970	29.2		
Average: 38.1 mm					
Standard deviation: 12.92 mm					

Annex 2 – Frequency factors

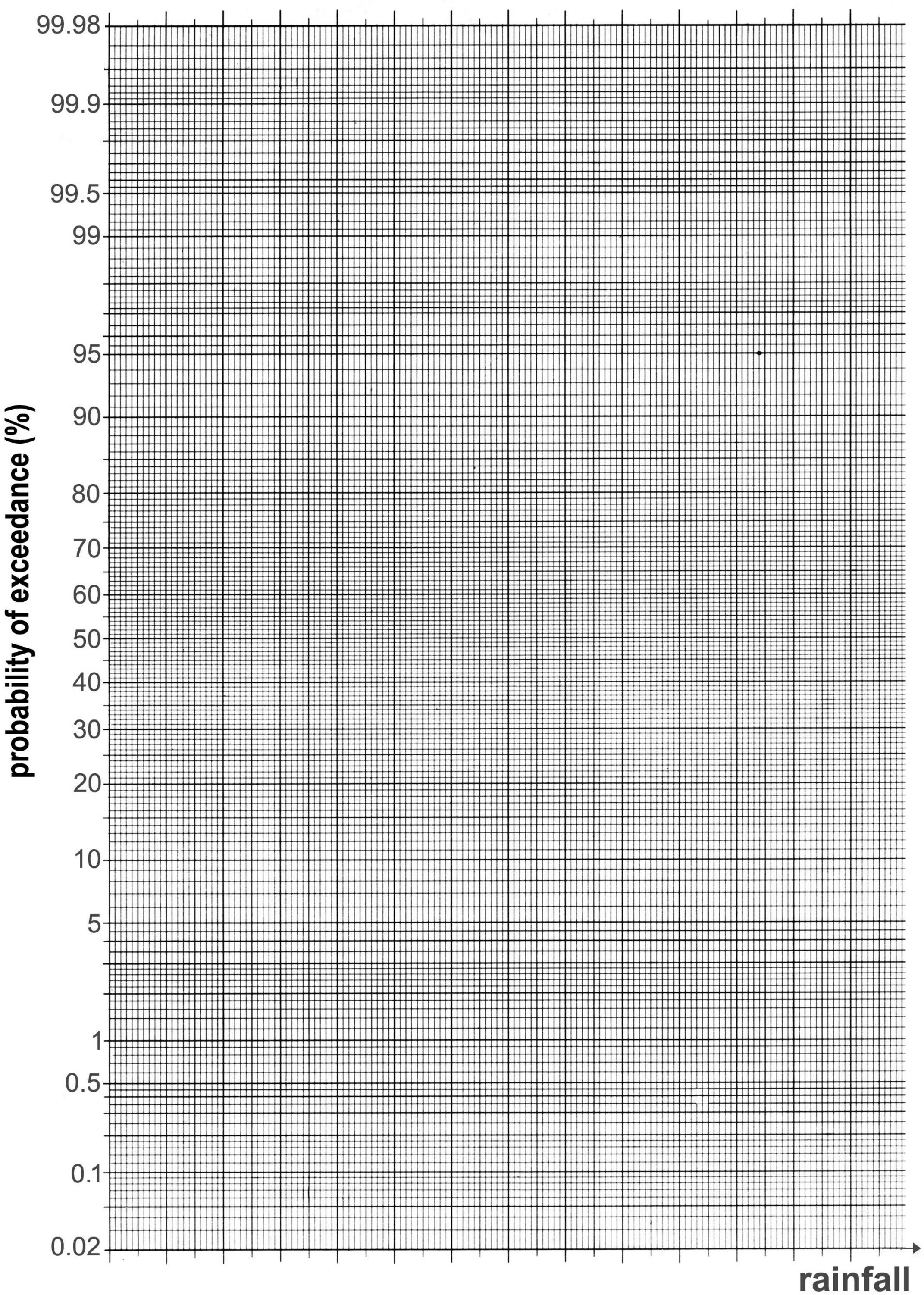
Table A2.1 Values for the frequency factor k for normal frequency distribution for various probabilities of exceedance (extracted from Snedecor and Cochran, 1980).

Probability of exceedance (%)	+ k	Probability of exceedance (%)	- k
5	+ 1.64	50	- 0.000
10	+ 1.28	55	- 0.125
15	+ 1.04	60	- 0.255
15.87	+ 1.00	65	- 0.39
20	+ 0.84	70	- 0.53
25	+ 0.66	75	- 0.66
30	+ 0.53	80	- 0.84
35	+ 0.39	84.13	- 1.00
40	+ 0.255	85	- 1.04
45	+ 0.125	90	- 1.28
50	+ 0.000	95	- 1.64

Table A2.2 Values for the frequency factor k for extreme value distributions for specific return periods and sample size (Source: WMO, 1983).

Sample Size	Return period (years)					
n	**5**	**10**	**15**	**20**	**50**	**100**
5	1.31	2.26	2.79	3.17	4.34	5.22
10	1.06	1.85	2.29	2.61	3.59	4.32
15	0.97	1.70	2.12	2.41	3.32	4.01
20	0.89	1.58	1.96	2.30	3.09	3.73
50	0.82	1.47	1.83	2.09	2.89	3.49
100	0.78	1.40	1.75	2.00	2.77	3.35

Annex 3 – Normal probability paper



Annex 4 – Irrigation chart

Irrigation guidelines for:

Potatoes

Soil type: clay (Harare 5E2)
Irrigation application gross depth: 25 mm

Irrigation interval given in days

Month		December			January			February			March		
Decade (10 days)		1	2	3	1	2	3	1	2	3	1	2	3
Actual weather condition	Hot + dry	4 days					3 days			4 days			
	Dry	7 days								4 days		7 days	
	Normal	20 days								7 days			
	Humid	20 days											
Growing stage		Germi nation	Vegetative development					Flowering		Yield formation + ripening			
Sensitivity to water-stress		very (a)	moderate sensitive (b)				sensitive (c)	very sensitive (a)		sensitive (c)			not (d)

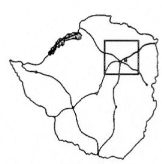
(a) The crop is extremely sensitive to water-stress during establishment.
Also during flowering water shortage should be avoided.
(b) During the early vegetative stage, the crop is moderate sensitive to water-stress and furthermore, root development is encouraged by a limited water supply.
(c) When the crop reaches its maximum crop development, the highest values of transpiration can be observed. This makes the crop very sensitive to water-stress. Also during yield formation sufficient water should be applied to avoid a decrease in yield.
(d) During ripening, less water is required.
Mind! Do not irrigate at the end of the day, so that leaves are dry when night falls. If they are wet during the night, Late Blight can occur!

Irrigation conditions

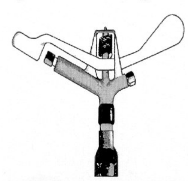
Growing season

- planting: begin December
- harvesting: end March

Region II



Sprinkler irrigation



75% field application efficiency

Example of an Irrigation Chart (front side) for potatoes cultivated in the region of Harare on a clay soil, presenting indicative irrigation intervals for four weather conditions (Holvoet, 2002).

Annex 5 – Structure of RAINBOW data files

The data files of RAINBOW are stored in the subdirectory 'DATA'. They are ASCII (text) files and have the extension 'DTA'.

The structure of the data file (Table A5) is illustrated for the example of the total annual rainfall for Bombay (Table A1.1):

- On the first line, the description of the data set (maximum length is 28 characters) as well as (from position 29 onwards) a set of 5 integer and 2 real values are specified. The values refer to selections made by the user at run time for all kinds of program parameters. The default setting for the set is ' 1' (normal distribution), ' 0' (no transformation), ' 3' (Weibull method for estimating cumulative probabilities), ' 1' (fitting a line through the points in the probability plot), ' 0' (no outliers), ' 0.0' (transformation constant), and ' 0.0' (value for nil). Since the program automatically updates the values in accordance to selections made by the user at run time, the initial setting of these parameters is of no importance.
- The first year and the last year of the record are specified respectively on the second and third line of the data file.
- The total annual rainfall depths for each year are listed from line four onwards till the end. Only one value is specified per line. Any postion on the line and any type of format (decimal point, scientific, …) is valid. The values can be integers (without decimal point) or reals. If data for a particular year is missing '-999' should be specified.

Table A5. Example of the data file containing the total annual rainfall for Bombay for the 37-year time period 1960-1996 (Table A1.1).

	Data string (description)	Program parameters
	← max. 28 characters →	from position 29 onwards
1st line	Bombay total rainfall	1 0 3 1 0 0.0 0.0
2nd line (from year)	1960	
3rd line (to year)	1996	
4th (1960 rainfall)	2119	
5th (1961 rainfall)	2205	
6th (1962 rainfall)	2029	
…	….	
31st (1987 rainfall)	1937	
32nd (1988 missing rainfall)	- 999	
33rd (1989 rainfall)	1900	
…	….	
38th (1994 rainfall)	2031	
39th (1995 missing rainfall)	-999	
40th (1996 missing rainfall)	-999	