551.577.31:551.501.45:519.216

MEDEDELINGEN LANDBOUWHOGESCHOOL
WAGENINGEN · NEDERLAND · 77-3 (1977)

STOCHASTIC MODELLING OF DAILY RAINFALL SEQUENCES

T. A. BUISHAND

Department of Mathematics and
Department of Land and Water Use,
Agricultural University, Wageningen, The Netherlands

(Received 1-X-1976)

H. VEENMAN & ZONEN B.V. - WAGENINGEN - 1977

Mededelingen Landbouwhogeschool
Wageningen 77-3 (1977)
Communications Agricultural University
(is also published as a thesis)

CONTENTS

I. INTRODUCTION

1. OBJECT AND SURVEY OF DATA

Decisions in water management are often based on rainfall series. From historical data it is possible to get some insight into problems about the amount of water available. The historic series can also be routed through a rainfall-runoff relation to obtain a streamflow series which can be used for planning water system projects. Working in this way gives a solution which is based on one realization of the rainfall process only. But what would be the solution if another series with the same properties as the historic series was used? Or more generally, how reliable is the solution? To answer these questions one must know the stochastic process underlying the sequence of rainfall data. This process, however, is very complicated; for instance, it usually exhibits seasonal variation and, when working within small time-increments, one encounters the problem of serial correlation and complicated marginal distributions. It is, therefore, often impossible to obtain direct analytical solutions for hydrological problems. For a better insight into a particular problem, one is usually forced to generate synthetic sequences based on a stochastic model for the rainfall process; and even that is often difficult.

The aim of this study is to construct a stochastic model for daily rainfall sequences. The time-increment of a day is chosen because rainfall is mostly recorded once a day and because a day seems a suitable choice for solving many problems in hydrology. It is necessary that the model is such that statistical simulation of synthetic sequences can easily be done.

Daily rainfall sequences are usually characterized by serial correlation and many observations with zero rainfall amount. It is the combination of these two facts which makes the generation of daily rainfall sequences complicated. When dealing with serial correlation only, one can apply, for instance, autoregressive models. But these models become very cumbersome if one is dealing with non-negative variables with a lot of zero values. It is also possible to fit theoretical distributions to daily data. For the Netherlands the fit of many distributions to daily rainfall data has been discussed by VAN MONTFORT (1968), but it is very hard to make a model with serial correlation and one of the marginal distributions, proposed by this author.

A widely used technique for handling daily rainfall series is to analyse first the occurrence of rain and non-rain days separately. In a second stage the behaviour of the non-zero rainfall amounts is studied. This technique will also be used here.

The process of rainfall occurrence can be taken in continuous time or in discrete time. Processes in continuous time have been discussed by GREEN (1964), TODOROVIC and YEVJEVICH (1969), QUÉLENNEC (1973) and KAVVAS and DELLEUR (1975). Working with processes in continuous time may have some drawbacks, namely:

a. The main body of rainfall data is given in units of one day. When a wet day is observed there can be more than one rainy period during such a day. Therefore, observing rainfall amounts for discrete time units can lead to quite another process. Moreover, when a model in continuous time is proposed it is often difficult to derive the statistical properties of daily rainfall amounts. This derivation is essential since estimates of the parameters should be based on the observed daily data.

b. There might be a daily cycle in the rainfall process. For instance, in some parts of the world rainfall only occurs during some fixed hours. To obtain a sound model such a diurnal variation should be incorporated.

Because of these disadvantages a rainfall process in discrete time is preferred here.

A great deal of this study deals with rainfall data from the Netherlands and the adjacent Belgian and German areas. The various rainfall stations which are taken into consideration are mentioned in Table 1.1. The geographical position

TABLE 1.1. Survey of Belgian, German and Dutch stations used in this study. The abbreviations between brackets are used in Figure 1.1.

Belgian stations:		
Ghent (Gt)	Moerbeke-Waas (MW)	Sint Andries-Brugge (SAB)
German stations:		
Ahaus (As)	Laar (Lar)	Ringenberg (Rg)
Bracht (Brt)	Lathen (Lan)	Schöppingen (Scn)
Dersum (Dm)	Leer (Ler)	Schüttorf (Sf)
Düren (Dn)	Lingen (Len)	Venhaus (Ves)
Herzogenrath (Hh)	Norden (Non)	Weener (Wer)
Jülich (Jh)	Norderney (Ny)	Widdelswehr (Wir)
Kleve (Ke)	Rheine (Re)	
Dutch stations:		
Aalten (Aan)	Groede (Ge)	Roggel (Rel)
Almelo (Ao)	Groningen (Gn)	Roodeschool (Rol)
Almen (Aln)	Haarlem (Hm)	Schaesberg (Srg)
Axel (Al)	Heino (Hno)	Scheveningen (Sen)
Biervliet (Bit)	Hellendoorn (Hn)	Schiermonnikoog (Sog)
Borculo (Bo)	Hengelo (Hlo)	Sint-Kruis (SK)
Cadzand (Cd)	Hoofddorp (Hp)	Stein (Stn)
Castricum (Cm)	Leiduin (Lin)	Ter Apel (TA)
De Bilt (DB)	Lettele (Le)	Terneuzen (Tn)
Delfzijl (Dl)	Leyden (Lyn)	Twente (Te)
Denekamp (Dp)	Lijnden (Ljn)	Vaals (Vas)
Den Helder (DH)	Lochem (Lm)	Valkenburg (Vg)
Deventer (Dr)	Nieuw-Beerta (NB)	Vroomshoop (Vp)
Dirksland (Dd)	Nijmegen (Nin)	Warffum (Wm)
Emmen (En)	Oldenzaal (Ol)	Winschoten (Wn)
Enschede (Ee)	Rekken (Rn)	Winterswijk (Wk)
Finsterwolde (Fe)	Roermond (Rd)	Zwanenburg (Zg)
Gouda (Ga)		

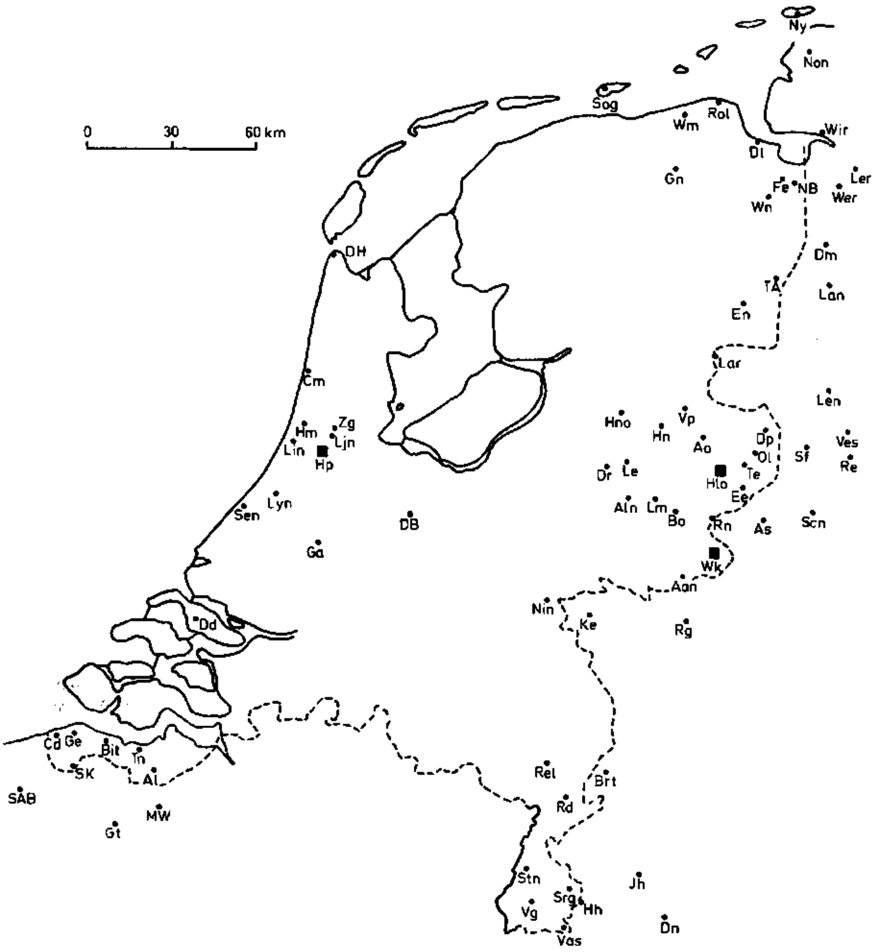


FIG. 1.1. Geographical position of Belgian, German and Dutch rainfall stations used in this study. The full names of the stations are given in Table 1.1.

of these stations is denoted in Figure 1.1. Attention will also be paid to some rainfall stations from other climatic regions, namely:

a. From India: Bangalore (12°58′ N, 77°35′ E), Calcutta (Alipore, 22°32′ N, 88°20′ E) and New Delhi (28°35′ N, 77°12′ E).
b. From Indonesia: Jakarta-27 (6°11′ S, 106°50′ E), Pasar Minggu (15 km south of Jakarta-27).
c. From Surinam: Paramaribo (5°51′ N, 55°10′ W), Domburg (5°42′ N, 55°05′ W).
d. From Sudan: Khartoum (15°37′ N, 32°33′ E).

e. From Egypt: Alexandria (Kôm el Nadûra, 32°12′ N, 29°53′ E).

Daily rainfall observations are usually given in tenths of millimeters. Two exceptions are:

a. Indonesian data. These are given in millimeters.
b. Indian data before 1958. These are given in hundredths of inches, but are converted to tenths of millimeters.

The source of the data is given in the last chapter (Chapter VI). This chapter also summarizes supplements of missing data.

When analysing rainfall series for a considerable number of years one must be aware that the recorded rainfall amounts have not always been obtained in the same way. There can be large differences in the mean rainfall amount due to changes in the way of measuring. If one neglects such non-homogeneities one gets a stochastic process which is not representative for the present situation but for a mixture of the many different situations in the past. Moreover, non-homogeneity can lead to a serious bias in estimates of parameters. Therefore, the problem of non-homogeneity is discussed in detail in Chapter II. In Chapter III the analysis of daily observations of Winterswijk, Hoofddorp and Hengelo is described. A stochastic model is developed and features of this model are compared with those of the historic series. Theoretical considerations about this model, based on the theory of stochastic processes, are given in Chapter IV. In this chapter formulas for the calculation of correlograms, variance-time curves and the cumulative distribution of k-day totals are derived. The application of the model to other climatic regions is presented in Chapter V.

2. Notation and abbreviations

Each chapter usually contains a number of sections. Formulas, tables and figures are numbered within these sections. For instance, (5.1) means Equation (1) of Section 5. Some chapters contain one or more appendices. These appendices are numbered within the chapter to which they belong; equations are numbered within appendices. That is (A3.2) means Equation (2) of Appendix A3. When reference is made to a formula, table or figure of another chapter, the number of this chapter is included. For instance, IV, (4.10) means Equation (10) of Section 4 of Chapter IV and II, (A4.3) means Equation (3) of Appendix A4 of Chapter II.

The rth moment about zero is denoted by μ'_r; the rth central moment by μ_r and the rth factorial moment by $\mu_{[r]}$. For the mean (μ'_1) usually the symbol μ is used only and the variance (μ_2) is often denoted by σ^2.

An estimate of a particular parameter is denoted by placing a caret above the parameter. So $\hat{\alpha}_2$ means an estimate of α_2. However, estimates of correlation coefficients (ρ) are denoted by r and estimates of the variance (σ^2) are usually denoted by s^2.

Random variables are underlined; $\underline{x} \simeq \underline{y}$ means that $\underline{x}$ and $\underline{y}$ are identically distributed. Logarithms are assumed to be to the base e (natural logarithms) and are denoted by log.

The major abbreviations are:

ML	maximum likelihood.
AML	approximate maximum likelihood.
LR	likelihood ratio.
OLS	ordinary least squares.
CL	critical level.

cdf	cumulative distribution function.
gf	generating function.
pgf	probability generating function.
scc	serial correlation coefficient.
Rel. freq.	relative frequency.
iid	independently and identically distributed.
NBD	negative binomial distribution.
SNBD	shifted negative binomial distribution.
TNBD	truncated negative binomial distribution.
GD	geometric distribution.
LSD	logarithmic series distribution.
LDF	'loi des fuites'.
SGD	shifted gamma distribution.

A, B	method of analysis in which a wet or dry spell is assigned to the period in which it begins (A) or ends (B). The capitals A and B are usually followed by the height (in tenths of millimeters) of the threshold defining a wet day (see III, 2).
WD, DW	method of analysis by wet-dry cycles (WD) or dry-wet cycles (DW). The capitals WD and DW are usually followed by the height (in tenths of millimeters) of the threshold defining a wet day (see III, 2).

Sd	dry season.
Sw	wet season.
Sdw	transition period from the dry to the wet season.
Swd	transition period from the wet to the dry season.

KNMI	Royal Netherlands Meteorological Institute (Koninklijk Nederlands Meteorologisch Instituut).

II. HOMOGENEITY OF DUTCH RAINFALL SERIES

1. Introduction

In this chapter the homogeneity of Dutch rainfall series is discussed. A rainfall series is called homogeneous if for each year rainfall on a particular calendar day or month is a realization of the same random variable. A homogeneous rainfall series is not necessarily a realization of a stationary stochastic process because it may exhibit seasonal variation. In fact the definition of a homogeneous series concerns the whole probability distribution of rainfall amounts. In practice, however, homogeneity of the mean is considered only, since departures from homogeneity in higher order moments can hardly be detected because of large sample variations.

Non-homogeneity can be a consequence of a gradual change in the meteorological situation, but can also be purely man-made, e.g. due to changes of site, or to changes in instructions to observers. Here it is assumed that departures from homogeneity are man-made and therefore non-homogeneities in the mean usually consist of jumps.

Tests for homogeneity were done with annual or monthly totals. Methods for testing homogeneity often make assumptions about the distribution of the rainfall data. Therefore the distribution of annual and monthly totals is discussed in Sections 2 and 3.

The probability of success in detecting jumps in the mean of a given rainfall series depends on the situation of neighbouring rainfall stations. For instance, when some station changes its way of measuring, the best way to study the effect of such a change is to compare the rainfall series with that of another station in the direct neighbourhood with no changes. Even small departures from homogeneity can be detected if the two stations are close together. An example of how a jump is determined by comparing altered rainfall stations with unaltered rainfall stations is given in Section 4. The problem discussed in that section is a possible jump in the mean of Dutch rainfall series due to a change in height of the rain gauge, during the period 1946–1954. The significance of a jump is tested by comparing rainfall series in the Netherlands with those of neighbouring countries.

It often happens, however, that the homogeneity of rainfall series of neighbouring stations is also doubtful or that there is no neighbouring rainfall station at all. Then only use can be made of a single rainfall series and many possible jumps will be passed unnoticed. An analysis of homogeneity making use of only one rainfall series is described in Section 5, where homogeneity of the Zwanenburg-Hoofddorp series is investigated.

When using a separate process for the occurrence of wet and dry days the homogeneity of the sequence of wet and dry days is also important. Statistical methods for detecting jumps in such a situation are given in Section 6.

2. The distribution of annual totals

When rainfall is recorded within small time increments (a day or shorter), the marginal distribution of the rainfall amounts is markedly skew, because zero values and some very large values occur with relatively large probability. Rainfall amounts over longer periods are less skew because of the effect of the central limit theorem and it is well known that the skewness of annual totals is so small that they can be assumed to be approximately Gaussian (cf. de Boer (1956, 1958)). Here a statistical support for this assumption is given on the basis of annual data of some Dutch and German rainfall series. Though many tests for normality exist only a few of them are considered here, namely a test on the coefficient of skewness, the Shapiro-Wilk test, the Kolmogorov-Smirnov test and the Kuiper test.

The coefficient of skewness γ is defined by:

$$(2.1) \qquad \gamma = \mu_3/\mu_2^{3/2}.$$

An estimate of γ can be obtained by replacing the central moments in the right side of (2.1) by unbiased sample estimates (cf. Ven te Chow (1964), 8-I-II C3):

$$(2.2) \qquad \hat{\gamma} = \frac{N\sqrt{N-1}}{N-2} \frac{\sum_{i=1}^{N} (x_i-\bar{x})^3}{\left\{\sum_{i=1}^{N} (x_i-\bar{x})^2\right\}^{3/2}} = \frac{\sqrt{N(N-1)}}{N-2} \sqrt{b_1}$$

with N: the number of data,
$\bar{x}$: the sample mean.

For a one-sided test the upper and lower 5 and 1 per cent critical values of $\underline{\hat{\gamma}}$ can be obtained from the corresponding percentage points of the $\sqrt{\underline{b}_1}$ statistic, given by Pearson and Hartley (1962, Table 34B). One can also use the normal approximation of the $\sqrt{\underline{b}_1}$ statistic given by D'Agostino (1970).

The test based on $\hat{\gamma}$ is only sensitive to skewed alternatives; the other tests given here are sensitive to many different kind of alternatives (so called omnibus tests).

The Shapiro-Wilk test is based on the ratio of the best linear unbiased estimate of the standard deviation calculated from an ordered sample to the sample standard deviation (Shapiro and Wilk (1965)).

Let $\mathbf{x}$ denote the vector of ordered observations

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(N)} \text{ and let } \mathbf{m} = (m_1, m_2, \ldots, m_N)'$$

denote the vector of expected values of standard normal order statistics. For the Shapiro-Wilk test one starts with the regression equation:

$$(2.3) \qquad \underset{\sim}{x}_{(i)} = \mu + \sigma m_{(i)} + \sigma \underset{\sim}{e}_i \qquad i = 1, \ldots, N$$

where μ and σ are unknown parameters for location and scale respectively. The error terms $\underset{\sim}{e}_i$ are assumed to have mean zero and covariance matrix $\mathbf{V}$.

The method of generalized least squares gives as an estimate of σ:

$$(2.4) \qquad \hat{\sigma} = \frac{\mathbf{m}'\mathbf{V}^{-1}\mathbf{x}}{\mathbf{m}'\mathbf{V}^{-1}\mathbf{m}}$$

which is the best linear unbiased estimate of σ based on the ordered sample. Equation (2.4) can be written as:

$$(2.5) \qquad \hat{\sigma} = \sqrt{\mathbf{b}'\mathbf{b}}\, \mathbf{a}'\mathbf{x}$$

with $\qquad \mathbf{b} = \dfrac{\mathbf{V}^{-1}\mathbf{m}}{\mathbf{m}'\mathbf{V}^{-1}\mathbf{m}}$

and $\qquad \mathbf{a} = \dfrac{\mathbf{V}^{-1}\mathbf{m}}{(\mathbf{m}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{m})^{1/2}} \quad \left(= \dfrac{\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{b}}}\right).$

The test statistic is:

$$(2.6) \qquad W = \frac{(\mathbf{a}'\mathbf{x})^2}{\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2 / N}.$$

For normal samples W is close to its maximum value 1; for non-normal samples W tends to smaller values.

The elements of the vector **a** in (2.5) and (2.6) depend on first and second moments of standard normal order statistics. The expectations can be obtained from HARTER (1961), but variances and covariances (the elements of **V**) are more difficult to obtain, especially for large N. Therefore, SHAPIRO and FRANCIA (1972) proposed to base the numerator of (2.6) on the ordinary least squares (OLS) estimate of σ, that is one has to substitute the identity matrix **I** for the matrix **V** in (2.4). They gave percentage points of the null distribution of the modified statistic W' for $N = 35, 50, 51(2)99$. The lengths of most annual rainfall series under investigation lie in this range. The table by SHAPIRO and FRANCIA (1972) has been extrapolated for series which are a bit longer than 99 years.

The Kolmogorov-Smirnov test and the Kuiper test are based on differences between the empirical and the theoretical distribution function. The empirical distribution function $F_N(x)$ is defined as:

$$(2.7) \qquad F_N(x) = \frac{\text{number of observations} \leqslant x}{N}$$

where N is the sample size.

Denote the cumulative distribution function by $F_0(x)$ and define:

$$(2.8a) \qquad D^+ = \sup_{-\infty < x < \infty} (F_N(x) - F_0(x)) = \max_{1 \leqslant i \leqslant N} \left\{ \frac{i}{N} - F_0(x_{(i)}) \right\}$$

$$(2.8b) \qquad D^- = \sup_{-\infty < x < \infty} (F_0(x) - F_N(x)) = \max_{1 \leqslant i \leqslant N} \left\{ F_0(x_{(i)}) - \frac{i-1}{N} \right\}$$

where $x_{(i)}$ denotes, as before, the ith order statistic of the sample.

The Kolmogorov-Smirnov statistic D is defined by:

$$(2.9) \qquad D = \max(D^+, D^-)$$

and the Kuiper statistic K is defined by:

$$(2.10) \qquad K = D^+ + D^-.$$

TABLE 2.1. Mean, standard deviation and realizations of test statistics for tests for normality of annual data of some Dutch and German rainfall series. Realizations of test statistics which are significant at the 5 per cent level are denoted by an asterisk.

Rainfall station	Period	Mean (mm)	Standard deviation (mm)	$\hat{\gamma}$	W'	$D\sqrt{N}$	$K\sqrt{N}$
Norderney	1881–1973	701	113	–0.015	0.992	0.554	1.139
Leer	1891–1970	744	106	–0.840	0.948*	0.851	1.388
Weener	1897–1970	735	108	–0.326	0.982	0.605	1.202
Laar	1903–1970	710	123	–0.050	0.992	0.401	0.915
Lingen	1855–1973	743	123	0.075	0.990	0.622	1.258
Rheine	1891–1970	746	128	–0.149	0.994	0.379	0.843
Ahaus	1891–1970	789	128	–0.051	0.988	0.615	1.196
Ringenberg	1893–1970	743	132	–0.114	0.986	0.582	1.071
Kleve	1851–1972	779	131	–0.120	0.985	0.532	1.147
Jülich	1894–1970	636	119	0.147	0.982	0.723	1.447
Herzogenrath	1894–1970	782	150	1.085*	0.934*	0.951*	1.415
Delfzijl	1872–1970	721	107	–0.231	0.989	0.529	0.998
Warffum	1893–1970	723	117	–0.140	0.991	0.579	1.129
Ter Apel	1892–1972	711	123	0.337	0.962*	0.504	1.093
Finsterwolde	1892–1972	687	117	0.150	0.984	0.695	1.317
Winschoten	1923–1972	751	110	–0.554	0.973	0.588	1.108
Enschede	1881–1972	751	122	0.332	0.964*	0.883	1.632*
Hengelo	1887–1972	748	134	0.390	0.975	0.544	0.930
Winterswijk	1880–1972	761	127	0.098	0.969*	0.895*	1.806*
Valkenburg	1904–1972	773	130	0.383	0.982	0.636	1.068
Roermond	1869–1972	657	117	0.116	0.992	0.588	1.216
Schaesberg	1909–1972	754	123	–0.084	0.975	0.477	1.027

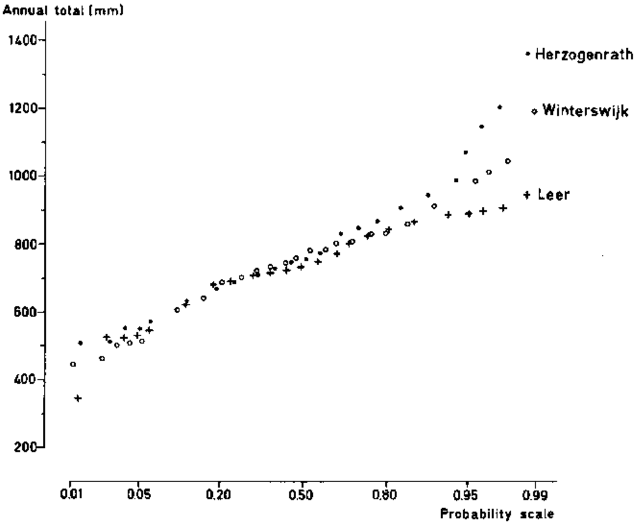


FIG. 2.1. Normal probability plots of annual totals of Leer, Herzogenrath and Winterswijk.

The tabulated percentage points of the null distribution of these statistics are only applicable if $F_0(x)$ is completely specified, which is not so here because mean and variance have to be estimated from the sample. Then percentage points, obtained by Monte Carlo simulation, are given by LILLIEFORS (1967) for the statistic $\underline{D}$ and by LOUTER and KOERTS (1970) for the statistic $\underline{K}$. Percentage points for $\underline{D}$ and $\underline{K}$ are also given in Table 54 (Case 2) of PEARSON and HARTLEY (1972). For large values of N the critical value at the 5 per cent level is approximately $0.886/\sqrt{N}$ for the Kolmogorov-Smirnov statistic and $1.450/\sqrt{N}$ for the Kuiper statistic which is about 65 and 85 per cent, respectively, of the commonly tabulated values in the situation of known parameters.

Table 2.1 shows realizations of the statistics $\hat{\gamma}$, W', $\underline{D}\sqrt{N}$ and $\underline{K}\sqrt{N}$ for rainfall series which will be used again for the analysis of homogeneity in Section 4. The denominator of the estimator of the standard deviation is $N-1$ in Table 2.1, because tables of percentage points of the statistics $\underline{D}$ and $\underline{K}$ are also based on this estimator. Realizations of the test statistic which are significant at the 5 per cent level are denoted by an asterisk. The test based on the coefficient of skewness is one-sided (test on positive skewness); annual totals of the station of Leer have a negative coefficient of skewness with a critical level of more than 0.99 and hence a negative coefficient of skewness seems to be possible. In general there is no evidence for departures from normality. Normal probability plots of annual totals of Leer, Herzogenrath and Winterswijk (see Figure 2.1) show that departures from normality are caused by extremely high values (Herzogenrath, Winterswijk) or extremely low values (Leer).

It is well-known that the Shapiro-Wilk test is more powerful than tests based on differences between the empirical and the theoretical distribution function. This is illustrated in Table 2.1, the Shapiro-Wilk test giving the larger number of significant values.

3. THE DISTRIBUTION OF MONTHLY TOTALS

Figure 3.1 shows estimates of the monthly mean and standard deviation for a number of stations with observations over a long period. The monthly mean reaches its minimum in February, March or April and its maximum in July or August, though stations near the west coast (Hoofddorp) also have a high October mean. Another feature of the monthly mean is the comparatively low September value. Stations remote from the coast (Enschede, Winterswijk, Roermond) are characterized by high standard deviations in February, July and August. The high February standard deviation is mainly caused by the high monthly total of February 1946. For coastal stations there is a nearly sinusoidal change of the standard deviation. The coefficient of variation is nearly constant (about 0.5) during the year.

In Section 3.1 a possible serial correlation of monthly totals is investigated. The marginal distribution of monthly totals is discussed in Section 3.2.

3.1. *Serial correlation of monthly totals*

Because there is a seasonal change in mean and standard deviation, the original totals x were standardized to u, with the formula:

$$(3.1) \qquad u_{12\,l+m} = \frac{x_{12\,l+m} - \bar{x}_m}{s_m}$$

with m : index of the month (1, ..., 12),
l : index of the year $(0, \ldots, n{-}1)$, n being the number of years,
$\bar{x}_m$: mean rainfall amount of month m,
s_m^2 : traditional (unbiased) estimate of the variance of the rainfall amount of month m.

A test for serial correlation can be based on the serial correlation coefficient (scc). The lag k scc was estimated by

$$(3.2) \qquad r_k = \frac{\sum_{i=1}^{N-k} (u_i - \bar{u})(u_{i+k} - \bar{u})}{\sum_{i=1}^{N} (u_i - \bar{u})^2}$$

where $N = 12 \times n$, the number of observations,
$\bar{u}$: mean of the N u_is.

Using $\bar{u} = 0$ and $\sum_{i=1}^{N} (u_i - \bar{u})^2 = N - 12$, Equation (3.2) becomes

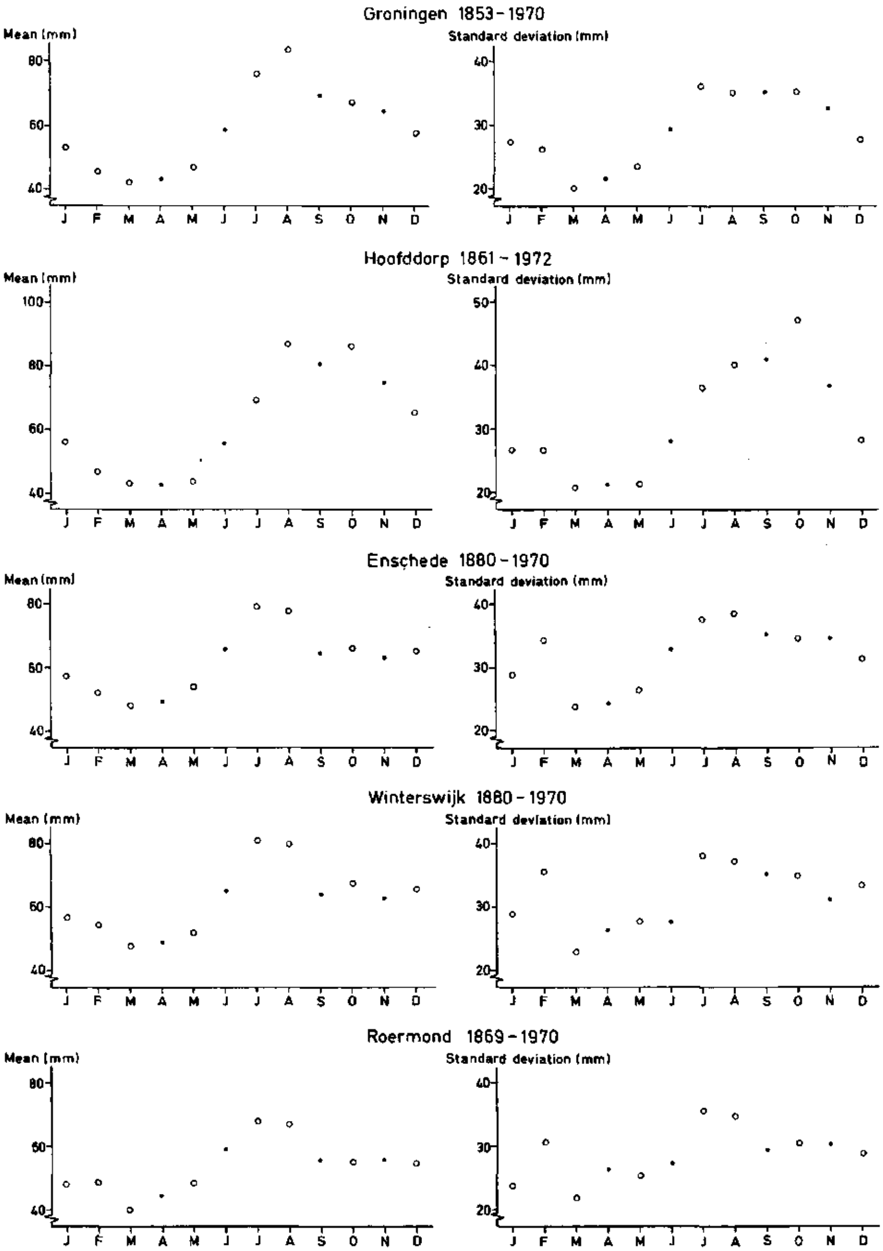


FIG. 3.1. Estimates of the mean and standard deviation of monthly totals of some long-term Dutch rainfall series. Values denoted by an open dot are proportional values for a 30-day period.

(3.3)
$$r_k = \frac{\sum_{i=1}^{N-k} u_i u_{i+k}}{N-12} .$$

For sufficiently large N (N about 75) the distribution of $\underline{r}_k \sqrt{N}$ is approximately standard normal if the observations are independent and normally distributed (cf. JENKINS and WATTS (1969), 5.3.5). Moreover, then the different $\underline{r}_k$s are approximately uncorrelated. Therefore, a rough test for serial correlation can be based on the statistic:

(3.4)
$$\underline{X}^2 = N \sum_{k=1}^{\nu} \underline{r}_k^2$$

which is approximately a $\underline{\chi}_\nu^2$-variable (chi-square with ν degrees of freedom) under the null hypothesis.

What should one do when dealing with non-normal data, as is the case here (see 3.2). BARTLETT (1946) pointed out that the asymptotic variances and covariances of the $\underline{r}_k$s do not depend on the marginal distribution. Moreover, the joint distribution of the $\underline{r}_k$s for normal data seems often to be a good approximation when dealing with non-normal variables (cf. YEVJEVICH (1972), Section 2.2). However, one should be very careful in applying the test to non-normal data because the convergence of the test statistic to its asymptotic distribution can be very slow. It is, therefore, advisable to repeat the test with a normalizing transformation on the data.

One can also base the test on r_1 alone or equivalently on the Von Neumann's ratio, which is defined as:

(3.5)
$$d = \frac{1}{2} \frac{\sum_{i=1}^{N-1} (u_{i+1} - u_i)^2}{\sum_{i=1}^{N} (u_i - \bar{u})^2} .$$

From (3.2) and (3.5) it is verified that:

(3.6)
$$d \approx 1 - r_1$$

(cf. SNEYERS (1957)).

Realizations of the test statistics mentioned above and their corresponding critical levels are given in Table 3.1 for the Hoofddorp and Winterswijk series. The test based on the statistic X^2 has been repeated with the square roots of the monthly totals which are approximately normally distributed (see Section 3.2). There is no evidence for serial correlation at the 5 per cent level either for transformed or untransformed data (the critical levels are always larger than 0.05). Taking the values 3,6 or 12 for ν in (3.4) leads to the same conclusions.

TABLE 3.1. Realizations of test statistics and their corresponding critical level (C.L.) for tests on serial correlation of monthly totals. The statistic X^2 in (3.4) is based on $\nu = 36$.

Rainfall series	Original data				Transformed data	
	d	C.L.	X^2	C.L.	X^2	C.L.
Winterswijk 1880–1970	0.971	0.141	22.70	0.959	28.73	0.800
Hoofddorp 1861–1972	0.962	0.105	40.44	0.280	41.76	0.235

3.2. *The marginal distribution of monthly totals*

In contrast with annual totals monthly totals have a markedly skew distribution. The monthly mean of the coefficient of skewness is 0.774 for the Winterswijk series and 0.562 for the Hoofddorp series and it can be shown from Table 34B of PEARSON and HARTLEY (1962) that these values show evidence for a positive skewness at the 5 per cent level.

There are many distributions which are positively skewed. Two of them will be examined in more detail, namely the gamma distribution and a distribution which will be denoted as 'loi des fuites' (LDF).

The gamma variable $\underline{\gamma}(\lambda,\nu)$ is defined by its probability density:

$$(3.7) \qquad f(x) = \frac{\lambda^\nu x^{\nu-1} e^{-\lambda x}}{\Gamma(\nu)} \qquad x>0,\ \lambda>0,\ \nu>0$$

where Γ stands for the gamma function. The parameter λ is a scale parameter and the parameter ν is a shape parameter.

If $\nu > 1$ it follows by differentiation of (3.7) that there is a mode at $(\nu-1)/\lambda$. If $\nu < 1$ the density is J-shaped and is infinite at the origin. For $\nu = 1$ one gets the exponential distribution, which has probability density:

$$(3.8) \qquad f(x) = \lambda e^{-\lambda x}.$$

Another special case of the gamma variable is the $\underline{\chi}^2_n$-variable, namely:

$$(3.9) \qquad \underline{\chi}^2_n \simeq \underline{\gamma}(\tfrac{1}{2},\tfrac{1}{2}n).$$

Moments of the gamma variable are:

$$(3.10a) \qquad \mu_1' = \nu/\lambda$$

$$(3.10b) \qquad \mu_2 = \nu/\lambda^2$$

$$(3.10c) \qquad \mu_3 = 2\nu/\lambda^3$$

$$(3.10d) \qquad \mu_4 = (6\nu + 3\nu^2)/\lambda^4$$

$$(3.10e) \qquad C = \sqrt{\mu_2}/\mu_1' = 1/\sqrt{\nu} \qquad (C \text{ is the coefficient of variation})$$

$$(3.10f) \qquad \gamma = 2/\sqrt{\nu}.$$

From (3.10e and f) it follows that the quotient γ/C is always 2, irrespective of the parameters of the distribution.

A normalizing transform of the gamma variable is the Wilson-Hilferty transform (cf. KENDALL and STUART (1969), 16.7):

$$(3.11) \qquad 3\sqrt{\nu}\left\{\left(\frac{\lambda\underline{\gamma}(\lambda,\nu)}{\nu}\right)^{1/3} - 1 + \frac{1}{9\nu}\right\}$$

which is asymptotically standard normal. The third central moment of the transformed variable is of order ν^{-3}; so the Wilson-Hilferty transform may give a good normal approximation when the shape parameter is large.

Estimates of ν and λ can be obtained, for example, by the method of moments or the method of maximum likelihood (ML). The moment estimates are:

$$(3.12a) \qquad \hat{\lambda} = \bar{x}/s^2$$

$$(3.12b) \qquad \hat{\nu} = \bar{x}^2/s^2$$

where $\bar{x}$ and s^2 are the sample mean and variance (unbiased version), respectively.

The ML estimate $\hat{\nu}$ of the parameter ν follows from (cf. THOM (1958)):

$$(3.13a) \qquad \psi(\hat{\nu}) - \log\hat{\nu} = \frac{1}{N}\sum_{i=1}^{N}\log x_i - \log\bar{x}$$

where N is the number of observations and ψ stands for the digamma function, which is the first derivative of the logarithm of the gamma function. The iterative solution of (3.13a) was described by CHOI and WETTE (1969). An initial estimate of ν can be based on the moment estimate or an approximative solution of (3.13a), e.g. the one given by THOM (1958) or the one given by GREENWOOD and DURAND (1960). The initial estimate given by CHOI and WETTE (1969) is only a simplified form of Thom's estimate.

After a solution of (3.13a) has been found, the ML estimate of the scale parameter can be obtained from:

$$(3.13b) \qquad \hat{\lambda} = \hat{\nu}/\bar{x}.$$

As a measure for the efficiency of a moment estimator with respect to a ML estimator one can take the ratio of the large-sample variances of the ML and the moment estimator. Because for large samples there is no other estimator with smaller variance than the ML estimator, this ratio is called the asymptotic efficiency of the moment estimator. Expressions for the large-sample variances of moment and ML estimators of the parameters of the gamma distribution are given in Appendix A1. The asymptotic efficiencies of the moment estimators of λ and ν are given in Table 3.2. Notice from that table that the asymptotic efficiency only depends on ν. For small values of ν (skew distributions) the method of moments gives very inefficient estimates.

A general measure for the asymptotic efficiency of the method of moments is the ratio of the determinants of the large-sample covariance matrices of the ML and moment estimators. For the gamma distribution this ratio equals the

TABLE 3.2. Asymptotic efficiency of the moment estimators of the gamma distribution as a function of the shape parameter.

	Estimator of			Estimator of	
ν	λ	ν	ν	λ	ν
0.1	0.347	0.050	2.0	0.636	0.575
0.2	0.363	0.098	3.0	0.712	0.676
0.3	0.382	0.144	4.0	0.763	0.739
0.4	0.401	0.187	5.0	0.798	0.782
0.5	0.420	0.227	6.0	0.825	0.812
0.6	0.440	0.264	7.0	0.845	0.835
0.7	0.458	0.299	8.0	0.861	0.853
0.8	0.476	0.331	9.0	0.874	0.868
0.9	0.494	0.360	10.0	0.885	0.880
1.0	0.510	0.388	100.0	0.987	0.987

asymptotic efficiency of ν, which follows from the formulas for the large-sample variances and covariances in Appendix A1.

The second probability distribution which is considered here can be described as follows. Suppose that rainfall occurs in instantaneous showers according to a Poisson process with mean intensity or rate $1/\mu$, that is the number of showers in a time interval with length t is Poisson distributed with mean t/μ. Rainfall amounts of single showers are assumed to be:

a. Independent of the process of occurrence.

b. Mutually independent.

c. Exponentially distributed with mean $1/\rho$.

The process described here was suggested as a model for rainfall over arid regions by FISHER and CORNISH (1960). BERNIER and FANDEUX (1970) applied this process successfully to fit the distribution of monthly totals of French rainfall series and because it was used earlier to describe the distribution of escape flows of gas conduits they called the distribution of a Poisson distributed sum of iid exponential variables the 'loi des fuites' (iid stands for independently and identically distributed). This name will also be used here and will be abbreviated as LDF.

DE BOER (1956, 1957, 1958) applied a slight modification of the LDF to describe the distribution of rainfall totals over a period of at least 30 days by taking a constant rainfall amount for each shower instead of exponentially distributed rainfall amounts.

Let $\underline{x}_t$ be the total rainfall amount in a period of length t. The probability distribution of $\underline{x}_t$ is derived in Appendix A2. For the derivation of the moments of $\underline{x}_t$ use can be made of the moment generating function (cf. COX (1962), Equation (8.3.4)):

(3.14) $$f(s) = E(e^{-s\underline{x}_t}) = e^{-\theta} \exp\left(\frac{\rho\theta}{\rho + s}\right)$$

with $\theta = t/\mu$. From (3.14) it follows:

(3.15) $$\log f(s) = \theta \left(-1 + \frac{1}{1 + s/\rho}\right) = \theta \sum_{m=1}^{\infty} (-1)^m (\tfrac{s}{\rho})^m.$$

On the other hand:

(3.16) $$\log f(s) = \sum_{m=1}^{\infty} (-1)^m \varkappa_m \frac{s^m}{m!}$$

where $\varkappa_m$ is, by definition, the mth cumulant of $\underline{x}_t$. So $\varkappa_m$ satisfies the relation:

(3.17) $$\varkappa_m = \frac{m!\theta}{\rho^m}$$

(cf. FISHER and CORNISH (1960)).

From (3.17) expressions for moments and central moments can be obtained:

(3.18a) $\mu_1' = \varkappa_1 = \theta/\rho$

(3.18b) $\mu_2 = \varkappa_2 = 2\theta/\rho^2$

(3.18c) $\mu_3 = \varkappa_3 = 6\theta/\rho^3$

(3.18d) $\mu_4 = \varkappa_4 + 3\mu_2^2 = (24\theta + 12\theta^2)/\rho^4$

(3.18e) $C = \sqrt{2/\theta}$

(3.18f) $\gamma = 3/\sqrt{2\theta}$.

From (3.18e and f) it follows that the quotient γ/C is always 1.5, irrespective of the parameters of the distribution.

A normalizing transform of the LDF is:

(3.19) $$\sqrt{2\rho}\,\{\sqrt{\underline{x}_t} - \sqrt{\theta/\rho}\}$$

which is asymptotically standard normal. It can be shown that the third central moment of the transformed variable is of order $1/\theta^3$, so the transformation may give a good normal approximation when θ is large. For monthly totals of French rainfall series the approximation (3.19) works quite well (cf. BERNIER and FANDEUX (1970)).

The moment estimates of the parameters ρ and θ follow from the equations:

(3.20a) $\hat{\rho} = 2\bar{x}/s^2 \quad (= 2\hat{\lambda})$

(3.20b) $\hat{\theta} = 2\bar{x}^2/s^2 \; (= 2\hat{\nu})$.

So the moment estimates of the parameters of the LDF differ only a factor from those of the gamma distribution.

Estimation of the parameters of the LDF by the ML method is complicated. The likelihood equations and their solution are given in Appendix A3.

Table 3.3 gives estimates of the parameters of the gamma distribution and of the LDF. The estimate $1/\hat{\mu}$ of the LDF was obtained from $\hat{\theta}$ by assuming t to be equal to the number of days of the month (for February t was set equal to 28.2). The magnitude of the estimated parameters changes considerably from month to month, which is partly due to their large standard deviations. For instance, for the Winterswijk series the monthly mean of the standard deviation of moment estimates of ρ and $1/\mu$ is 0.020 mm^{-1} and 0. 040 days^{-1}, respectively, which can be obtained from (A1.6a and b). ML estimates of ρ and $1/\mu$ have a somewhat smaller standard deviation, namely 0.018 mm^{-1} and 0.037 days^{-1},

TABLE 3.3. Estimates of the parameters of the gamma distribution and the LDF.

	Winterswijk (1880–1970)							
	Gamma distribution				LDF			
	$\hat{\nu}$		$\hat{\lambda}$ (mm^{-1})		$1/\hat{\mu}$ (days^{-1})		$\hat{\rho}$ (mm^{-1})	
Month	Moments	ML	Moments	ML	Moments	ML	Moments	ML
January	3.84	3.12	0.064	0.052	0.248	0.233	0.128	0.121
February	2.26	2.28	0.044	0.045	0.160	0.190	0.089	0.105
March	4.20	3.96	0.084	0.079	0.271	0.278	0.168	0.173
April	3.37	2.90	0.069	0.059	0.225	0.228	0.138	0.140
May	3.61	3.53	0.066	0.065	0.233	0.248	0.132	0.141
June	5.57	5.29	0.086	0.081	0.372	0.375	0.171	0.173
July	4.65	4.20	0.054	0.049	0.300	0.301	0.108	0.109
August	4.18	3.39	0.053	0.043	0.270	0.257	0.105	0.100
September	3.35	3.02	0.052	0.047	0.223	0.230	0.103	0.106
October	3.71	3.00	0.053	0.043	0.240	0.227	0.106	0.101
November	3.97	3.95	0.062	0.062	0.265	0.282	0.124	0.132
December	3.85	3.25	0.056	0.047	0.248	0.248	0.112	0.112
	Hoofddorp (1861–1972)							
January	4.07	3.93	0.072	0.069	0.263	0.273	0.144	0.150
February	2.90	2.37	0.068	0.055	0.206	0.203	0.135	0.133
March	4.42	4.22	0.096	0.092	0.285	0.293	0.193	0.198
April	3.94	3.28	0.093	0.077	0.262	0.256	0.186	0.181
May	4.02	3.98	0.088	0.087	0.260	0.275	0.176	0.187
June	3.94	3.47	0.071	0.062	0.262	0.260	0.142	0.140
July	3.60	3.00	0.049	0.041	0.232	0.222	0.099	0.095
August	4.73	3.88	0.052	0.043	0.305	0.286	0.104	0.098
September	3.92	3.06	0.048	0.038	0.262	0.241	0.096	0.089
October	3.32	2.44	0.037	0.027	0.214	0.194	0.074	0.067
November	4.04	4.11	0.054	0.055	0.270	0.292	0.108	0.118
December	5.35	3.88	0.078	0.057	0.345	0.299	0.156	0.135

respectively. The method for obtaining standard deviations of moment and ML estimates is given in Appendix A1.

The monthly mean of the ratio of the determinants of the estimated covariance matrices of the ML and moment estimators is 0.80 for both the Winterswijk and Hoofddorp series. For the gamma distribution this ratio is 0.71.

For the gamma distribution a honest comparison of the estimates of different months is not possible because these estimates are not corrected for the fact that different months can have different lengths. Another disadvantage of the gamma distribution is that its tail is too long for monthly data. This fact can be shown by computing moment estimates of the ratio γ/C, which has to be 1.5 for the LDF and 2 for the gamma distribution. The monthly mean of the estimate of this ratio is 1.46 for Winterswijk and 1.11 for Hoofddorp; only for the month of February of Winterswijk is this ratio larger than 2, namely 2.80. Perhaps this result explains why for the gamma distribution the ML estimate of the variance (obtained by substituting ML estimates in the right side of (3.10b)) tends to be larger than the moment estimate. This is seen from Table 3.4 where moment and ML estimates of the standard deviation are compared. The estimates of the standard deviation of the annual totals were obtained from the summation of monthly variance estimates. There is a good correspondence between moment estimates and ML estimates, when a LDF is assumed, but the ML estimate under assumption of a gamma distribution is larger in nearly all cases.

Cumulative frequencies of monthly totals and theoretical values, based on fitted distributions (gamma distribution, LDF), are compared in Figure 3.2.

TABLE 3.4. Comparison of different estimates of the standard deviation (in mm) of monthly totals.

	Winterswijk 1880–1970			Hoofddorp 1861–1972		
Month	Moments	ML gamma distribution	ML LDF	Moments	ML gamma distribution	ML LDF
January	30.5	33.9	31.4	28.0	28.5	27.5
February	33.8	33.7	31.1	25.2	27.9	25.4
March	24.3	25.0	24.0	21.8	22.3	21.5
April	26.7	28.9	26.5	21.4	23.4	21.6
May	28.7	29.0	27.8	22.7	22.8	22.1
June	27.6	28.3	27.5	28.0	29.9	28.2
July	39.8	41.8	39.7	38.4	42.0	39.2
August	38.9	43.2	39.9	41.8	46.2	43.2
September	35.4	37.3	34.9	41.0	46.5	42.8
October	36.2	40.3	37.2	49.1	57.2	51.6
November	32.1	32.2	31.1	37.0	36.8	35.6
December	34.9	38.0	34.9	29.6	34.8	31.8
Year	113.4	120.4	112.8	115.4	126.3	117.4

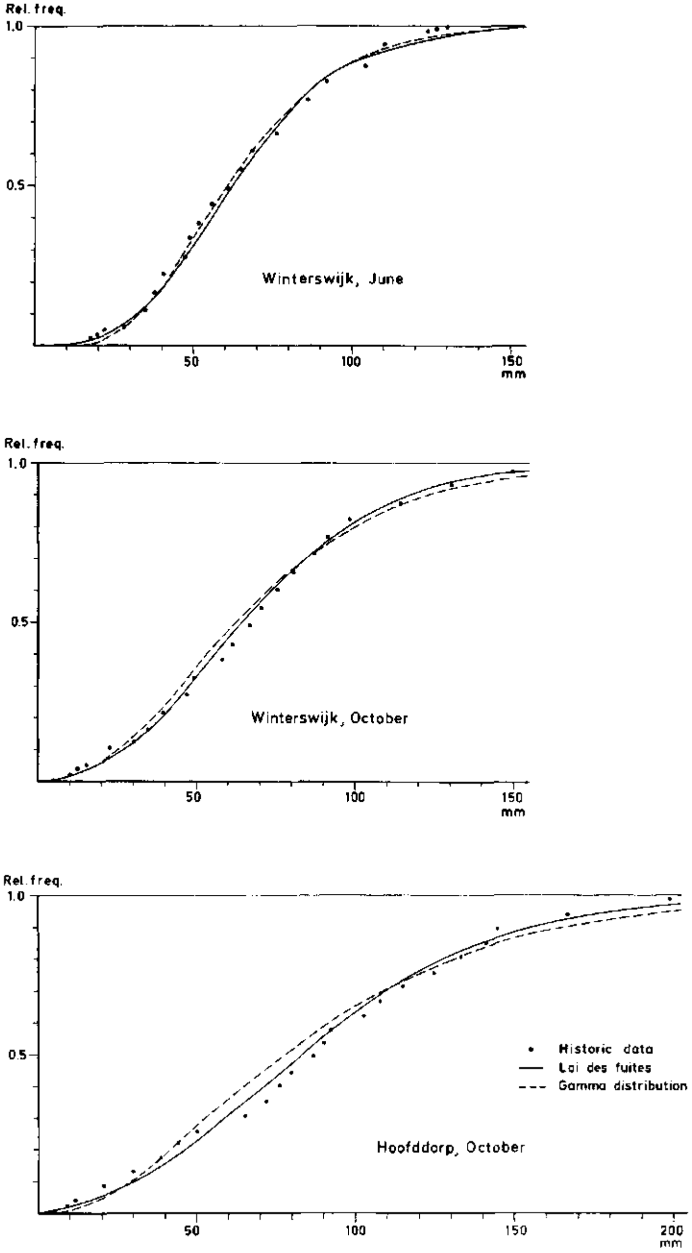


FIG. 3.2. Cumulative frequencies of monthly totals of Winterswijk (1880–1970) and Hoofddorp (1861–1972) and theoretical cumulative distribution functions.

TABLE 3.5. Critical levels of the X^2-test of goodness of fit for the LDF fitted to monthly totals.

Month	Winterswijk 1880–1970	Hoofddorp 1861–1972
January	0.216	0.158
February	0.346	0.137
March	0.765	0.374
April	0.416	0.316
May	0.117	0.481
June	0.573	0.983
July	0.601	0.551
August	0.249	0.076
September	0.148	0.299
October	0.519	0.095
November	0.249	0.016
December	0.601	0.394

The theoretical curves are based on ML estimates. The difference between the cumulative distribution functions of the gamma distribution and of the LDF is usually small, except for Hoofddorp, October. For this month the LDF gives a slightly better fit.

So the LDF could be preferred to the gamma distribution for fitting the distribution of monthly totals.

The critical levels (C.L.) of the X^2-test of goodness of fit for the LDF are given in Table 3.5.

For application of the test, the carrier of the distribution of the monthly totals was divided into classes in such a way that the expected cell frequency was the same for all classes and was as small as possible but at least 7. The expected cell frequencies were calculated with ML estimates based on the actual data instead of ML estimates based on observed cell frequencies; therefore the approximation of $\underline{X}^2$ at H_0 by $\underline{\chi}^2$ with parameter equal to the number of classes minus 3 gives a somewhat progressive test (cf. WATSON (1958) and HERMANS (1969)).

From the tabulated critical levels it is seen that the LDF fits the data well.

An attractive property of the LDF is that it can easily fit data with a high fraction of zeroes and therefore application of the LDF to monthly totals of stations with an arid or monsoon climate gives no special problems. An example is given in Figure 3.3 where the LDF is fitted to monthly totals of Bangalore (1879–1970).

The critical levels of the X^2-test of goodness of fit are 0.238, 0.002 and 0.414 for February, April and July, respectively; the ratio $\hat{\vartheta}/\hat{C}$ is 1.31, 1.23 and 1.89 for these months. The poor fit for the month of April is caused by the strange shape of the empirical distribution function. It may be assumed that most commonly used probability distributions do not fit these data well.

The LDF can be generalized in several ways. For instance, one can take gam-

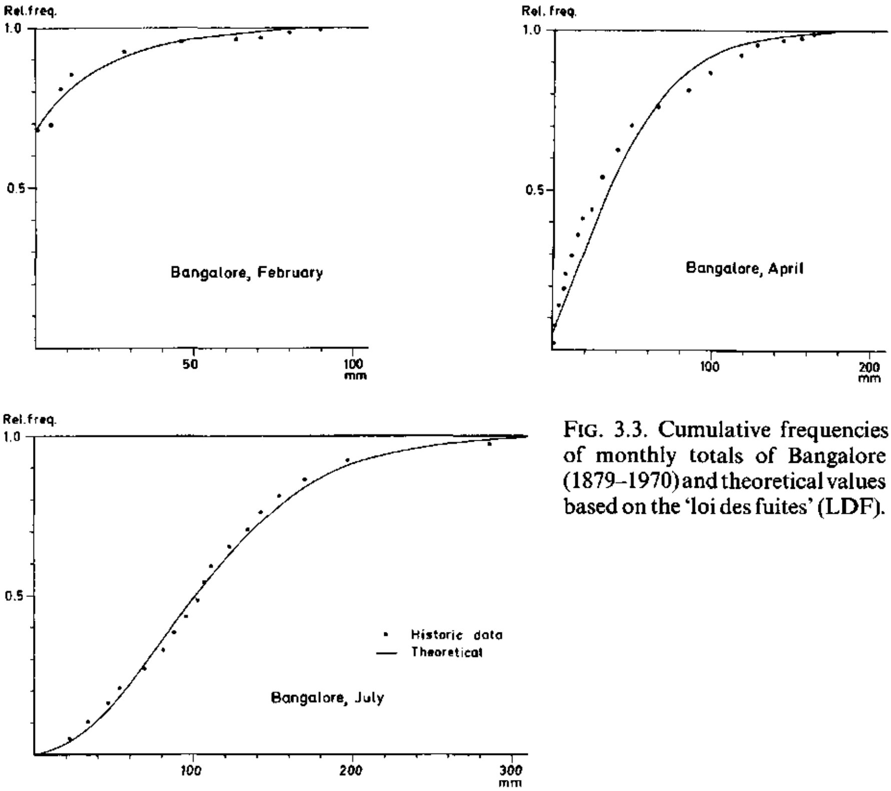


FIG. 3.3. Cumulative frequencies of monthly totals of Bangalore (1879–1970) and theoretical values based on the 'loi des fuites' (LDF).

ma distributed showers instead of exponential ones. Then the moment generating function of $\underline{x}_t$ is

$$(3.21)\qquad f(s) = E(e^{-s\underline{x}_t}) = e^{-\theta}\exp\left\{\theta\;\left(\frac{\rho}{\rho+s}\right)^{\nu}\right\}$$

where ν and ρ are the shape and scale parameter, respectively, of the gamma distribution for the showers. Taking logarithms in (3.21) and expanding log $f(s)$ in powers of s/ρ gives for the mth cumulant of $\underline{x}_t$:

$$(3.22)\qquad \varkappa_m = (-1)^m \binom{-\nu}{m}\frac{m!\theta}{\rho^m} = \frac{\theta}{\rho^m}\prod_{i=1}^{m}(\nu+i-1).$$

From (3.22) it can be deduced that:

$$(3.23)\qquad \gamma/C = 1 + \frac{1}{\nu+1}.$$

Since $\nu > 0$ the ratio γ/C can take values in the range [1,2).

When $\lambda, \nu \rightarrow \infty$ so that $\nu/\lambda \rightarrow \mu$ one gets the distribution which was proposed by DE BOER (1956, 1957, 1958) for rainfall totals over a period of at least 30 days. For this distribution the ratio γ/C equals 1.

4. Non-homogeneities due to a change in height of the rain gauge

In the beginning of this century the rim of rain gauges of the Royal Netherlands Meteorological Institute (KNMI) was at 1.50 m above the ground; during the period 1946–1954 rain gauges were lowered to 0.40 m above the ground. This was done after research of Braak (1945) who compared rain gauges with different heights at various sites in the Netherlands. Some results of his research for rain gauges with their rim at 1.50 or 0.40 m are summarized in Table 4.1. This table shows marked differences between rainfall amounts from rain gauges at different heights. These differences are caused by the rain gauge influencing the air movement so that a part of the rainfall, which should be recorded, is blown over the gauge. The largest differences occur at Dirksland, which is an unsheltered coastal station. Differences are smaller at coastal stations, which are more or less sheltered (Leiduin and Castricum) and at stations remote from the coast (De Bilt).

A drawback of Braak's research is that the rain gauges were only compared for a few years. Therefore, in this section the influence of the lowering of rain gauges is studied over a longer period of observation. Because the height of all rain gauges has been lowered in the Netherlands, a comparison between changed and unchanged rain gauges can only be based on rainfall data of neighbouring countries with no changes of height in the same period. For rainfall records of neighbouring countries one has the following possibilities:

a. Rainfall observations of the German Meteorological Institute. The observations are very suitable for this research, because near the Dutch border no change in height or type of rain gauge has occurred since 1883.

Table 4.1. Comparison of rain gauges at two different heights for various sites, after Braak (1945). The height of the rim of rain gauge R1 is 1.50 m; for rain gauge R2 this height is 0.40 m for the sites Castricum, Leiduin and De Bilt, and 0.35 m for the site Dirksland. (The rainfall amount of rain gauge R2 of Dirksland is assumed to be 37.2 mm in May 1940.)

		Dirksland	Castricum	Leiduin	De Bilt
Number of months		41	42	34	23
Monthly mean (mm)	R1	52.5	64.9	60.6	57.8
	R2	57.8	65.8	63.2	58.8
Monthly standard deviation (mm)	R1	36.1	45.5	42.8	43.5
	R2	39.0	46.3	44.6	43.9
Correlation coefficient (see (A4.1)) of R1 and R2 data		0.9984	0.9998	0.9992	0.9999
Number of times that the monthly total of	R1 > R2	0	6	2	1
	R1 = R2	0	1	0	1
	R1 < R2	41	35	32	21

b. Rainfall observations of the Belgian Royal Meteorological Institute. A drawback of these observations is that the type of rain gauge was changed about 1950. Besides, it is only since 1951 that measurement of rainfall in Belgium, has been well organized.

c. Rainfall observations at the Observatory of Ghent University. In the period 1921–1972 no change in height took place.

The detection of jumps in the mean, using annual totals, is discussed in Section 4.1. In Section 4.2 a seasonal change of jumps is investigated with monthly totals.

4.1 *Detection of jumps using annual totals*

The probability of success in detecting jumps strongly depends on the quality of the rainfall data at different sites. It is possible to get an idea about the quality of the data by comparing cross correlation coefficients of annual totals (see Section 4.1.1). The estimating and testing of jumps with a multivariate regression model for point rainfall data is discussed in Section 4.1.2. An analysis with partial sums of differences of averages of point rainfall data is described in Section 4.1.3. Section 4.1.4 deals with regression models using averages of point rainfall data and finally, in Section 4.1.5, the results of this research are compared with Braak's results.

4.1.1. Cross correlation coefficients of annual totals

Let ρ_{xy} be the correlation coefficient of two stations X and Y and r_{xy} the sample correlation coefficient (to be defined in Appendix A4). If there are N simultaneous observations at the two stations, it can be shown that for the sample correlation coefficient:

$$E(\underline{r}_{xy}) \approx \rho_{xy} \tag{4.1a}$$

and

$$\text{var}\,(\underline{r}_{xy}) \approx \frac{(1-\rho_{xy}^2)^2}{N}\,. \tag{4.1b}$$

The expressions only hold for homogeneous series. For the validity of (4.1b) also normality and absence of serial correlation have to be assumed (cf. KENDALL and STUART (1969), 10.9). These assumptions seem reasonable for annual totals on the basis of the results in previous sections.

In the case of non-homogeneous rainfall series the sample correlation coefficient can be strongly biased. The bias of the sample correlation coefficient is investigated in Appendix A4 for one jump in the mean in one of the two series. The numerical examples given in this appendix show that only very large jumps can lead to a serious underestimation of the theoretical correlation coefficient.

For 23 stations, correlation coefficients of annual totals were estimated for the period 1894–1970. Figure 4.1 shows the relation between the estimated correlation coefficients and the distances between the stations. Distances were

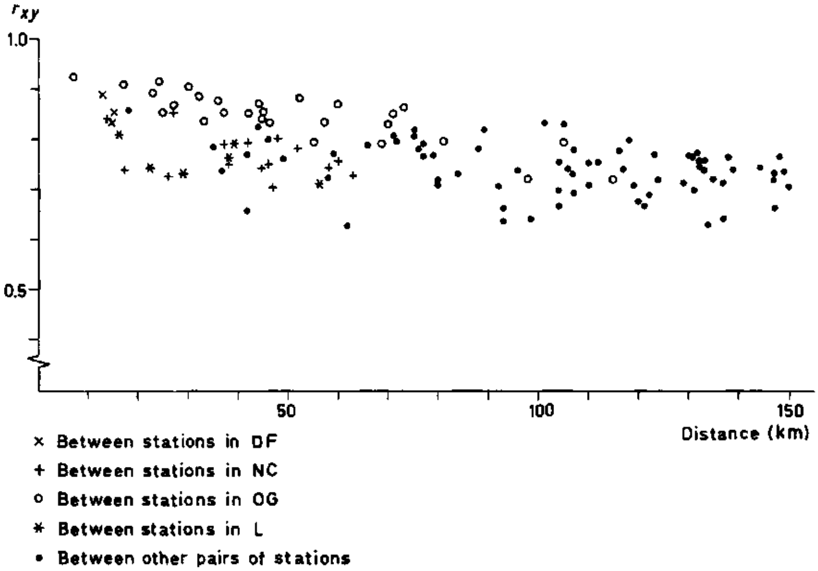


FIG. 4.1. Estimated correlation coefficients (r_{xy}) of annual totals for the period 1891–1970. The stations considered are:

- Biervliet, Groede, St. Kruis (Dutch-Flanders (DF)),
- Norderney, Norden, Leer, Warffum, Delfzijl, Finsterwolde (Northern Coastal area (NC)),
- Lingen, Rheine, Ahaus, Kleve, Ringenberg, Enschede, Hengelo, Winterswijk (Overijssel, Gelderland and adjacent German area (OG)),
- Jülich, Herzogenrath, Düren, Roermond (Limburg and adjacent German area (L)),
- Lathen, Ter Apel.

Correlation coefficients are only given for distances less than 150 km.

obtained from a list of coordinates. It is somewhat surprising that for small distances there is a considerable variation in the values of r_{xy}, which is much larger than could be expected from (4.1b). This large variation can be due to non-homogeneity or anisotropy of the considered area. Large differences can also be caused by changes in the rain gauge installation which usually give rise to a (negative) bias in correlation estimates (see Appendix A4). Not only the reduction in height of rain gauges in the Netherlands is important, but also the frequent changes of site. Changes of site can cause serious departures from homogeneity in the coastal area (local differences of the wind effect) and the southern part of the Netherlands (orographic effect). It is seen from Figure 4.1 that low values for r_{xy} are mainly found for stations in the northern coastal area and for Limburg and adjacent German area.

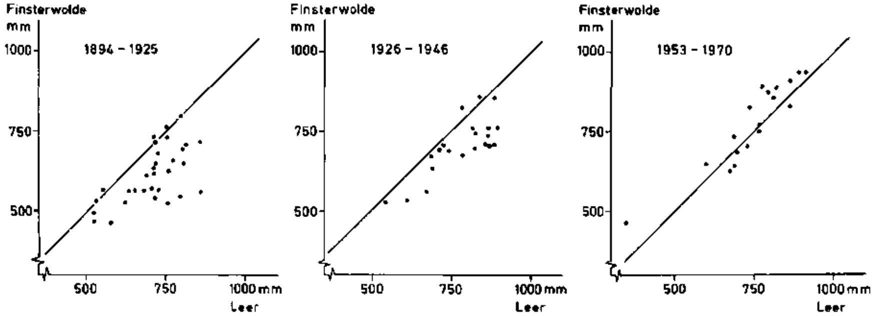


FIG. 4.2. Annual totals of Finsterwolde and Leer for three different periods.

Figure 4.2 compares annual totals of Finsterwolde and Leer for the periods 1894–1925, 1926–1946 and 1953–1970. The estimated correlation coefficients for these periods are 0.704, 0.794 and 0.913, respectively. A test on equality of correlation coefficients can be based on Fisher's z-transform:

$$z_{xy} = \tfrac{1}{2}\log\left(\frac{1+r_{xy}}{1-r_{xy}}\right) \tag{4.2}$$

which is the inverse of the hyperbolic tangent of the sample correlation coefficient.

Under the assumption of normality the mean and variance of $\underline{z}_{xy}$ are approximately (cf. KENDALL and STUART (1969), 16.33):

$$E(\underline{z}_{xy}) \approx \tfrac{1}{2}\log\left(\frac{1+\rho_{xy}}{1-\rho_{xy}}\right) + \frac{\rho_{xy}}{2(N-1)} \tag{4.3a}$$

$$\operatorname{var}(\underline{z}_{xy}) \approx \frac{1}{N-1} + \frac{4-\rho_{xy}^2}{2(N-1)^2}\,. \tag{4.3b}$$

The z-transform has an advantage because its distribution tends much faster to normality than the distribution of the sample correlation coefficient. Under the assumption of equal correlation for two different periods, the difference of the z-transforms is approximately normally distributed with mean zero (cf. KENDALL and STUART (1973), 26.19). The standard deviation of this difference follows from (4.3b) and is about 0.35 when both series have a length of 20 years.

The z-transforms for the three periods in Figure 4.2 are 0.704, 0.794 and 1.545, respectively; thus, there is some evidence for a better correlation in the most recent period.

The poor correspondence between simultaneous annual totals during the period 1894–1925 is partly because from 1911 up to 1924 the rain gauge of Finsterwolde was surrounded by huge elm-trees.

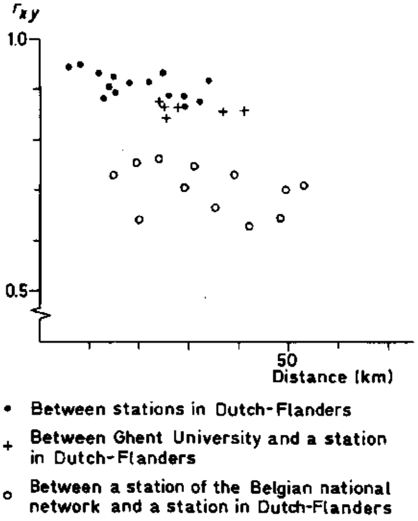


FIG. 4.3. Estimated correlation coefficients (r_{xy}) of annual totals of rainfall stations in Belgium and Dutch-Flanders for the period 1931–1972.

Figure 4.3 shows correlation coefficients of annual totals of rainfall stations in Belgium and Dutch-Flanders for the period 1931–1972. This figure reveals the following facts:

a. There is a strong correlation between annual totals from different stations in Dutch-Flanders. The estimated correlation coefficients are in general larger than correlation coefficients between rainfall stations in the northern coastal area, which is also seen from Figure 4.1.

b. There is a reasonable correlation between annual totals of Ghent University and those of stations in Dutch-Flanders.

c. There is a poor correspondence between rainfall totals of stations of the Belgian and Dutch national networks.

For annual totals of rainfall stations in this area Table 4.2 gives estimated correlation coefficients and their z-transforms for two different periods. The

TABLE 4.2. Correlation coefficients (r_{xy}) and their z-transforms (z_{xy}) of annual totals of some Belgian and Dutch stations for two different periods.

Station X	Station Y	r_{xy}		z_{xy}	
		1931–1946	1952–1972	1931–1946	1952–1972
St. Andries–Brugge	St. Kruis	0.636	0.905	0.752	1.501
Moerbeke–Waas	St. Kruis	0.378	0.851	0.406	1.259
Cadzand	St. Kruis	0.836	0.940	1.208	1.740
Ghent	St. Kruis	0.739	0.867	0.948	1.321
Ghent	Axel	0.814	0.891	1.139	1.427
St. Kruis	Axel	0.723	0.939	0.914	1.730

poor correspondence between rainfall data of the Belgian and Dutch national networks is due to a poor correspondence in the period before 1950. For all pairs of stations given in Table 4.2 correlation coefficients of the first period are smaller than those of the second period.

Because of the poor correspondence with rain gauges of the Dutch national network the rain gauges of the Belgian Meteorological Institute will not be considered.

4.1.2. A multivariate regression model

The effect of a reduction in height of rain gauges depends on the wind exposure and because the degree of sheltering against the wind differs from station to station possible jumps in the mean need not be the same for all stations. Therefore, in the first instance use is made of a multivariate regression model for estimating and testing jumps at various Dutch stations. A model could be:

$$(4.4)\qquad \begin{cases} \underline{y}_{kj} = \sum\limits_{i=1}^{p} \beta_{ij}\, x_{ki} + \beta_{0j} + \underline{e}_{kj} & k = 1,\ldots,n;\ j = 1,\ldots,q \\ \underline{y}_{kj} = \sum\limits_{i=1}^{p} \beta_{ij}\, x_{ki} + \beta_{0j} + \delta_j + \underline{e}_{kj} & k = n+1,\ldots,N; j=1,..,q \end{cases}$$

with p : number of foreign (Belgian, German) stations,
q : number of Dutch stations,
x_{ki}: rainfall amount in the kth year of the ith foreign station,
y_{kj}: rainfall amount in the kth year of the jth Dutch station.

During the first n years (period 1) the height of Dutch rain gauges is 1.50 m; during the last N-n years (period 2) this height is 0.40 m. The error terms $\underline{e}_{1j},\ldots,\underline{e}_{Nj}$ are assumed to be iid for every j. The marginal distribution is assumed to be Gaussian with mean zero.

The use of this model underlies the assumption that only a change in height can cause a jump in the mean. The facts that there are changes of site and that in 1962 a new type of rain gauge was introduced in the Netherlands are not considered. Therefore care is needed in the interpretation of the results of the regression analysis.

Estimates of the regression coefficients can be obtained by applying the method of least squares for each Dutch station separately (cf. RAO (1973), 8c.1 and 8c.2).

Figure 4.4 shows estimates of the jump δ_j, which were obtained by applying the regression model for four different regions. The periods for which the regression model was applied are given in Table 4.3. The largest values for the $\hat{\delta}_j$s belong to coastal stations. Jumps of more than 10 per cent are found in the northern coastal area (Warffum, Schiermonnikoog, Roodeschool), but they are in general much smaller for stations in the south-western coastal area. Possible explanations for this phenomenon are given in Section 4.1.3. The

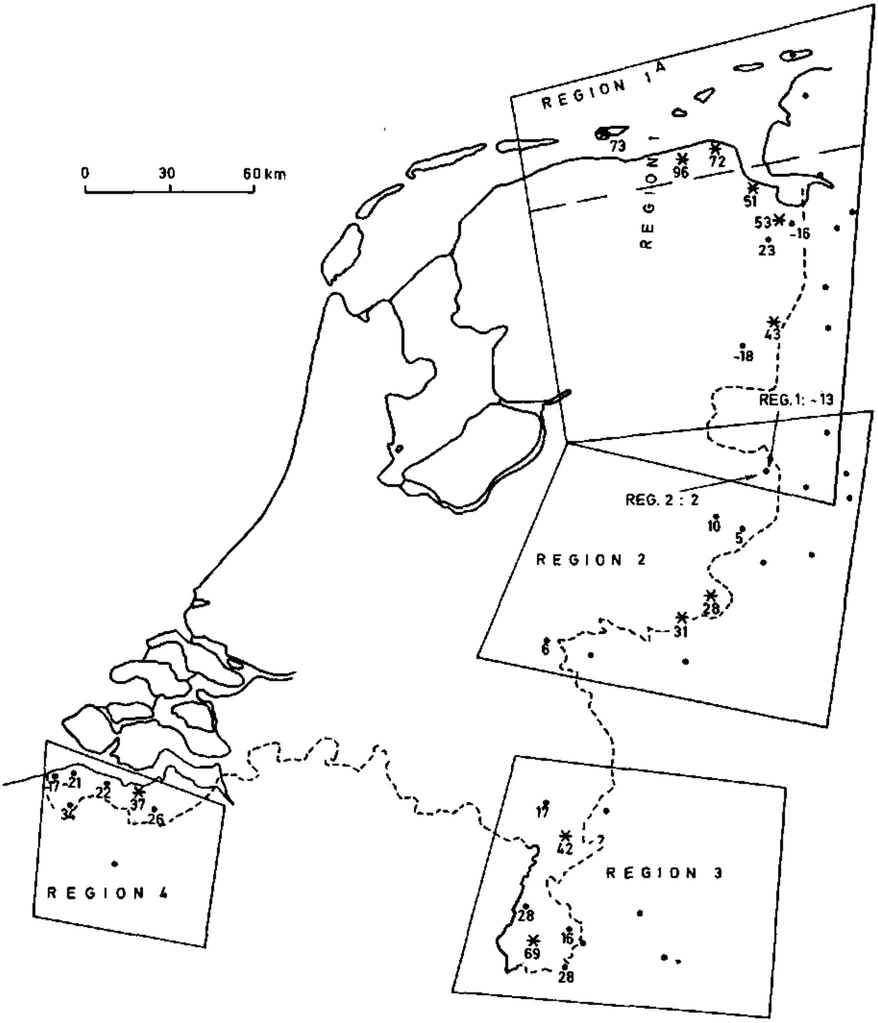


FIG. 4.4. Estimates of jumps δ_j (see Equation (4.4)) for different regions (in mm). Stations for which the jump differs significantly from zero are denoted by an asterisk.

TABLE 4.3. Realizations of the U-statistic (Equation (4.5)) for testing significance of jumps δ_j in the regression model (4.4). The different regions are given in Figure 4.4.

Region	Period 1	Period 2	U	Critical level
1	1926–1946	1953–1970	0.27	0.002
2	1926–1945	1953–1970	0.79	0.405
3	1926–1945	1955–1970	0.69	0.094
4	1931–1946	1952–1972	0.63	0.029

height of estimated jumps rapidly decreases, as the distance to the coast increases. There are, however, some stations for which the height of the estimated jump strongly deviates from jumps of adjacent stations (Nieuw Beerta, Ter Apel, Valkenburg, Groede). This deviation can be due to changes in the instrument or changes of site.

Whether or not the jump of a particular Dutch rainfall station differs from zero can be tested with a Student test. The test is one-sided, because it may be assumed that reduction of height of rain gauges leads to an increase in the recorded rainfall amounts. Stations for which a significant value is found at the 5 per cent level are denoted by an asterisk in Figure 4.4.

For a particular region one can also look at all stations simultaneously and test the hypotheses:

- H_0: all δ_js are equal to zero, and
- H_1: not all δ_js are equal to zero.

The test statistic, which has to be used in this case is (cf. RAO (1973), 8c.4)

$$(4.5) \qquad U = \frac{|\mathbf{C}_1|}{|\mathbf{C}_0|}$$

with $|\mathbf{C}_0|$: determinant of the sample covariance matrix under H_0,
$|\mathbf{C}_1|$: determinant of the sample covariance matrix under H_1.

Under H_0 the test statistic is close to its maximum value 1; values much smaller than 1 lead to rejection of H_0. For the null distribution of the statistic $\underline{U}$ holds (cf. RAO (1973), Table 8c.5β):

$$(4.6) \qquad \frac{1-\underline{U}}{\underline{U}} \simeq \frac{q}{N-p-q-1} \; \underline{F}(q, N-p-q-1)$$

where $\underline{F}$ stands for Snedecor's F-variable.

Realizations of the U-statistic and their critical levels are given in Table 4.3. At the 5 per cent level H_0 is rejected for the coastal regions 1 and 4. For region 4 this is somewhat surprising, because most jumps are small in this region and due to some negative values the average jump does not differ very much from the average jump of region 2. However, not only the height of jumps is important for the power of the U-statistic, but also the structure of the covariance matrix.

4.1.3. Analysis with partial sums

For the detection and quantification of jumps cumulative sum techniques can also be used. This section deals with partial sums of differences of annual averages of Dutch and foreign stations.

The ith partial sum S_i of a sequence of numbers $\{a_k\}_{k=1}^{N}$ is defined as:

$$(4.7) \qquad \begin{cases} S_0 = 0 \\ S_i = \sum_{k=1}^{i} a_k \qquad\qquad i = 1, \ldots, N. \end{cases}$$

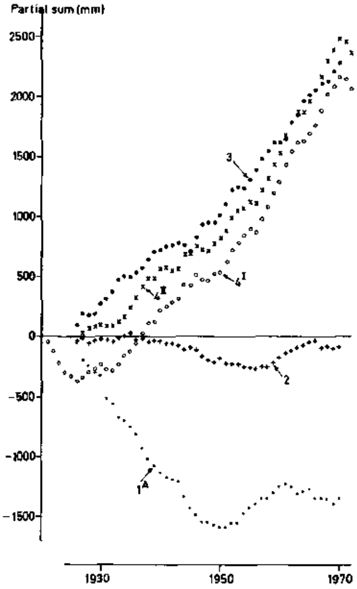


FIG. 4.5. Partial sums of differences of annual averages from two different countries. For region 4 the curve 4^{I} is based on the stations Ghent, Axel and Biervliet, whereas the curve 4^{II} is based on the stations Ghent, Axel, Biervliet, St. Kruis and Terneuzen.

When dealing with differences of annual averages from two different countries, a_k is assumed to be:

$$a_k = \bar{y}_k - \bar{x}_k \tag{4.8}$$

with $\bar{x}_k$: average of annual totals of some foreign rainfall stations for the kth year,

$\bar{y}_k$: average of annual totals of some Dutch rainfall stations for the kth year.

The index i is chosen such that $i = 1$ corresponds to the first year of period 1 and $i = N$ corresponds to the last year of period 2.

Figure 4.5 shows the relation between i and S_i for four different regions, which are denoted in Figure 4.4. The direction of the curves is not the same. The curve goes upwards when the mean of Dutch stations is larger than the mean of foreign stations; a downward curve occurs when the opposite is true. During the period in which Dutch rain gauges were lowered there is a visible change in slope of the curves. This change is most evident for the stations in region 1^{A} and is less obvious for stations in regions 2 and 3, because of the small wind effect in these regions. A remarkable fact is the large difference between stations in region 1^{A} and those in region 4. Possible explanations for this phenomenon are:

a. The average wind velocity is somewhat smaller for region 4.
b. Stations in region 4 may be more sheltered against the wind. The importance of the degree of protection was shown in Table 4.1 and was also demonstrated by BRAZIER (1927).

c. Other departures from homogeneity can be important. For regions 1^A and 4 the quality of the foreign stations is very important, because their number is small. The curves of region 4 show a change in slope during the early thirties, which is an indication of other non-homogeneities. Doubtful is also the fact that after 1960 the curve of region 1^A is nearly flat.

The estimated correlation coefficient between averages of annual totals of Dutch and foreign stations is 0.870 for region 1^A, 0.977 for region 2, 0.939 for region 3, 0.867 for region 4, using the stations Ghent, Axel and Biervliet only, and 0.889 for region 4, when the stations Ghent, Axel, Biervliet, St. Kruis and Terneuzen are used. If one compares these correlation coefficients with correlation coefficients of annual totals of point rainfall data, which are given in Figures 4.1 and 4.3, it turns out that they are larger for regions 1^A, 2 and 3. Some caution is needed for this conclusion, because the correlation coefficients do not refer to the same period.

Two explanations can be given for this phenomenon:

a. A jump in the mean due to a change of site causes a negative bias in the correlation estimator, but this jump is much smaller when averaging over other stations and consequently, it follows from the considerations in Appendix A4 that the negative bias in the correlation estimator is small.

b. Assume that in a particular region there are three rainfall stations in each country, denoted as X_1, X_2, X_3 and Y_1, Y_2, Y_3, respectively. Further it will be assumed that the stations in each country lie in an equilateral triangle. When the triangles of the two different countries are congruent and the distance between stations of the same country is small in comparison with the distance between stations of different countries, a reasonable variance-covariance structure is:

(4.9a) $\quad \text{var}\,\underline{x}_i = \text{var}\,\underline{y}_i = \sigma^2 \qquad$ for all i

(4.9b) $\quad \text{cov}(\underline{x}_i, \underline{y}_j) = \rho_b\sigma^2 \qquad$ for all i and j

(4.9c) $\quad \text{cov}(\underline{x}_i, \underline{x}_j) = \text{cov}(\underline{y}_i, \underline{y}_j) = \rho_w\sigma^2 \qquad$ for all i and j with $i \neq j$

where $\underline{x}_i$ and $\underline{y}_i$ are the annual totals at the sites X_i and Y_i, respectively.

Usually ρ_b will be smaller than ρ_w, because the correlation coefficient is in general decreasing with the distance. From (4.9) it can be concluded that the variances and covariances of the averages are:

$$(4.10a) \quad \text{var}\,\bar{\underline{x}} = \text{var}\,\bar{\underline{y}} = \frac{1}{3}(1+2\rho_w)\sigma^2$$

$$(4.10b) \quad \text{cov}(\bar{\underline{x}}, \bar{\underline{y}}) = \rho_b\sigma^2$$

and so for the correlation coefficient of the averages

$$(4.11) \quad \rho_{\bar{x}\bar{y}} = \frac{3\rho_b}{1+2\rho_w}$$

which is larger than ρ_b for $\rho_w < 1$.

4.1.4. Regression models based on averages of point rainfall observations

In Section 4.1.2 jumps in the mean were estimated for rainfall stations in different regions of the Netherlands and use was made of the U-statistic to test the significance of jumps. For regions 1^A, 2 and 3 it was shown that correlation coefficients of averages were larger than correlation coefficients of individual stations and therefore, a more powerful test for significance of jumps for a

TABLE 4.4. Features of annual totals (in mm) for stations in regions 1^A, 2 and 3. The means of periods 1 and 2 are denoted by m_1 and m_2. For region 1^A period 1 corresponds to 1926–1948 and period 2 to 1953–1970. For the other regions periods 1 and 2 are given in Table 4.3. The difference between m_2 and m_1 is denoted by v and its estimated standard deviation by s_v. The last column gives the realization of Student's t-statistic ($= v/s_v$) for testing equality of means.

		m_1	m_2	v	s_v	t
Stations in region 1^A						
Germany	Norderney	716	743	27	38	0.70
	Norden	789	812	24	36	0.66
Netherlands	Warffum	695	806	111	34	3.24
	Schiermonnikoog	710	822	112	36	3.12
	Roodeschool	650	738	88	31	2.83
Stations in region 2						
Germany	Venhaus	757	777	20	44	0.47
	Rheine	749	773	24	47	0.51
	Schüttorf	782	823	41	49	0.84
	Schöppingen	807	806	0	53	-0.01
	Ahaus	810	835	25	46	0.54
	Ringenberg	733	772	39	48	0.81
	Kleve	802	792	-10	50	-0.19
Netherlands	Denekamp	770	799	28	46	0.61
	Enschede	785	814	29	46	0.63
	Hengelo	778	799	21	46	0.45
	Winterswijk	770	817	46	49	0.95
	Aalten	749	805	56	47	1.20
	Nijmegen	781	795	14	40	0.35
Stations in region 3						
Germany	Bracht	744	827	83	46	1.80
	Jülich	631	713	82	40	2.05
	Herzogenrath	741	856	115	53	2.16
	Düren	613	632	19	37	0.52
Netherlands	Schaesberg	734	804	69	38	1.82
	Roermond	638	745	106	41	2.60
	Roggel	679	757	78	45	1.74
	Stein	699	806	107	44	2.44
	Vaals	832	958	126	48	2.62
	Valkenburg	722	863	140	43	3.28

particular region could be obtained by taking the annual averages of the rainfall stations of each country.

Table 4.4. gives the annual averages of periods 1 and 2 for stations in regions 1^A, 2 and 3. This table also gives the differences between the means, their estimated standard deviations and realizations of Student's t-statistic for testing equality of means. At the 5 per cent level differences between the means of period 1 and 2 are significant for all Dutch stations in regions 1^A and 3 (the one-sided critical value of $\underline{t}$ is about 1.69). In region 3, however, there are also significant differences between the means of periods 1 and 2 for some German stations. Besides, this region is characterized by large differences between nearby stations (Düren, Herzogenrath).

The first regression model to be considered is:

$$(4.12)\qquad \begin{cases} \underline{y}_k = \alpha + \beta x_k + \underline{e}_k & k = 1, \ldots, n \\ \underline{y}_k = \alpha + \beta x_k + \delta + \underline{e}_k & k = n+1, \ldots, N \end{cases}$$

with x_k: mean rainfall amount for German (or Dutch) stations of a particular region in the kth year,

y_k: mean rainfall amount for Dutch (or German) stations of a particular region in the kth year.

The error terms $\underline{e}_k$ are assumed to be independent normal variates with mean zero and the same variance.

Regression coefficients were estimated by ordinary least squares (OLS). Table 4.5 gives estimates of the jumps ($\hat{\delta}$), their estimated standard deviations ($s_{\hat{\delta}}$) and t-values ($\hat{\delta}/s_{\hat{\delta}}$) with one-sided critical levels (C.L.). The square of the multiple correlation coefficient (R^2) is also given in this table. The results in Table 4.5 show that it does not matter very much whether one predicts German data from Dutch data or Dutch data from German data. For regions 1^A and 3 application of a Student test leads to rejection of the null hypothesis ($\delta = 0$) at the 5 per cent level. However, the results for region 3 strongly depend on the stations, used in the regression analysis. Here a large number of stations is included to get a large multiple correlation coefficient. Though the alternatives for the U-test and t-test are not the same, it is remarkable that there are large differences in critical levels for regions 2 and 3 (see Tables 4.3 and 4.5).

Up to now it has been assumed that the reduction of height of Dutch rain gauges leads to a jump in the mean that can be corrected by adding a constant rainfall amount to annual totals of period 1. One can also think of models with the property that annual totals of period 1 have to be multiplied by some factor.

The first multiplicative model (model 1) is:

$$(4.13)\qquad \begin{cases} f\underline{y}_k = \beta x_k + \underline{e}_k & k = 1, \ldots, n \\ \underline{y}_k = \beta x_k + \underline{e}_k & k = n+1, \ldots, N \end{cases}$$

TABLE 4.5. Results for the regression model (4.12).

	Y: German stations X: Dutch stations			X: German stations Y: Dutch stations		
Region	1^A	2	3	1^A	2	3
$\hat{\delta}$ (mm)	–80.9	–13.2	–29.9	83.0	13.6	36.8
$s_{\hat{\delta}}$ (mm)	15.9	8.7	14.3	12.8	8.3	12.9
$t^{\hat{\delta}}$	– 5.09	– 1.52	– 2.09	6.48	1.63	2.85
C.L.	0.000	0.069	0.022	0.000	0.055	0.004
R^2	0.848	0.965	0.915	0.879	0.966	0.922

with x_k: mean rainfall amount for German stations of a particular region in the kth year,

y_k: mean rainfall amount for Dutch stations of a particular region in the kth year.

The error terms $\underline{e}_k$ are assumed to be independent normal variates with mean zero and variance σ^2.

The regression coefficients can be obtained by the method of maximum likelihood (ML), see Appendix A5.

Equation (4.13) can also be written as:

$$(4.14) \qquad \begin{cases} \underline{y}_k = \dfrac{\beta}{f} x_k + \dfrac{1}{f} \underline{e}_k & k = 1, \ldots, n \\ \underline{y}_k = \beta x_k + \underline{e}_k & k = n+1, \ldots, N \end{cases}$$

and since $f \approx 1$ this model is nearly equivalent to (model 2):

$$(4.15) \qquad \begin{cases} \underline{y}_k = \alpha_1 x_k + \underline{e}_k & k = 1, \ldots, n \\ \underline{y}_k = \alpha_2 x_k + \underline{e}_k & k = n+1, \ldots, N. \end{cases}$$

Estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ of α_1 and α_2, respectively, can be obtained by OLS, and a Student test can be done to test the equality of the regression coefficients α_1 and α_2, that is $f = 1$. An estimate $\hat{f}$ of the factor f follows from:

$$(4.16) \qquad \hat{f} = \hat{\alpha}_2 / \hat{\alpha}_1.$$

Linearization of (4.16) gives for the variance of $\underline{\hat{f}}$:

$$(4.17) \qquad \operatorname{var} \underline{\hat{f}} \approx \frac{\alpha_2^2 \operatorname{var} \underline{\hat{\alpha}}_1}{\alpha_1^4} + \frac{\operatorname{var} \underline{\hat{\alpha}}_2}{\alpha_1^2}$$

(cf. KENDALL and STUART (1969), 10.6). This variance can be estimated by:

$$(4.18) \qquad s_{\hat{f}}^2 = \hat{\alpha}_2^2 \, s_{\hat{\alpha}_1}^2 / \hat{\alpha}_1^4 + s_{\hat{\alpha}_2}^2 / \hat{\alpha}_1^2$$

where $s_{\hat{\alpha}_1}^2$ and $s_{\hat{\alpha}_2}^2$ are estimates of the variances of $\underline{\hat{\alpha}}_1$ and $\underline{\hat{\alpha}}_2$, respectively.

In models 1 and 2 the standard deviations of the error terms do not depend on the variate x. A model, which has the property that larger annual totals of German stations lead to error terms with larger standard deviations, is (model 3):

$$(4.19) \qquad \begin{cases} f \underline{y}_k = \beta x_k (1 + \underline{e}_k) & k = 1, \ldots, n \\ \underline{y}_k = \beta x_k (1 + \underline{e}_k) & k = n+1, \ldots, N. \end{cases}$$

Here the symbols x_k and y_k have the same meaning as in Equation (4.13). The error terms $\underline{e}_k$ are assumed to be iid with mean zero. Taking logarithms at both sides of (4.19) gives:

TABLE 4.6. Some results for multiplicative models.

		Region		
		1^{A}	2	3
Model 1	$\hat{f}$	1.115	1.017	1.027
	$s_{\hat{f}}$	0.019	0.011	0.018
Model 2	$\hat{f}$	1.115	1.017	1.028
	$s_{\hat{f}}$	0.020	0.011	0.019
	t	6.08	1.61	1.50
	C.L.	0.000	0.058	0.071
Model 3	$\hat{f}$	1.115	1.016	1.037
	$s_{\hat{f}}$	0.020	0.011	0.020
	t	6.00	1.36	1.92
	C.L.	0.000	0.091	0.032

$$(4.20)\qquad \begin{cases} \log(\underline{y}_k/x_k) \approx \log\beta - \log f + \underline{e}_k & k = 1,\ldots,n \\ \log(\underline{y}_k/x_k) \approx \log\beta \qquad\qquad + \underline{e}_k & k = n+1,\ldots,N \end{cases}$$

because:

$$(4.21)\qquad \log(1+\underline{e}_k) \approx \underline{e}_k.$$

Estimates of the regression coefficients can be obtained by OLS and a Student test can be done to test the hypothesis $\log f = 0$, that is $f = 1$. For the estimation of the standard deviation of $\underline{\hat{f}}$ use can be made of:

$$(4.22)\qquad \mathrm{var}\,\underline{\hat{f}} \approx f^2 \mathrm{var}(\log \underline{\hat{f}}).$$

Results for multiplicative models are given in Table 4.6. In this table t stands for the realization of Student's test statistic for a test on $f = 1$. The critical levels (C.L.) are based on a one-sided test. The values in Table 4.6 show that the various multiplicative models lead to the same result. For region 1^{A} there is an obvious indication for a larger mean in period 2, but for the other regions this is less obvious. There is also a good correspondence between the results of the additive model, given in Table 4.5 and those of multiplicative models, given in Table 4.6. From a mathematical point of view it is difficult to decide which kind of model has to be preferred because there are no zero or nearly zero annual totals, but from a hydrological point of view a multiplicative model is more plausible. Besides, the factor in multiplicative models can also be applied to periods shorter than a year. If an additive model is used, a method has to be found for splitting up the jump in the mean for shorter periods.

4.1.5. Comparison with earlier research

In the previous sections it was investigated whether a change of height of Dutch rain gauges causes a jump in the mean of annual totals. Dutch rainfall observations were compared with observations of neighbouring countries for a period of about 40 years. Because of this rather long period this research might be more useful than BRAAK's (1945) research. However, Braak compared rain gauges at the same site, whereas here rainfall stations of different sites are used and consequently Braak's rainfall observations show a better correlation. This is seen from the estimated correlation coefficients in Table 4.1 and Figures 4.1 and 4.3.

It is, however, not only the distance between the rainfall stations which leads to smaller correlation coefficients. Non-homogeneities also give rise to a negative bias in the estimated correlation coefficients. When there are non-homogeneities, for instance due to changes of site, the results of the analysis may be biased. This bias can be reduced by taking into consideration many stations simultaneously, since such non-homogeneities occur locally.

Though different methods were followed here to determine the height of the jump, the result is nearly always the same. The height of the jump usually ranges from 2 per cent (stations remote from the coast) to a bit more than 10 per cent (coastal stations). These results correspond quite well to those of Braak, but it should be noted that Braak discarded months with snowfall.

4.2. *Seasonal changes of jumps in the mean*

In this section seasonal changes in a non-homogeneity due to a reduction in height of rain gauges are investigated. Seasonal changes can result from (cf. BRAZIER (1927)):

a. Seasonal changes in wind velocity. In the Netherlands large wind velocities occur more frequently during winter.
b. Seasonal changes in the degree of sheltering. During summer, rain gauges are in general better protected against the wind.
c. Seasonal changes in drop size. Rain drops are in general larger during the summer season, so that the wind has less influence on the movement of the rain drops.

Monthly totals were used to investigate seasonal changes of a jump in the mean.

The first model to be considered is model 2 of the previous section (see Equation (4.15)):

$$(4.23)\qquad \begin{cases} \underline{y}_{km} = \alpha_{1m}\, x_{km} + \underline{e}_{km} & k = 1, \ldots, n \\ \underline{y}_{km} = \alpha_{2m}\, x_{km} + \underline{e}_{km} & k = n+1, \ldots, N \end{cases}$$

with x_{km}: mean of the square root of the rainfall amounts for German stations of a particular region in the mth month ($m = 1$ corresponds to January) of the kth year,

y_{km}: mean of the square root of the rainfall amounts for Dutch stations of a particular region in the mth month of the kth year.

The square root of monthly totals is taken here as a normalizing transfor-

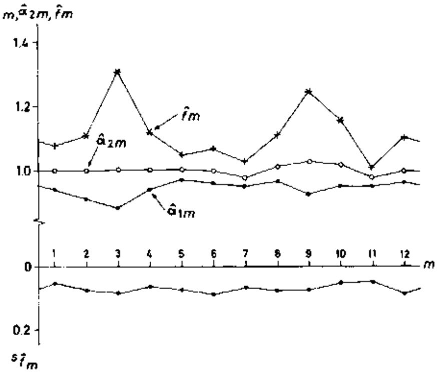


FIG. 4.6. Estimated regression coefficients and estimated multiplication factors with their estimated standard deviations for model 2 (see Equations (4.23)and (4.24)) and region 1^{A}. Factors which differ significantly from 1 are denoted by an asterisk.

mation (see 3.2). A multiplication factor for monthly totals in the *m*th month of period 1 can be obtained from:

$$(4.24) \qquad \hat{f}_m = \hat{\alpha}_{2m}^2/\hat{\alpha}_{1m}^2.$$

Since the coefficient of variation of monthly totals is larger than for annual totals (they differ by a factor of about $\sqrt{12}$) and since the cross correlation coefficients of monthly and annual totals are of the same order, the estimate of the factor f_m is less accurate than the estimate of f for annual totals (its standard deviation is about a factor $\sqrt{12}$ larger). Because of this large standard deviation only stations in region 1^{A} are considered. For stations in this region the estimates $\hat{\alpha}_{1m}$, $\hat{\alpha}_{2m}$, $\hat{f}_m$ and $s_{\hat{f}_m}$ (estimate of the standard deviation of $\hat{f}_m$) are given in Figure 4.6. Again use was made of a linearization for the determination of $s_{\hat{f}_m}$. Figure 4.6 also shows for which months a one-sided Student test leads to rejection of the hypothesis $\alpha_{1m} = \alpha_{2m}$ ($\alpha = 0.05$). The height of the factor ($\hat{f}_m$) changes irregularly from month to month, partly because of its rather large standard deviation. It is remarkable, however, that the annual variation of the factor $\hat{\alpha}_{1m}$ is much larger than the annual variation of $\hat{\alpha}_{2m}$. All factors turn out to be larger than 1, but the largest values occur during the winter season. Possible explanations for this phenomenon are given in the beginning of this section. Further it should be noticed that during the winter season a substantial part of the monthly totals can consist of snow. The amount of snow measured strongly depends on the type of rain gauge and this makes the regression analysis less accurate during winter, because Dutch and German rain gauges are not of the same type. Another problem is that the correlation between monthly averages of Dutch and German stations can differ considerably from month to month. This is seen from Figure 4.7 which shows averages of square roots of monthly totals of Dutch and German stations for March and April. The points are more scattered for period 1 of the month of March. Since there is an irregular change in the correlation between monthly totals the estimated standard deviation ($s_{\hat{f}_m}$) of the factor also changes irregularly (see Figure 4.6).

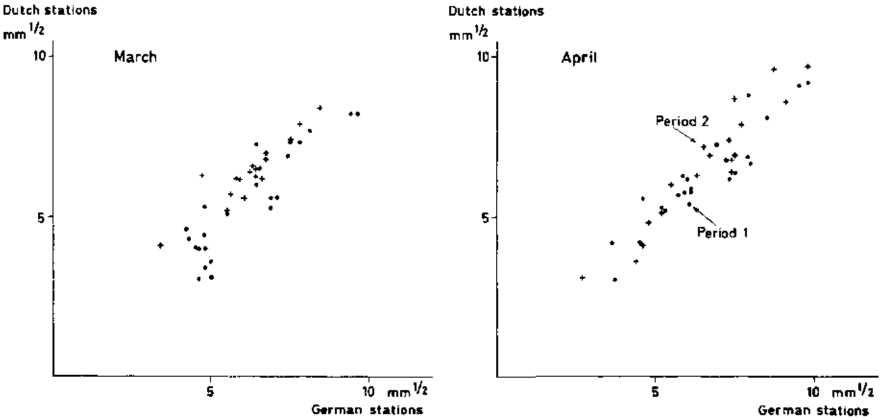


FIG. 4.7. Averages of square roots of monthly totals (in mm) of Dutch and German stations in region 1^A.

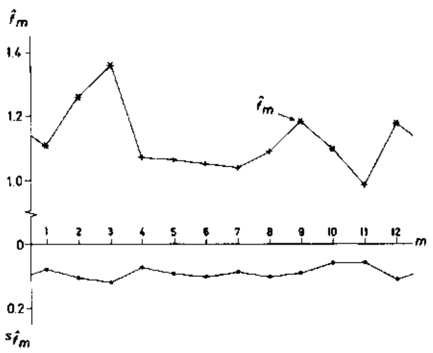


FIG. 4.8. Monthly estimates of the multiplication factor and their estimated standard deviations for model 3 and region 1^A. Factors which differ significantly from 1 are denoted by an asterisk.

For the estimation of the factor f_m use can also be made of model 3 of the previous section (see Equation (4.19)). Estimates of the factors and their estimated standard deviations ($s_{\hat{f}_m}$) are given in Figure 4.8. For the calculation of $s_{\hat{f}_m}$ Equation (4.22) was applied.

Figure 4.8 also shows for which months a one-sided Student test leads to rejection of the hypothesis $\log f_m = 0$ ($\alpha = 0.05$). Figures 4.6 and 4.8 only show small differences between the estimated factors of models 2 and 3. Only model 2 leads to a slightly larger value of $\hat{f}_m$ for the month of March and a slightly smaller value for the month of September. The estimated standard deviation, $s_{\hat{f}_m}$, is always somewhat smaller when model 2 is used. For this reason and also because there is no indication for an increase in the standard deviation with the height of the rainfall amount (see Figure 4.7), model 2 is preferable to model 3.

5. Homogeneity of the Zwanenburg–Hoofddorp series

The Zwanenburg–Hoofddorp series is one of the longest rainfall series in the world. In 1735 rainfall observations started in Zwanenburg and since February 1861 have been continued in Hoofddorp. Because of its length this rainfall series can be very important for water resources problems, but before using this series one must be sure of its homogeneity.

Several possible non-homogeneities of this series are discussed in Section 5.2 and 5.3. The analysis of homogeneity is preceded by a short review about the history of this rainfall series.

5.1. *Historical review of the rainfall observations at Zwanenburg and Hoofddorp*

5.1.1. Rainfall observations at Zwanenburg

Rainfall was observed three times a day. The original data from 1766 can be found in the archives of the 'Hoogheemraadschap Rijnland' at Leyden. The original data before 1766 were lost, but nearly all monthly totals of the period 1735–1860 are known. Supplements of gaps in the series were given by LABRIJN (1945).

Up till September 1787 the rain gauge had an orifice of 493 cm^2 and its rim was about 3 meter above the ground. From May 1788 the orifice of the rain gauge was 246 cm^2 and its height about 2.50 m. From that date there was also a frame of brass-wire in the funnel to prevent stoppages. Rainfall was recorded in units of 1 'lijn' (1 'lijn' is about 2.21 millimeters). Large errors can have occurred during frost and snowfall, because solid precipitation in the rain gauge was not melted. Also small rainfall amounts were not recorded.

To get a homogeneous Zwanenburg-Hoofddorp series, LABRIJN (1945) multiplied all monthly totals of Zwanenburg by a factor 1.11. The factor was based on a comparison of annual means of Zwanenburg and Den Helder for the period 1844–1858 with values of Hoofddorp, Lijnden and Den Helder for the period 1891–1930. The following remarks can be made on Labrijn's procedure:

a. The distance between Den Helder and Hoofddorp is quite large (nearly 75 km).

b. The type and the height of the rain gauge have not always been the same in Den Helder and in Hoofddorp.

c. The correction is the same for all months. Use of different factors for winter and summer is preferable, because during winter there is a large wind effect and there are problems with solid precipitation.

For rainfall observations before 1787 Labrijn justified the use of the factor 1.11 by comparing annual means of Zwanenburg-Hoofddorp with those of Haarlem for the periods 1735–1742 and 1891–1930. Indeed from the annual averages of Zwanenburg-Hoofddorp and Haarlem, given in Table 5.1, a factor 1.11 looks reasonable, but when comparing annual averages of Zwanenburg-Hoofddorp with those of Leyden (see Table 5.2) a larger factor would be more appropriate. Notice from Table 5.2, that the 1736–1758 mean of Leyden (ob-

TABLE 5.1. Annual averages (in mm) of Zwanenburg–Hoofddorp and Haarlem for different periods (after Labrijn (1945)).

Rainfall series	Average for the period	
	1735–1742	1891–1930
Zwanenburg–Hoofddorp	620	750
Haarlem	718	770

TABLE 5.2. Annual averages (in mm) of Zwanenburg-Hoofddorp and Leyden for different periods.

Rainfall series	Average for the period		
	1736–1758	1925–1946	1951–1970
Zwanenburg–Hoofddorp	658	746	825
Leyden	786	719	816

servations of Musschenbroek) does not differ very much from more recent means.

For tests on homogeneity of the Zwanenburg–Hoofddorp series, the Zwanenburg series given by Labrijn (1945) was used. Unless stated otherwise, rainfall observations before 1861 are divided by 1.11 to regain the original values.

5.1.2. Rainfall observations at Hoofddorp

Meteorological observations at Hoofddorp started in February 1861. The rain gauge had a square funnel with an area of 400 cm^2 (cf. LABRIJN (1945)) and its rim was at 1.55 m above the ground. From October 1907 onwards another type of rain gauge was used. This rain gauge, which was the standard gauge of the KNMI, had a round funnel with an area of 400 cm^2. The height of the rim was 1.50 m, but the most important difference with respect to the previous gauge was its shallow funnel. The rain gauge was lowered to a height of 0.40 m in March 1947. In 1964 the new standard gauge, with an orifice of 200 cm^2 and a better shape of funnel, was introduced. The site of the rainfall station was changed in October 1913, January 1961 and January 1973.

The original data of the Hoofddorp series can be found in the archives of the KNMI, except the data for the period 1867–1887. Daily, monthly and annual totals can also be found in publications of the KNMI.

5.2. *The usefulness of the Zwanenburg data*

In this and the next section, the following subseries are distinguished:

- Zwanenburg (1735–1860),
- Zwanenburg 1 (1735–1787),
- Zwanenburg 2 (1788–1860),
- Hoofddorp (1861–1972).

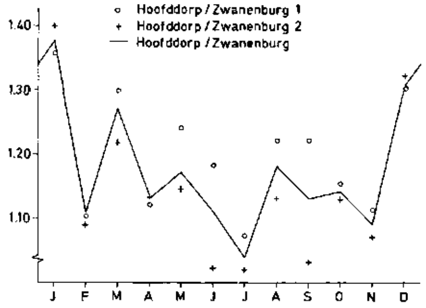


FIG. 5.1. Ratios of the monthly means of Hoofddorp and Zwanenburg.

TABLE 5.3. Critical levels of the Wilcoxon test for equality of means of Zwanenburg and Hoofddorp.

Month	Zwanenburg Hoofddorp	Zwanenburg 1 Hoofddorp	Zwanenburg 2 Hoofddorp
January	0.000	0.000	0.000
February	0.158	0.480	0.125
March	0.000	0.021	0.001
April	0.040	0.054	0.130
May	0.019	0.295	0.008
June	0.079	0.414	0.050
July	0.386	0.473	0.476
August	0.012	0.190	0.007
September	0.091	0.807	0.020
October	0.057	0.182	0.078
November	0.343	0.551	0.352
December	0.000	0.000	0.000

A subdivision of the Hoofddorp series is given in Section 5.3.

The Zwanenburg and Hoofddorp series are compared on the basis of monthly totals. There are no problems of a time-shift due to the transition from the Julian to the Gregorian calendar, because in this part of the Netherlands the Gregorian calendar was introduced in 1583. The ratios of the monthly means of the Hoofddorp and Zwanenburg series are given in Figure 5.1. A Wilcoxon test was done for testing differences in monthly means of the Hoofddorp and Zwanenburg series. The critical levels of this test are given in Table 5.3. The test is one-sided, because it may be assumed that the Zwanenburg mean could be smaller (larger height of the rain gauge, omission of small rainfall amounts and the measurement of snow). From Figure 5.1 and Table 5.3 it can be concluded:

a. Differences between monthly means of the Zwanenburg and Hoofddorp series can be quite large and are on the average a bit more than 10 per cent (this is about Labrijn's correction, but it should be noted that the Hoofddorp series is 25 years longer here). For months of the winter season the differences can be much larger, which is due to the wind effect and the measurement of snow.

b. For many months the Wilcoxon test leads to significant values at the 5 per cent level.

c. There are differences between Zwanenburg 1 and 2 during the summer season. For the period of April-September the monthly mean is 61.4 mm for Zwanenburg 1 and 55.6 mm for Zwanenburg 2. The differences may have been caused by interception losses due to the frame of brass-wire.

d. The magnitude of the differences changes irregularly from month to month. Especially the small differences for the month of February are remarkable. Before 1905 the February monthly totals in the archives of the KNMI refer to the period January 31st–March 1st, but Labrijn's series has been corrected for this.

To get an idea about the reliability of old rainfall series, correlation coefficients of annual totals of Zwanenburg–Hoofddorp and Leyden were calculated for the periods 1736–1758, 1925–1946 and 1951–1970. The estimated correlation coefficients for these periods are respectively 0.792, 0.911 and 0.918; annual totals for the first and third period are plotted in Figure 5.2. From this figure it is seen that the points of the first period are more scattered. Something similar holds for the series of Zwanenburg–Hoofddorp and Den Helder (see Figure 5.3) and therefore the Zwanenburg data are useless for the solution of present-day hydrological problems. May be the best place for such data is a museum.

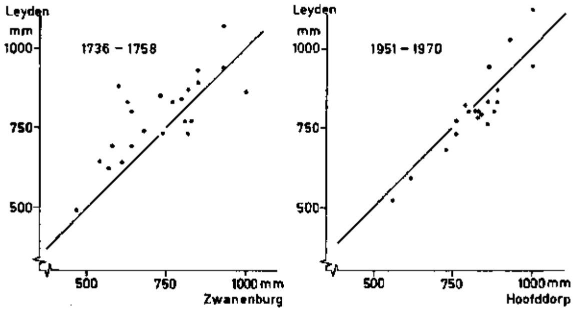


FIG. 5.2. Annual totals of Leyden and Zwanenburg-Hoofddorp for two different periods. The Zwanenburg-Hoofddorp data are those given by LABRIJN (1945).

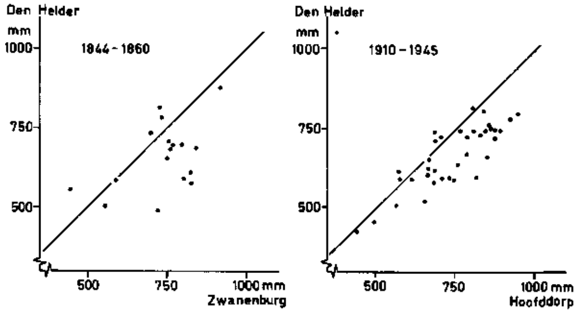


FIG. 5.3. Annual totals of Den Helder and Zwanenburg-Hoofddorp for two different periods. The Zwanenburg-Hoofddorp data are those given by LABRIJN (1945).

5.3. *Homogeneity of the Hoofddorp series*

In Section 5.1.2 various changes in rainfall measurements at Hoofddorp were mentioned. A first idea about departures from homogeneity due to these changes can be obtained by plotting the partial sums of the departures from the mean ($a_k = x_k - \bar{x}$ in (4.7), where $\bar{x}$ is the mean of the x_ks) on the basis of annual totals (see Figure 5.4). There are some changes in slope of the curve of S_i versus i, namely:

a. About 1880. No indications for this change can be found in the archives of the KNMI.

b. Somewhere between 1905 and 1910. This change may be attributed to the change in the type of instrument in 1907. The change in slope could also be ascribed to the change of site in 1913, but the positions of the rain gauges before and after 1913 (see archives of the KNMI) do not give any reason for a smaller mean precipitation after 1913.

c. About 1950. This change is ascribed to the reduction in height of the rain gauge.

As a consequence of these results the Hoofddorp series is split up into 3 subseries, namely Hoofddorp 1, 2 and 3, which refer to the periods 1861–1907, 1908–1946 and 1947–1972, respectively. Estimates of the annual means, μ_1, μ_2 and μ_3 of Hoofddorp 1, 2 and 3 are 776, 738 and 790 mm, respectively.

Homogeneity of the Hoofddorp series is tested under the assumption of equal variances and normality of the annual totals. The following tests were done:

a. $H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_1, \mu_2$ and μ_3 are not mutually equal. The realization of Snedecor's F-statistic is 1.81 which is not significant at the 5 per cent level (critical level is 0.168).

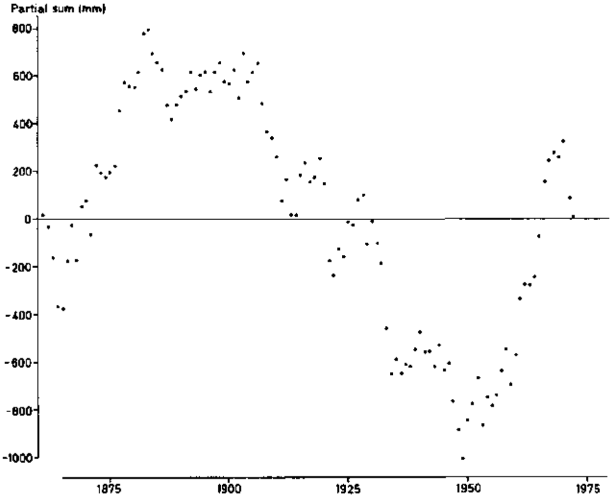


FIG. 5.4. Partial sums of departures from the mean of annual totals of Hoofddorp (1861–1972).

b. $H_0: \mu_1 = \mu_2$
$H_1: \mu_1 > \mu_2$
which gives a realization of 1.06 for Student's t-statistic (critical level is 0.147).

c. $H_0: \mu_2 = \mu_3$
$H_1: \mu_3 > u_2$
which gives a realization of 1.67 for Student's t-statistic (critical level is 0.050).

So, when no series of neighbouring stations are used, differences in the mean of about 6 per cent do not give evidence for non-homogeneity at the 5 per cent level (one-sided).

An interesting point of investigation is the change of site in 1961. During the period November 1958 – December 1961 observations were made at the old and new site, which are denoted as X and Y, respectively. For the 38 monthly totals there are 22 positive differences between Y and X, 15 negative differences and 1 tie. A sign test does not give evidence for differences in the mean at the 5 per cent level. The mean difference between monthly totals of Y and X is only 0.23 mm.

6. Homogeneity with respect to the number of wet days

The daily rainfall model which is described in Chapters III and IV contains a separate process for the occurrence of wet and dry days (shortly denoted as wet-dry process). A dry day is defined as a day on which rainfall does not exceed some threshold δ. In the model rainfall amounts on dry days are set to zero. An important problem is the choice of δ, since a large value of δ can result in a stochastic model which is a bad approximation of the real rainfall process. On the other hand, if a low value for δ is taken the wet-dry series can be non-homogeneous, due to the quality of different observers.

To demonstrate this, correlation coefficients of the annual number of wet days were estimated for 15 stations in a small area in the east of the Netherlands with thresholds of 0.3 and 0.8 mm. The value 0.3 mm is a kind of minimum value, because rainfall amounts smaller than this value can also be due to fog or dew. The relation between the estimated correlation coefficients and the distances between the stations is given in Figure 6.1. There is only a weak dependence between the height of the correlation coefficient and the distance, but the most important fact is the large scatter of the points for the lower threshold. Even though the number of years (19) is small, the points should be close together because the correlation coefficients are heavily correlated, as a consequence of the large correlation between the data. A threshold of 0.8 mm looks more acceptable, though there is a small reduction in the mean (about 3 per cent) when values smaller than 0.8 mm are set to zero.

In Section 3.1 the Von Neumann's ratio was introduced for testing serial correlation in homogeneous series. The numerator in (3.5) is hardly influenced by a jump in the mean, but the denominator usually tends to be much larger,

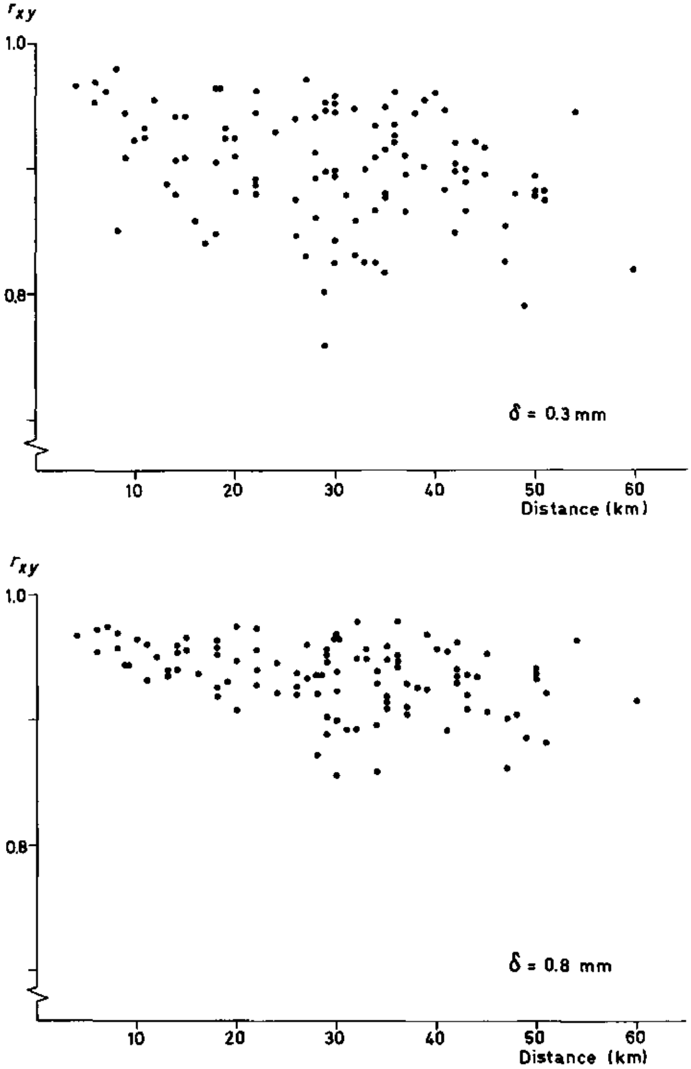


FIG. 6. 1. Correlation coefficients of the annual number of days with a rainfall amount of at least δ mm for the stations of Heino, Vroomshoop, Almelo, Enschede, Hengelo, Twente, Hellendoorn, Oldenzaal, Lettele, Lochem, Winterswijk, Borculo, Rekken, Deventer and Almen.

which is seen from (A4.5b) and (A4.8b). Therefore the Von Neumann's ratio tends to be smaller than 1 for a non-correlated rainfall series with a jump in the mean. Also, for more than one jump in the mean the denominator tends to be larger (cf. YEVJEVICH and JENG (1969), Equations (31) and (52)) and consequently the Von Neumann's ratio tends to be smaller than 1. The annual number of wet days of successive years can be considered as independent (otherwise there would also be an indication for serial correlation in monthly and annual rainfall amounts) and therefore the Von Neumann's ratio can be used as a test for homogeneity. This test is less powerful than the F-test and t-tests considered in Section 5.3, because it does not assume any knowledge about the position of possible jumps. Yet, it is possible to get a clear idea about the homogeneity of the wet-dry process as will be seen below.

For annual data the number of observations is quite small and therefore the null distribution of the Von Neumann's ratio could be sensitive to departures from normality. Therefore a Monte Carlo experiment was done to investigate this influence. Samples of size 70 were generated from a normal distribution and two special cases of the λ-distribution. The variable $\underline{x}$ has a λ-distribution if:

$$(6.1) \qquad \begin{cases} \underline{x} = (\underline{u}^{\lambda} - (1-\underline{u})^{\lambda})/\lambda & \lambda \neq 0 \\ \underline{x} = \log(\underline{u}/(1-\underline{u})) & \lambda = 0 \end{cases}$$

where $\underline{u}$ is standard uniform (uniform on (0,1)). For $\lambda = 0$ one has the logistic distribution.

The cases of the λ-distribution, which are considered here are those for which $\lambda = 0$ and $\lambda = -1$. For these values of λ the distributions have longer tails than the normal distribution; in the case $\lambda = -1$ the distribution is even so long-tailed that no moments exist (as in the Cauchy distribution). Normal probability plots of the empirical distributions of the Von Neumann's ratio, based on 1,000 series for each type of distribution, are given in Figure 6.2.

Pseudo-random standard uniform variates were obtained from the function RAN of the DEC 10 computer, which is based on an article by PAYNE et al. (1969). The series for different types of distributions are based on the same standard uniform variates.

The empirical distributions of the Von Neumann's ratio coincide for normal and logistic variates, so it can be concluded that small departures from normality are not important. When $\lambda = -1$ a test based on the Von Neumann's ratio is conservative at the 5 per cent level.

For the annual number of wet days of the rainfall series of Winterswijk, Hengelo and Hoofddorp realizations of the Von Neumann's ratio are given in Table 6.1, together with estimates of the mean, the standard deviation and the

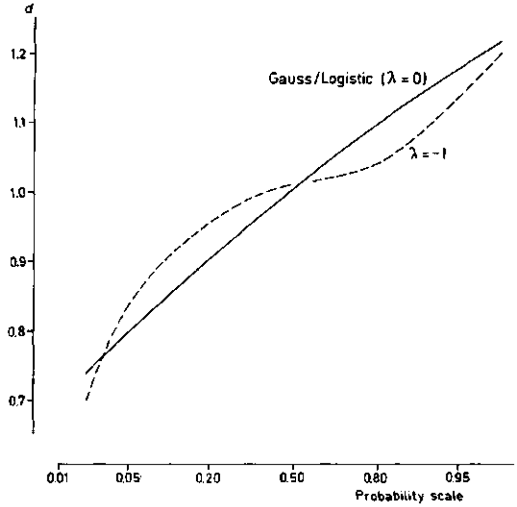


FIG. 6.2. Normal probability plots of the Von Neumann's ratio d for independent processes with a Gaussian or λ-distribution (see Equation (6.1)). The plots are based on 1,000 samples of size 70.

TABLE 6.1. Mean (m), standard deviation (s), coefficient of skewness ($\hat{\gamma}$) and Von Neumann's ratio (d) of the annual number of wet days.

			m	s	$\hat{\gamma}$	d
$\delta = 0.3$ mm	Winterswijk	1881–1973	169.5	20.2	–0.662	0.639
	Winterswijk	1908–1973	175.2	17.2	–0.853	0.954
	Hengelo	1908–1973	166.4	19.3	–0.310	0.821
	Hoofddorp	1867–1971	168.2	19.3	–0.212	0.800
$\delta = 0.8$ mm	Winterswijk	1881–1973	140.0	15.6	–0.363	0.930
	Winterswijk	1908–1973	142.3	15.6	–0.422	1.018
	Hengelo	1908–1973	139.9	17.3	–0.487	0.904
	Hoofddorp	1867–1971	138.9	16.9	–0.116	0.977

coefficient of skewness. These values are given for two different thresholds, namely $\delta = 0.3$ and 0.8 mm. For the lower threshold the Von Neumann's ratio shows evidence for non-homogeneity ($\alpha = 0.05$, two-sided).

For Winterswijk and Hoofddorp partial sums of departures from the mean are given in Figure 6.3, to get an idea about the positions of jumps in the mean. Years with changes of observer are also denoted in this figure, except for the Hoofddorp series before 1910. When $\delta = 0.3$ mm, there is a change in slope of the curve about 1910 for Winterswijk and about 1925 for Hoofddorp. These changes in slope can be ascribed to the change of observer in 1907 and 1922. For Winterswijk, STOL (1970) showed that the number of small rainfall amounts is only comparatively low during the winter season of the period before 1908.

It is possible to test homogeneity with these partial sums. If S_i denotes the ith partial sum of the departures from the mean and if N is the length of the series, the adjusted range is defined as:

$$R_N^a = \max_{0 \leqslant i \leqslant N} S_i - \min_{0 \leqslant i \leqslant N} S_i \tag{6.3a}$$

and the rescaled range as:

$$R_N = R_N^a / s \tag{6.3b}$$

where s denotes the sample standard deviation.

When there is a jump in the mean the adjusted range tends to be larger, usually in such a way that also the rescaled range tends to be larger. For homogeneous normal independent processes cumulative distribution functions of $\underline{R}_N$, based on Monte Carlo simulations, are given by WALLIS and O'CONNEL (1973) for $N = 20\ (10)\ 50\ (25)\ 100$.

To investigate the sensitivity to departures from normality the rescaled ranges were computed for the synthetic series on which Figure 6.2 was based. Normal probability plots of the empirical distributions of R_N are given in Figure 6.4. For the normal and logistic distribution the distributions coincide, while for $\lambda = -1$ tests based on R_N are conservative when the percentage points by Wallis and O'Connel are used.

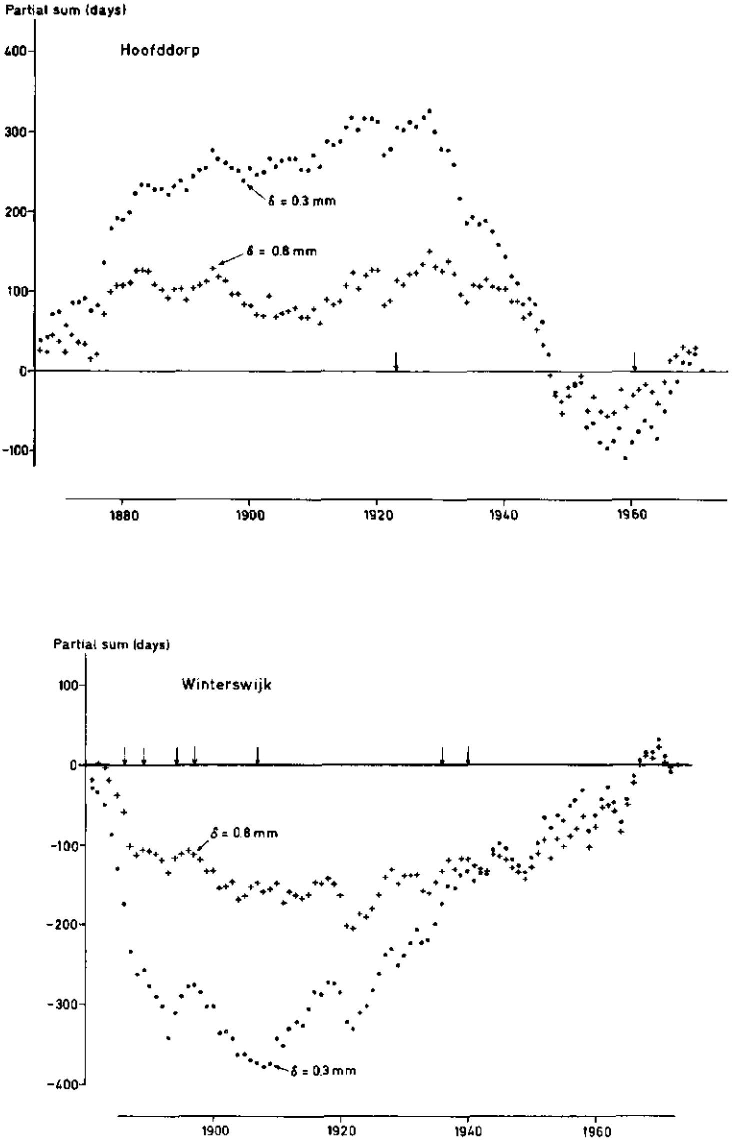


FIG. 6.3. Partial sums of departures from the mean of the annual number of wet days of Winterswijk (1881–1973) and Hoofddorp (1867–1971). Years with a change of observer are denoted by an arrow, except for Hoofddorp (1867–1909).

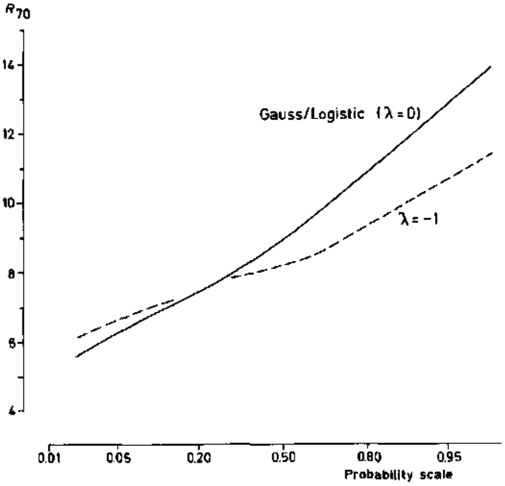


FIG. 6.4. Normal probability plots of R_{70} (see Equation (6.3)) for independent processes with a Gaussian or λ-distribution, based on 1,000 samples.

For Winterswijk (1881–1973) and Hoofddorp (1867–1971) the hypothesis of homogeneity is rejected at the 5 per cent level for $\delta = 0.3$ mm.

So far tests for homogeneity were based on one rainfall series only. More powerful tests can be obtained when the annual number of wet days at different stations are compared. For instance, realizations of the Von Neumann's ratio for the differences of the annual number of wet days of Winterswijk and Hengelo (1908–1973) are 0.334 and 0.541 for $\delta = 0.3$ and 0.8 mm, respectively. Even for the larger threshold there is evidence for non-homogeneity at the 5 per cent level, that is at least one series is non-homogeneous. A comparison of the annual number of wet days of Hoofddorp with averages of De Bilt, Gouda and Scheveningen for the period 1953–1971 leads to a significant jump in the mean of the Hoofddorp series, when the threshold is 0.3 mm. The stations De Bilt, Gouda and Scheveningen are chosen here because they have no missing data and no change of observer. So non-homogeneity of the wet-dry series of Hoofddorp for $\delta = 0.3$ mm is mainly due to the observations in the period 1922–1960.

7. SUMMARY

In this chapter the homogeneity of some Dutch rainfall series was investigated. A rainfall series was called homogeneous if the distribution of rainfall amounts is the same for every year. Homogeneity was investigated for the mean rainfall amount and for the mean of the wet-dry series.

Tests for homogeneity of the mean rainfall amount were based on monthly and annual totals. It was shown that annual totals are approximately Gaussian; for monthly totals the 'loi des fuites' (LDF) gives a good fit. Besides, absence of serial correlation in monthly or annual totals can be assumed. Two subjects about homogeneity of monthly and annual totals were considered, namely the influence of the reduction in height of Dutch rain gauges in the period 1946–1954 and the homogeneity of the Zwanenburg-Hoofddorp series.

Jumps in the mean during the period 1946–1954 were estimated and tested, with regression models for annual totals of Dutch and Belgian or German stations. For stations remote from the coast an increase in the mean of about 2 per cent was found; for stations in the coastal area the increase in the mean was sometimes more than 10 per cent. However, there was a large variation in the height of estimated jumps, due to differences in the degree of protection against the wind. A slight indication for larger jumps during the winter season was found from a comparison of monthly data of Dutch and German stations in the northern coastal area.

Significant differences between the means of the Zwanenburg series (1735–1860) and Hoofddorp series (1861–1972) were found. Besides, a poor correlation was found between simultaneous rainfall observations of Zwanenburg and those of other stations. The Hoofddorp series showed no significant departures from homogeneity.

Homogeneity of wet-dry series was tested with the annual number of wet days. Departures from homogeneity are possible because small rainfall amounts are often registered as zero. There are only a few long-term rainfall series in the Netherlands for which the wet-dry series is homogeneous, if a wet day is defined as a day with at least 0.3 mm rainfall. A lower bound of 0.8 mm for rainfall amounts on wet days seems to be more appropriate for the Netherlands.

APPENDICES

A1. Variances and covariances of estimators of the parameters of the gamma distribution and of the 'loi des fuites'

A1.1. *Variances and covariances of moment estimators*

An approximation of the covariance matrix of the moment estimators of the parameters of the gamma distribution can be obtained from a linearization of the relations (3.12) (cf. Kendall and Stuart (1969), 10.6):

$$\text{(A1.1)} \qquad \mathbf{COV}(\underline{\hat\lambda},\underline{\hat\nu}) \approx \mathbf{T} \cdot \mathbf{COV}(\underline{\bar x},\underline{s}^2) \cdot \mathbf{T}'$$

with $\mathbf{COV}(\underline{\hat\lambda},\underline{\hat\nu})$: covariance matrix of moment estimators,
$\mathbf{COV}(\underline{\bar x},\underline{s}^2)$: covariance matrix of the sample mean and variance,
$\mathbf{T}$: transformation matrix, which has the form:

$$\text{(A1.2)} \qquad \mathbf{T} = \begin{pmatrix} \dfrac{\delta\hat\lambda}{\delta\bar x} & \dfrac{\delta\hat\lambda}{\delta s^2} \\ & \\ \dfrac{\delta\hat\nu}{\delta\bar x} & \dfrac{\delta\hat\nu}{\delta s^2} \end{pmatrix}_{\bar x = \mu_1',\, s^2 = \mu_2} = \begin{pmatrix} \dfrac{1}{\mu_2} & -\dfrac{\mu_1'}{\mu_2^2} \\ & \\ \dfrac{2\mu_1'}{\mu_2} & -\dfrac{\mu_1'^2}{\mu_2^2} \end{pmatrix}.$$

When N, the number of observations, is large, the matrix $\mathbf{COV}(\underline{\bar x},\underline{s}^2)$ is approximately:

$$\text{(A1.3)} \qquad \mathbf{COV}(\underline{\bar x},\underline{s}^2) \approx \frac{1}{N}\begin{pmatrix} \mu_2 & \mu_3 \\ & \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}$$

(cf. Thom (1958)).

Substitution of (A1.2) and (A1.3) in (A1.1) gives for the second moments of the moment estimators:

$$\text{(A1.4a)} \qquad \operatorname{var}(\underline{\hat\lambda}) \approx \frac{1}{N}\left\{\frac{1}{\mu_2} + \frac{\mu_1'^2(\mu_4-\mu_2^2)}{\mu_2^4} - \frac{2\mu_1'\mu_3}{\mu_2^3}\right\}$$

$$\text{(A1.4b)} \qquad \operatorname{var}(\underline{\hat\nu}) \approx \frac{1}{N}\left\{\frac{4\mu_1'^2}{\mu_2} + \frac{\mu_1'^4(\mu_4-\mu_2^2)}{\mu_2^4} - \frac{4\mu_1'^3\mu_3}{\mu_2^3}\right\}$$

$$\text{(A1.4c)} \qquad \operatorname{cov}(\underline{\hat\lambda},\underline{\hat\nu}) \approx \frac{1}{N}\left\{\frac{2\mu_1'}{\mu_2} + \frac{\mu_1'^3(\mu_4-\mu_2^2)}{\mu_2^4} - \frac{3\mu_1'^2\mu_3}{\mu_2^3}\right\}.$$

For the approximations given above it does not matter whether one takes N of N-1 in the denominator of the variance estimate.

Substitution of (3.10a, b, c and d) in the right sides of (A1.4) gives:

$$\text{(A1.5a)} \qquad \operatorname{var}(\hat{\underline{\lambda}}) \approx \frac{\lambda^2}{N}(\frac{3}{\nu} + 2)$$

$$\text{(A1.5b)} \qquad \operatorname{var}(\hat{\underline{\nu}}) \approx \frac{2\nu}{N}(\nu + 1)$$

$$\text{(A1.5c)} \qquad \operatorname{cov}(\hat{\underline{\lambda}},\hat{\underline{\nu}}) \approx \frac{2\lambda}{N}(\nu + 1).$$

Equations (A1.5a and b) are also given by THOM (1958).

Because of the relations between moment estimators of the gamma distribution and those of the LDF, see (3.20), approximations of the variances of the moment estimators of the LDF follow from (A1.4) by replacing $\hat{\underline{\lambda}}$ by $\frac{1}{2}\hat{\underline{\rho}}$ and $\hat{\underline{\nu}}$ by $\frac{1}{2}$ $\hat{\underline{\theta}}$. After substitution of (3.18a, b, c and d) one obtains:

$$\text{(A1.6a)} \qquad \operatorname{var}(\hat{\underline{\rho}}) \approx \frac{2\rho^2}{N}(\frac{1}{\theta} + 1)$$

$$\text{(A1.6b)} \qquad \operatorname{var}(\hat{\underline{\theta}}) \approx \frac{2\theta^2}{N}(\frac{1}{\theta} + 1)$$

$$\text{(A1.6c)} \qquad \operatorname{cov}(\hat{\underline{\rho}},\hat{\underline{\theta}}) \approx \frac{\rho}{N}(2\theta + 1).$$

A1.2. *Variances and covariances of maximum likelihood estimators*

The asymptotic covariance matrix of the maximum likelihood estimators can be obtained from the expectations of the second derivatives of the logarithm of the likelihood function (cf. KENDALL and STUART (1973), 18.15, 18.16 and 18.26). This gives the following results for the ML estimators of the gamma distribution (cf. THOM (1958) and JOHNSON and KOTZ (1970), Chapter 17 (43)):

$$\text{(A1.7a)} \qquad \operatorname{var}(\hat{\underline{\lambda}}) \approx \frac{\lambda^2\psi'(\nu)}{N(\nu\psi'(\nu) - 1)}$$

$$\text{(A1.7b)} \qquad \operatorname{var}(\hat{\underline{\nu}}) \approx \frac{\nu}{N(\nu\psi'(\nu) - 1)}$$

$$\text{(A1.7c)} \qquad \operatorname{cov}(\hat{\underline{\lambda}},\hat{\underline{\nu}}) \approx \frac{\lambda}{N(\nu\psi'(\nu) - 1)}$$

where ψ' stands for the trigamma function (second derivative of the logarithm of the gamma function).

Explicit formulas for variances and covariances of ML estimators of the

LDF cannot be obtained easily. Estimates of variances and covariances were obtained from a numerical evaluation of the second derivatives of the logarithm of the likelihood function. Formulas for the second derivatives are given in Appendix A3.

A2. The probability distribution of the 'loi des fuites'

In this appendix the probability density and the cumulative distribution function (cdf) of the LDF are derived.

Let $\underline{n}_t$ be the number of showers in an interval of length t and let $\underline{x}_t$ be the rainfall amount in that interval, then:

$$\text{(A2.1)} \qquad P(\underline{x}_t \leqslant x) = P(\underline{n}_t = 0) + \sum_{k=1}^{\infty} P(\underline{x}_t \leqslant x | \underline{n}_t = k)\, P(\underline{n}_t = k).$$

If $\underline{n}_t = k$, then $\underline{x}_t$ is a sum of k iid exponential variables with scale parameter ρ. It is well known that this sum is gamma distributed with shape parameter k and scale parameter ρ. Further, $\underline{n}_t$ is Poisson distributed with mean $\theta = t/\mu$. So (A2.1) becomes:

$$\text{(A2.2)} \qquad P(\underline{x}_t \leqslant x) = e^{-\theta} + \sum_{k=1}^{\infty} \frac{\theta^k e^{-\theta}}{k!} \int_0^x \frac{\rho^k y^{k-1}}{\Gamma(k)} e^{-\rho y}\, dy.$$

For the integral on the right side of (A2.2) one can write:

$$\text{(A2.3)} \qquad \int_0^x \frac{\rho^k y^{k-1}}{\Gamma(k)} e^{-\rho y}\, dy = 1 - \sum_{i=0}^{k-1} \frac{(\rho x)^i}{i!} e^{-\rho x}$$

which can be obtained by integration by parts or using the fact that (A2.3) represents the cdf of the waiting time to the kth event in a Poisson process with rate ρ.

From (A2.2) it follows (by differentiation):

$$\text{(A2.4)} \qquad \begin{cases} P(\underline{x}_t = 0) = e^{-\theta} \\ P(x < \underline{x}_t < x + dx) = e^{-\theta} \displaystyle\sum_{k=1}^{\infty} \frac{(\rho\theta)^k}{k!} \frac{x^{k-1} e^{-\rho x}}{(k-1)!}\, dx \qquad x > 0. \end{cases}$$

This expression can also be obtained by expanding the Laplace-Stieltjes transform (3.14) in powers of $1/(\rho + s)$ and inverting the series term by term (cf. Cox (1962), Exercise 27).

Using Equation 9.6.10 of Abramowitz and Stegun (1970), one gets for $x > 0$:

$$\text{(A2.5)} \qquad P(x < \underline{x}_t < x + dx) = e^{-\theta - \rho x} \sqrt{\frac{\rho\theta}{x}}\, I_1 (2\sqrt{\rho\theta x})\, dx$$

where I_1 stands for a modified Bessel function of order 1 (cf. FISHER and CORNISH (1960), COX (1962), Equation (8.3.6) and BERNIER and FANDEUX (1970)).

A3. ESTIMATION OF THE PARAMETERS OF THE 'LOI DES FUITES' BY THE METHOD OF MAXIMUM LIKELIHOOD

Suppose there are N independent observations of which n are zero and $m = N - n$ are positive. The non-zero observations are denoted by $x_1, \ldots, x_m$.

The likelihood $L^*(\rho, \theta)$ follows from (A2.4):

$$\text{(A3.1)} \qquad L^*(\rho,\theta) = e^{-N\theta} \prod_{i=1}^{m} e^{-\rho x_i} \left\{ \sum_{k=1}^{\infty} \frac{(\rho\theta)^k x_i^{k-1}}{k!\,(k-1)!} \right\}.$$

Instead of maximizing $L^*(\rho, \theta)$ with respect to ρ and θ one can also maximize $L(\rho,\theta) = \log L^*(\rho,\theta)$. Taking logarithms in (A3.1) gives:

$$\text{(A3.2)} \qquad L(\rho,\theta) = -N\theta - \rho \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} \log \left\{ \sum_{k=1}^{\infty} \frac{(\rho\theta)^k x_i^{k-1}}{k!(k-1)!} \right\}.$$

Let $\lambda = \rho\theta$, then:

$$\text{(A3.3)} \qquad K(\lambda,\theta) = L(\lambda/\theta,\theta) = -N\theta - \frac{\lambda}{\theta} \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} \log h_i(\lambda)$$

where $h_i(\lambda)$ is given by:

$$\text{(A3.4)} \qquad h_i(\lambda) = \sum_{k=1}^{\infty} \frac{\lambda^k x_i^{k-1}}{k!(k-1)!}.$$

Maximization of $K(\lambda,\theta)$ proceeds in two stages. First, the log likelihood is maximized for fixed λ with respect to θ, and second, the result in the first step is maximized with respect to λ.

For fixed λ, one gets:

$$\text{(A3.5)} \qquad G(\lambda) = \max_{\theta} K(\lambda,\theta) = \max_{\theta} \left(-N\theta - \frac{\lambda}{\theta} \sum_{i=1}^{m} x_i \right) + \sum_{i=1}^{m} \log h_i(\lambda)$$

and it follows, by differentiation, that the maximum is attained for $\theta = \sqrt{\lambda \sum_{i=1}^{m} x_i/N}$. Substituting this value in (A3.5) gives:

$$\text{(A3.6)} \qquad G(\lambda) = -2\sqrt{\lambda N \sum_{i=1}^{m} x_i} + \sum_{i=1}^{m} \log h_i(\lambda).$$

The maximum of $G(\lambda)$ can be found by the iteration formula of Newton-Raphson:

$$\text{(A3.7)} \qquad \lambda_l = \lambda_{l-1} - G'(\lambda_{l-1})/G''(\lambda_{l-1}) \qquad\qquad l = 1, 2, \ldots$$

with:

(A3.8a) $$G'(\lambda) = -\lambda^{-1/2} \sqrt{N \sum_{i=1}^{m} x_i} + \sum_{i=1}^{m} \frac{h'_i}{h_i}$$

(A3.8b) $$G''(\lambda) = \tfrac{1}{2}\lambda^{-3/2} \sqrt{N \sum_{i=1}^{m} x_i} + \sum_{i=1}^{m} \left\{ \frac{h''_i}{h_i} - \left(\frac{h'_i}{h_i}\right)^2 \right\}$$

and:

(A3.9a) $$h'_i = \sum_{k=1}^{\infty} \frac{\lambda^{k-1} x_i^{k-1}}{(k-1)!\,(k-1)!}$$

(A3.9b) $$h''_i = \sum_{k=2}^{\infty} \frac{\lambda^{k-2} x_i^{k-1}}{(k-2)!\,(k-1)!} = \sum_{k=1}^{\infty} \frac{\lambda^{k-1} x_i^{k}}{(k-1)!\,k!} = \frac{x_i}{\lambda} h_i.$$

A remark from a computational point of view is that a too large starting value gives an overcorrection in the first iteration; so protection against negative values of λ_1 seems advisable.

For the evaluation of the variances and covariances of the ML estimators one needs expressions for the second derivatives of the log likelihood (see A1.2). From (A3.3) and (A3.6) it follows:

(A3.10) $$L(\rho,\theta) = -N\theta - \rho \sum_{i=1}^{m} x_i + G(\rho\theta) + 2c\sqrt{\rho\theta}$$

with $c = \sqrt{N \sum_{i=1}^{m} x_i}$.

Differentiation of (A3.10) gives:

(A3.11a) $$L_{\rho\rho} = \theta^2 G''(\rho\theta) - \tfrac{1}{2}c\sqrt{\theta/\rho^3}$$

(A3.11b) $$L_{\theta\theta} = \rho^2 G''(\rho\theta) - \tfrac{1}{2}c\sqrt{\rho/\theta^3}$$

(A3.11c) $$L_{\rho\theta} = \rho\theta G''(\rho\theta) + G'(\rho\theta) + \tfrac{1}{2}c/\sqrt{\rho\,\theta}.$$

At the optimum $G'(\rho\theta) = 0$; $G''(\rho\theta)$ can be obtained from (A3.8b).

A4. Bias in the sample correlation coefficient due to a jump in the mean

Suppose there are N simultaneous observations at the sites X and Y, then the sample correlation coefficient is:

(A4.1) $$r_{xy} = \frac{\tilde{c}_{xy}}{\sqrt{\tilde{s}_x^2 \tilde{s}_y^2}}$$

with: $$\tilde{c}_{xy} = \sum_{i=1}^{N} x_i y_i - \frac{1}{N} \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i,$$

$$\tilde{\underline{s}}_x^2 = \sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2 ,$$

$$\tilde{\underline{s}}_y^2 = \sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2 .$$

The quantities $\tilde{c}_{xy}$, $\tilde{s}_x^2$ and $\tilde{s}_y^2$ differ a factor N from the sample covariance of $\underline{x}$ and $\underline{y}$, the sample variance of $\underline{x}$ and the sample variance of $\underline{y}$, respectively.

For the expectation of $\underline{r}_{xy}$ one has approximately:

$$\text{(A4.2)} \qquad E(\underline{r}_{xy}) \approx \frac{E(\tilde{\underline{c}}_{xy})}{\sqrt{E(\tilde{\underline{s}}_x^2)\,E(\tilde{\underline{s}}_y^2)}} .$$

When dealing with two homogeneous series this relation leads to (4.1a). In this appendix approximations of $E(\underline{r}_{xy})$ are given for one jump in the mean at the site X.

Two models are considered, namely a model in which a jump in the mean is created by adding a constant to a part of the series (model a) and a model in which the realizations of a part of the series are multiplied by some factor (model b).

Model a reads:

$$\text{(A4.3)} \qquad \begin{cases} \underline{x}_i = \underline{\varepsilon}_i & i = 1, \ldots, n \\ \underline{x}_i = \underline{\varepsilon}_i + \delta & i = n+1, \ldots, N \\ \underline{y}_i = \underline{\eta}_i & i = 1, \ldots, N \end{cases}$$

with: $E(\underline{\varepsilon}_i) = \mu_\varepsilon ; \mathrm{cov}(\underline{\varepsilon}_i,\underline{\varepsilon}_j) = \sigma_\varepsilon^2 \delta_{ij}$

$E(\underline{\eta}_i) = \mu_\eta ; \mathrm{cov}(\underline{\eta}_i,\underline{\eta}_j) = \sigma_\eta^2\, \delta_{ij} ; \mathrm{cov}(\underline{\varepsilon}_i,\underline{\eta}_j) = \sigma_{\varepsilon\eta}\delta_{ij}$

$(\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j)$.

Taking expectations of sample variances and covariances gives:

$$\text{(A4.4a)} \qquad E(\tilde{\underline{c}}_{xy}) = \frac{N-1}{N} \sum_{i=1}^{N} E(\underline{x}_i \underline{y}_i) - \frac{1}{N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} E(\underline{x}_i)\, E(\underline{y}_j)$$

$$\text{(A4.4b)} \qquad E(\tilde{\underline{s}}_x^2) = \frac{N-1}{N} \sum_{i=1}^{N} E(\underline{x}_i^2) - \frac{2}{N} \sum_{i=1}^{N} \sum_{j=1}^{i-1} E(\underline{x}_i)\, E(\underline{x}_j)$$

$$\text{(A4.4c)} \qquad E(\tilde{\underline{s}}_y^2) = \frac{N-1}{N} \sum_{i=1}^{N} E(\underline{y}_i^2) - \frac{2}{N} \sum_{i=1}^{N} \sum_{j=1}^{i-1} E(\underline{y}_i)\, E(\underline{y}_j) .$$

Substitution of (A4.3) in (A4.4) gives, after some algebra:

(A4.5a) $$E(\underline{\tilde{c}}_{xy}) = (N-1)\,\sigma_{\varepsilon\eta}$$

(A4.5b) $$\begin{aligned} E(\underline{\tilde{s}}_x^2) &= (N-1)\,\sigma_\varepsilon^2 + Nq(1-q)\,\delta^2 \\ &\approx (N-1)\,\sigma_\varepsilon^2\,[1 + q\,(1-q)\,\delta^2/\sigma_\varepsilon^2] \end{aligned}$$

(cf. YEVJEVICH and JENG (1969), Equation (18))

(A4.5c) $$E(\underline{\tilde{s}}_y^2) = (N-1)\,\sigma_\eta^2$$

where $q = (N-n)/N$.

Substitution of (A4.5) in (A4.2) finally yields:

(A4.6) $$E(\underline{r}_{xy}) \approx \frac{\sigma_{\varepsilon\eta}}{\sigma_\varepsilon\sigma_\eta\,\sqrt{1+q(1-q)\delta^2/\sigma_\varepsilon^2}} = \frac{\rho_{\varepsilon\eta}}{\sqrt{1+q(1-q)\delta^2/\sigma_\varepsilon^2}}\,.$$

So the (negative) bias in $\underline{r}_{xy}$ depends on the ratio $\delta/\sigma_\varepsilon$ and q (influence is maximal if $q = \frac{1}{2}$). For $q = \frac{1}{3}$ (this is about the value for q in Figure 4.1, when a jump in the mean of Dutch rainfall series is assumed in the period 1946–1954) one finds for $E(\underline{r}_{xy})$:

$\rho_{\varepsilon\eta}$	$\delta^2/\sigma_\varepsilon^2 = 0.60$	$\delta^2/\sigma_\varepsilon^2 = 0.15$
0.90	0.85	0.88
0.70	0.66	0.69
0.50	0.47	0.49

For Dutch rainfall series the value 0.60 for $\delta^2/\sigma_\varepsilon^2$ corresponds to a jump in the mean which is a bit more than 10 per cent. Even then, the difference between $E(\underline{r}_{xy})$ and $\rho_{\varepsilon\eta}$ is hardly noticeable, because the standard deviation of $\underline{r}_{xy}$ is quite large as a consequence of the small number of observations. Only when $\rho_{\varepsilon\eta}$ is very large the standard deviation of $\underline{r}_{xy}$ is sufficiently small (see Equation (4.1b)).

If one takes non-overlapping sums of p x_is, the ratio $\delta^2/\sigma_\varepsilon^2$ changes proportional with p. For monthly totals this ratio differs about a factor 12 from that of annual totals and consequently, the bias in $\underline{\tilde{s}}_x^2$ and $\underline{r}_{xy}$ is negligible for monthly totals.

Model b reads:

(A4.7) $$\begin{cases} \underline{x}_i = \underline{\varepsilon}_i & i = 1, \ldots, n \\ \underline{x}_i = (1+f)\underline{\varepsilon}_i & i = n+1, \ldots, N \\ \underline{y}_i = \underline{\eta}_i & i = 1, \ldots, N \end{cases}$$

where $\underline{\varepsilon}_i$ and $\underline{\eta}_i$ have the same properties as in model a.

Substitution of (A4.7) in (A4.4) gives, after some algebra:

$$\text{(A4.8a)} \qquad E(\underline{\tilde{c}}_{xy}) = (N-1)\,\sigma_{\varepsilon\eta}\,[1+qI]$$

$$\text{(A4.8b)} \qquad E(\underline{s}_x^2) = (N-1)\,[1+2qI+qI^2]\,\sigma_\varepsilon^2 + Nq(1-q)I^2\mu_\varepsilon^2$$

$$\approx (N-1)\,\sigma_\varepsilon^2\,[1+2qI+qI^2+q(1-q)I^2\mu_\varepsilon^2/\sigma_\varepsilon^2]$$

(cf. YEVJEVICH and JENG (1969), Equation (40)).

$$\text{(A4.8c)} \qquad E(\underline{s}_y^2) = (N-1)\,\sigma_\eta^2$$

with $q = (N-n)/N$.

Substitution of (A4.8) in (A4.2) gives:

$$\text{(A4.9)} \qquad E(\underline{r}_{xy}) \approx \frac{(1+qI)\rho_{\varepsilon\eta}}{\sqrt{1+2qI+qI^2+(1-q)qI^2\mu_\varepsilon^2/\sigma_\varepsilon^2}}\,.$$

For this model the bias in $\underline{r}_{xy}$ depends on μ_ε. If $q = \frac{1}{3}$ and $\mu_\varepsilon^2/\sigma_\varepsilon^2 = 45$ (cf. Table 2.1) one finds for $E(\underline{r}_{xy})$:

$\rho_{\varepsilon\eta}$	$I = 0.10$	$I = 0.05$
0.90	0.84	0.88
0.70	0.65	0.69
0.50	0.47	0.49

So the results of this model are nearly the same as those of model a.

A5. Estimation of the parameters of model 1 by the method of maximum likelihood

In this appendix the ML estimates of the parameters in the regression model (4.13) are derived. The $\underline{y}_k$s are independently and normally distributed; for $k = 1, \ldots, n$ the mean is $\beta x_k/f$ and the variance is σ^2/f^2; for $k = n+1, \ldots, N$ the mean is βx_k and the variance is σ^2. Hence, the likelihood is:

$$\text{(A5.1)} \qquad L^*(\beta,\sigma,f) = \left(\frac{\sigma}{f}\sqrt{2\pi}\right)^{-n} \exp\left\{-\tfrac{1}{2}f^2 \sum_{k=1}^{n} (y_k-\beta x_k/f)^2/\sigma^2\right\} \times$$

$$\times\ (\sigma\sqrt{2\pi})^{-(N-n)}\exp\left\{-\tfrac{1}{2}\sum_{k=n+1}^{N}(y_k-\beta x_k)^2/\sigma^2\right\}.$$

Taking logarithms in (A5.1) gives:

$$\text{(A5.2)} \qquad L(\beta,\sigma,f) = \log L^*(\beta,\sigma,f) = -N\log\sqrt{2\pi} - N\log\sigma + n\log f +$$

$$-\frac{1}{2\sigma^2}\left\{\sum_{k=1}^{n}(fy_k-\beta x_k)^2 + \sum_{k=n+1}^{N}(y_k-\beta x_k)^2\right\}\,.$$

In the first instance $L(\beta,\sigma,f)$ is maximized for fixed f with respect to β and σ, and afterwards the log likelihood is maximized with respect to f.

Let:

$$\text{(A5.3)} \qquad G(f) = \max_{\beta,\sigma} L(\beta,\sigma,f).$$

For fixed f, (4.13) is a linear regression model and consequently the likelihood attains its maximum value for:

$$\text{(A5.4)} \qquad \hat{\beta} = \left(f \sum_{k=1}^{n} x_k y_k + \sum_{k=n+1}^{N} x_k y_k\right) / \sum_{k=1}^{N} x_k^2$$

and

$$\text{(A5.5)} \qquad N\hat{\sigma}^2 = f^2 \sum_{k=1}^{n} y_k^2 + \sum_{k=n+1}^{N} y_k^2 - \left(f \sum_{k=1}^{n} x_k y_k + \sum_{k=n+1}^{N} x_k y_k\right)^2 / \sum_{k=1}^{N} x_k^2 .$$

Substitution of (A5.4) and (A5.5) in (A5.3) gives:

$$\text{(A5.6)} \qquad G(f) = -N \log\sqrt{2\pi} - N \log\hat{\sigma} + n \log f - N/2$$

using (A5.2).

The value $\hat{f}$ for which $G(f)$ attains its maximum can be found by the iteration formula of Newton-Raphson:

$$\text{(A5.7)} \qquad f_l = f_{l-1} - (1-\omega^l)\, G'(f_{l-1})/G''(f_{l-1}) \qquad l = 1, 2, \ldots$$

For the relaxation factor ω the value 0.9 was chosen and the starting value f_0 was taken to be 1. The first and second derivatives of $G(f)$ are:

$$\text{(A5.8)} \qquad G'(f) = -\frac{N}{\hat{\sigma}}\hat{\sigma}_f + \frac{n}{f}$$

where

$$\text{(A5.9)} \qquad N\hat{\sigma}_f = \frac{1}{\hat{\sigma}}\left\{f \sum_{k=1}^{n} y_k^2 - \sum_{k=1}^{n} x_k y_k \left[f \sum_{k=1}^{n} x_k y_k + \sum_{k=n+1}^{N} x_k y_k\right] / \sum_{k=1}^{N} x_k^2\right\} = \frac{A}{\hat{\sigma}}$$

and

$$\text{(A5.10)} \qquad G''(f) = -\frac{1}{\hat{\sigma}^2} A_f + \frac{2A^2}{N\hat{\sigma}^4} - \frac{n}{f^2}$$

with

$$\text{(A5.11)} \qquad A_f = \sum_{k=1}^{n} y_k^2 - \left(\sum_{k=1}^{n} x_k y_k\right)^2 / \sum_{k=1}^{N} x_k^2 .$$

It can be shown that the variance of $\hat{\underline{f}}$ follows from (cf. RICHARDS (1961)):

(A5.12) $\mathrm{var}\,\hat{\underline{f}} \approx 1/E(\underline{G}''(f))$

which is estimated as:

(A5.13) $s_{\hat{f}}^2 = -1/G''(\hat{f})$.

III. ANALYSIS OF DAILY RAINFALL DATA FROM DUTCH STATIONS

1. Introduction

In this chapter the analysis of the daily rainfall series of Winterswijk (1908–1973) and Hoofddorp (1867–1971) is discussed. The Winterswijk series from 1881 up to 1908 was not analysed because of its poor quality (see II, 6). Besides, some results are given for the series of Hengelo (1908–1973) which was analysed in a later stage.

The data are analysed in two steps. First, the occurrence of wet and dry days is described and second, the modelling of rainfall amounts on wet days is studied. A wet day is defined as a day with a rainfall amount of at least δ millimeters. To study the influence of the height of the threshold the rainfall series of Winterswijk was analysed with thresholds of $\delta = 0.3$ mm and $\delta = 0.8$ mm. The series of Hoofddorp and Hengelo were only analysed with $\delta = 0.8$ mm for quality reasons (see II, 6).

In the subsequent sections different aspects of the daily rainfall model are discussed. Section 2 describes how the influence of seasonal changes was reduced. In Section 3, 4 and 5 the parameters are estimated and some assumptions underlying the model are tested. Characteristics of the historic sequence and the model are compared in Section 6 and 7. Some features of the model (correlograms, variance-time curves and for simple cases the cumulative distribution functions of k-day totals ($k = 1, 2, \ldots$) can be obtained by numerical computations. The formulas underlying these computations will be derived in Chapter IV. Characteristics which could not be easily derived by numerical computations were obtained by Monte Carlo simulation.

2. The reduction of seasonal variation

In II,3.1 the estimation of serial correlation coefficients was based on standardized values to reduce the effect of seasonal changes in mean and standard deviation. However, the use of such transformations is not attractive, when dealing with a stochastic model with a separate process for the occurrence of wet and dry days (shortly denoted as wet-dry process). Therefore, the estimation of the parameters and the testing of some assumptions underlying the model were done for each month separately. Sometimes periods of three months were combined to seasons. The seasons, which are distinguished here, are: winter (December–February), spring (March–May), summer (June–August) and autumn (September–November).

Special rules have to be devised for wet or dry spells which extend across the limits of a period (month or season). A wet interval or spell is defined here as a sequence of wet days, on each side bounded by a dry day. A dry spell can be defined analogously. The analysis by month (or season) should be done in such a way that wet or dry spells are not split up. To satisfy this requirement different methods can be followed. Those used here are:

A Wet or dry intervals are assigned to the period in which they begin.

B Wet or dry intervals are assigned to the period in which they end.

WD A wet-dry cycle is assigned to the period in which the dry spell begins. A wet-dry cycle consists of a wet spell and its following dry spell.

DW A dry-wet cycle is assigned to the period in which the wet spell begins. A dry-wet cycle can be defined in the same manner as a wet-dry cycle.

Rainfall amounts (on wet days) are assigned to the period to which the corresponding wet interval belongs, except for the estimation of the correlation coefficient between the length of a dry spell and the rainfall amount on the day following that spell (see 4.1). Days before the first complete spell or cycle in a historic record are discarded in the analysis. The same is done with days after the last complete spell or cycle.

For Winterswijk and Hengelo the analysis by dry-wet or wet-dry cycles relates to the period December 1907–November 1973; for Hoofddorp it relates to the period March 1867–February 1972.

In the following text, tables and figures the capitals A, B, WD or DW are usually followed by the height of the threshold in tenths of millimeters.

The methods A and B are used for parameter estimation. These methods are preferred for this purpose because the stochastic model will be used to generate synthetic sequences. This is explained further in Section 7. For estimating serial correlation coefficients, use is made of the methods WD and DW because it is desirable that if a particular period ends with a wet spell in some year it begins with a dry spell in the next year.

3. Analysis of the occurrence of wet and dry days

It is well known that the probability of a day being wet or dry generally depends on past conditions. In statistical models usually one of the following stochastic processes is used to describe the persistence in the occurrence of wet and dry days:

a. Two-state (namely wet or dry) Markov chains of a certain order (cf. Lowry and Guthrie (1968) and Dumont and Boyce (1974)). Here the assumption is made that the probability of some state on any day only depends on the states of a certain number (the order of the chain) of previous days.

b. Alternating renewal processes (cf. Cole and Sherriff (1972) and Quélennec (1973)). Here it is assumed that the lengths of successive spells are independent. A more complete définition of an alternating renewal process will be given in Chapter IV.

There is an overlap between these methods because first and second order two-state Markov chains are also alternating renewal processes. In this study the alternating renewal process is preferred, because:

a. It was expected to give a better fit for the occurrence of very long spells.

b. The number of parameters in a model with such a process can be reduced more easily.

In Section 3.1 the adequacy of the alternating renewal process is tested. The distribution of the lengths of wet and dry spells is discussed in Section 3.2.

3.1. *A test for an alternating renewal process*

Tests for renewal processes of which the alternating renewal process is a generalization (see Chapter IV) are usually based on second-moment properties (serial correlation, spectrum) of the intervals (cf. Cox and Lewis (1966), 6.4). The assumption that the wet-dry process is an alternating renewal process is tested here with correlation coefficients between lengths of successive wet and dry spells (these correlation coefficients should be zero for an alternating renewal process).

For the estimation of the correlation coefficient between the length of a wet spell and the length of its following dry spell, the rainfall series was analysed by wet-dry cycles (WD). When the historic series is analysed in this manner, all wet and dry spells belonging to a particular period can be used for the estimation of the correlation coefficient of that period. An analysis by dry-wet cycles (DW) should lead to the problem that for every year the last wet spell of some period has no successive dry spell, belonging to the same year and period. For this reason a DW analysis was used to estimate the correlation coefficient between the length of a dry spell and its following wet spell.

A survey of estimated correlation coefficients is given in Table 3.1. The number of cycles in a season is about 1000 for the Winterswijk and Hengelo series and about 1600 for the Hoofddorp series. If the intervals are independently and normally distributed, it follows from II, (4.1b) that the standard deviation of the estimated correlation coefficient should be about 0.033 for the Winterswijk and Hengelo series and about 0.025 for the Hoofddorp series. These approximated values of the standard deviation can also be used when the distribution of the intervals is non-normal, but the usefulness of the asymptotic normality of the correlation estimator can be doubtful in this case. The tabulated correlation coefficients and their standard deviations do not support a

Table 3.1. Estimated correlation coefficients between lengths of wet and dry spells.

Season	Winterswijk ($\delta = 0.3$ mm)		Winterswijk ($\delta = 0.8$ mm)		Hengelo ($\delta = 0.8$ mm)		Hoofddorp ($\delta = 0.8$ mm)	
	WD	DW	WD	DW	WD	DW	WD	DW
Winter	-0.047	0.001	-0.008	-0.016	-0.056	-0.044	-0.033	-0.080
Spring	-0.007	0.001	0.054	0.013	0.051	-0.028	-0.018	-0.042
Summer	0.005	-0.035	0.042	-0.068	0.012	-0.054	-0.046	-0.012
Autumn	-0.060	-0.045	-0.054	-0.075	-0.027	-0.062	-0.042	0.034

real correlation, so the assumption of an alternating renewal process seems reasonable. The same conclusion was reached by COLE and SHERRIFF (1972) for rainfall series in the River Dee catchment (Wales) and by QUÉLENNEC (1973) for French rainfall series.

Therefore it will be assumed that the wet-dry process is an alternating renewal process.

3.2. *The distribution of the lengths of wet and dry spells*

The probability distributions, which are fitted here to the lengths of wet and dry spells are all modifications of the negative binomial distribution (NBD). These modifications are defined at the beginning of this section. Then the parameters are estimated and the goodness of fit is tested. Finally the seasonal variation of the parameters is investigated for a particular case.

The NBD with parameters p and r can be defined by its probability function:

$$(3.1) \qquad P(\underline{x} = k) = \binom{k+r-1}{k} p^r (1-p)^k = \binom{-r}{k} p^r (-q)^k \qquad k = 0, 1, \ldots$$

with $0 < p \leqslant 1$, $r \geqslant 0$ and $q = 1-p$.

This distribution cannot be applied directly to describe the distribution of the lengths of weather spells, because these spells have always a length of at least one day. Therefore, one of the following modifications of (3.1) can be used:

a. A shift of the origin by one day which gives the probability function:

$$(3.2) \qquad P(\underline{y} = k) = P(\underline{x} = k-1) = \binom{k+r-2}{k-1} p^r q^{k-1} \qquad k = 1, 2, \ldots$$

where $\underline{y}$ is the random length (in days) of a spell. This distribution will be called the shifted negative binomial distribution (SNBD).

b. Truncation at zero which leads to the following probability function:

$$(3.3) \qquad P(\underline{y} = k) = P(\underline{x} = k | \underline{x} \geqslant 1) = \binom{k+r-1}{k} \frac{p^r q^k}{1-p^r} \qquad k = 1, 2, \ldots$$

Equation (3.3) still defines a probability function, for $-1 \leqslant r < 0$. The distribution defined by this equation will be called the truncated negative binomial distribution (TNBD).

The probability functions of the SNBD and the TNBD are monotonically decreasing in k if $rq < 1$. This is usually so for the distribution of lengths of weather spells, as shown by Table 3.2 where frequency distributions of lengths of weather spells from Winterswijk A8 are given. When $rq > 1$ there is a mode for $k \neq 1$.

There are some special cases of the SNBD and the TNBD which could be of interest for the distribution of lengths of weather spells:

a. If $r = 1$ in (3.2) or (3.3) one gets the geometric distribution (GD):

$$(3.4) \qquad P(\underline{y} = k) = pq^{k-1} \qquad k = 1,2, \ldots$$

TABLE 3.2. Number of weather spells with length k (days) for Winterswijk A8.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	>15
Dry spells																
January	134	59	37	21	27	17	8	7	3	2	3	7	4	0	3	7
July	126	71	52	21	20	13	9	10	4	4	5	6	1	1	1	7
Wet spells																
January	149	69	44	26	24	7	9	3	1	3	0	0	0	0	1	2
July	152	83	54	21	15	8	8	4	3	2	0	0	0	1	1	0

Wet and dry intervals have geometric distributions if the wet-dry process is a first order Markov chain (see Chapter IV).

b. If $r \to 0$ in (3.3) one gets (cf. KENDALL and STUART (1969), 5.16) the logarithmic series distribution (LSD):

$$P(\underline{y}=k)=\alpha\frac{q^k}{k} \qquad k=1,2,\ldots \tag{3.5}$$

with $\alpha = -1/\log p$. The LSD has been widely used to fit the distribution of lengths of weather spells (cf. WILLIAMS (1952), COOKE (1953), LAWRENCE (1954) and RAMABHADRAN (1954)).

The mean, the variance and the third central moment of the distributions, cited above, are given in Table 3.3.

Parameters of the different modifications of the NBD can be estimated by the method of maximum likelihood (ML). For the SNBD the likelihood equations have the same form as those of the NBD, because this distribution only

TABLE 3.3. Moments of the negative binomial distribution and some of its modifications.

Name	μ_1'	μ_2	μ_3
NBD	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\frac{q(1+q)r}{p^3}$
SNBD	$\frac{rq}{p}+1$	$\frac{rq}{p^2}$	$\frac{q(1+q)r}{p^3}$
TNBD	$\frac{rq}{p(1-p^r)}$	$\frac{rq[1-p^r(1+rq)]}{p^2(1-p^r)^2}$	$\frac{rq+3r^2q^2+rq^2+r^3q^3}{p^3(1-p^r)}-3\frac{r^2q^2+r^3q^3}{p^3(1-p^r)^2}+$ $+\ 2\ \frac{r^3q^3}{p^3(1-p^r)^3}$
GD	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{q(1+q)}{p^3}$
LSD	$\frac{\alpha q}{p}$	$\frac{\alpha q(1-\alpha q)}{p^2}$	$\frac{\alpha q(1+q-3\alpha q+2\alpha^2q^2)}{p^3}$

involves a shift of one unit. Likelihood equations of the NBD have been given by Haldane (1941), Wise (1946), Fisher (1953), van Montfort (1966) and Johnson and Kotz (1969). The iterative solution of the likelihood equations was discussed in detail by Wise (1946) and van Montfort (1966).

From (3.3) one gets for the log likelihood L of the TNBD:

$$L = N\log\left(\frac{r}{1-p^r}\right) + Nr\log p + \log(1-p)\sum_{k=1}^{\infty} kn_k - \sum_{k=1}^{\infty} n_k \log k! + \tag{3.6}$$

$$+ \sum_{k=2}^{\infty} n_k \sum_{j=2}^{k} \log(r+j-1) \qquad r > -1 \text{ and } r \neq 0$$

with n_k = number of observations equal to k,

$N = \sum_{k=1}^{\infty} n_k$ = total number of observations.

For $r = 0$ one must use the log likelihood of the LSD. Differentiation of (3.6) yields the likelihood equations as given by Sampford (1955). Initial estimates for the iterative solution can be based on the relations:

$$p = \mu_1' \left(1-P(\underline{y} = 1)\right)/\mu_2 \tag{3.7a}$$

$$r = \frac{p\mu_1' - P(\underline{y} = 1)}{1-p} \tag{3.7b}$$

which follow from (3.3) and the expressions for μ_1' and μ_2 in Table 3.3 (cf. Brass (1958)).

The ML estimate of the parameter p of the GD is the reciprocal of the sample mean. The solution of the likelihood equations of the LSD was discussed by Barton et al. (1963) and Johnson and Kotz (1969). An initial estimate can be based on $P(\underline{y} = 1)$ or a quadratic approximation of the equation given by Barton et al. (1963).

A relaxation factor of 0.9 was used when the likelihood equations were solved iteratively by the Newton-Raphson procedure.

Some critical levels of the X^2-test of goodness of fit are given in Table 3.4 for the modifications of the NBD.

For the application of the X^2-test, the range of interval lengths $(1, 2, \ldots, \infty)$ was divided into classes. The partitioning started from $y = 1$ and ran upwards in such a way that the estimated expected number of intervals in a certain class was at least 5 and as small as possible; if the remaining expected number of intervals was less than 5 the last class was extended to infinity. For the GD and the LSD the class intervals were the same as for the TNBD, which possibly gives a better comparison of the lack of fit of these three distributions.

Table 3.4 leads to the following three conclusions:

a. The SNBD and the TNBD fit well in nearly all months. For Dutch rainfall

TABLE 3.4. Critical levels of the X^2-test of goodness of fit for different distributions for the lengths of dry and wet spells. Small values indicate poor fit.

	Dry spells							
	Winterswijk A3				Winterswijk A8			
Month	SNBD	TNBD	GD	LSD	SNBD	TNBD	GD	LSD
January	0.455	0.742	0.004	0.604	0.324	0.104	0.000	0.094
February	0.209	0.558	0.000	0.499	0.037	0.099	0.000	0.127
March	0.169	0.080	0.000	0.101	0.063	0.017	0.000	0.016
April	0.884	0.841	0.311	0.050	0.579	0.616	0.069	0.039
May	0.479	0.671	0.018	0.195	0.825	0.789	0.001	0.418
June	0.454	0.395	0.015	0.064	0.826	0.774	0.153	0.057
July	0.876	0.865	0.002	0.774	0.180	0.641	0.004	0.247
August	0.012	0.061	0.000	0.076	0.068	0.339	0.000	0.208
September	0.605	0.522	0.006	0.203	0.107	0.088	0.000	0.049
October	0.649	0.512	0.000	0.587	0.438	0.269	0.000	0.108
November	0.459	0.438	0.010	0.237	0.916	0.809	0.058	0.171
December	0.286	0.323	0.000	0.341	0.375	0.283	0.009	0.041
	Wet spells							
January	0.950	0.940	0.503	0.054	0.285	0.275	0.135	0.024
February	0.960	0.984	0.561	0.205	0.860	0.902	0.455	0.403
March	0.981	0.985	0.634	0.155	0.677	0.680	0.372	0.161
April	0.617	0.561	0.358	0.026	0.200	0.219	0.100	0.054
May	0.928	0.930	0.842	0.000	0.851	0.902	0.648	0.000
June	0.240	0.270	0.299	0.002	0.226	0.208	0.291	0.000
July	0.979	0.966	0.329	0.264	0.473	0.561	0.506	0.015
August	0.290	0.286	0.150	0.004	0.376	0.522	0.325	0.038
September	0.347	0.356	0.235	0.011	0.025	0.110	0.254	0.000
October	0.783	0.771	0.308	0.076	0.386	0.345	0.434	0.005
November	0.718	0.707	0.443	0.011	0.024	0.021	0.009	0.001
December	0.558	0.626	0.364	0.004	0.597	0.476	0.524	0.003

series both modifications of the NBD are nearly equivalent. The SNBD has a simpler form than the TNBD. The TNBD, however, has the advantage that it includes both the GD and the LSD as special cases. Besides, there are some tropical rainfall series for which the TNBD fits the length of dry spells better (see V, 2.2). In the remainder of this chapter only the TNBD is considered.

b. The GD gives a good fit for wet spells in nearly all months. For dry spells the fit is nearly always poor. In general, the GD gives too few long spells and too few spells with a length of one day.

c. The LSD fits lengths of dry spells well. For wet spells, however, the fit is mostly poor. In general the LSD gives too many long spells and too many spells with a length of one day.

The Winterswijk series, analysed by method B, and the series of Hoofddorp and Hengelo give similar results.

A possible seasonal dependence of the distribution of lengths of wet or dry spells can be tested by the likelihood ratio (LR) test. In general the LR test is based on:

$$(3.8) \qquad l^* = \sup_{H_0} L^*(\theta) / \sup_{H_0 \cup H_1} L^*(\theta)$$

where $L^*(\boldsymbol{\theta})$ denotes the likelihood at some point $\boldsymbol{\theta}$ of the parameter space. In the denominator the likelihood is maximized with no restriction on the parameter space, while in the numerator the likelihood is maximized with, say n, independent constraints on the parameter space. In consequence l^* is always less than or equal to 1. There is no evidence for H_1 when l^* is close to 1. As test statistic, however, one usually takes:

$$(3.9) \qquad l = -2\log l^*$$

which is under H_0 asymptotically a realization of a χ^2_n-variable, provided that some regularity conditions are fulfilled (cf. KENDALL and STUART (1973), 24.7).

The seasonal dependence of the parameters of the TNBD is discussed extensively in the remainder of this section. The test statistic in (3.9) involves the difference of the log likelihood maximized under the unrestricted model and the log likelihood maximized under the restricted model. Under the assumption of independence the log likelihood used for testing seasonal variation is the sum of the twelve monthly values, given by (3.6). In the unrestricted model there are 24 parameters, namely two for each month: p and r. If one wants to test the hypothesis that there is no seasonal variation in both p and r, there are only two parameters in the restricted model. The asymptotic χ^2-distribution of $\underline{l}$ has therefore 22 degrees of freedom.

Realizations of the LR statistic (l_1) and their critical levels (C.L.) are given in the first and third pair of columns of Table 3.5 for dry and wet spells respectively. In most cases H_0 is rejected at the 5 per cent level. Especially for the Hoofddorp series the values of the test statistic are very large, partly because the Hoofddorp series is longer than the Winterswijk series.

The estimated parameters are given in Figure 3.1 for some cases. An asterisk in this figure denotes the estimated parameters of the TNBD fitted to all wet or dry spells. There exists a strong correlation between $\underline{\hat{p}}$ and $\underline{\hat{r}}$. The smallest and the largest estimated correlation coefficient between the ML estimators of p and r are respectively 0.852 and 0.899 for dry intervals of Winterswijk A8; 0.918 and 0.977 for wet intervals of Winterswijk A8; 0.842 and 0.896 for dry spells of Hoofddorp A8, and 0.889 and 0.967 for wet spells of Hoofddorp A8. The estimates of this correlation coefficient were obtained from the second derivatives of the logarithm of the likelihood function (cf. KENDALL and STUART (1973), 18.15, 18.16 and 18.26). Figure 3.1 shows that the estimated parameters change irregularly from month to month, mainly because the estimates of the parameters are strongly correlated. For wet spells of Winterswijk A8, confidence regions for the parameters p and r are given for some months in Figure 3.2. These confidence regions give the points for which

TABLE 3.5. Results of LR tests for reduction of the number of parameters of monthly fitted TNBDs for lengths of weather spells. The realizations l_2 and l_3 (and their corresponding critical levels) are based on approximate ML estimates. For some cases the values based on the exact ML estimates are given between brackets, under the values based on the approximate ML estimates.

	Dry spells				Wet spells					
	p and r constant		r constant		p and r constant		r constant		r constant and p sinusoidal	
	l_1	C.L.	l_2	C.L.	l_1	C.L.	l_2	C.L.	l_3	C.L.
Winterswijk A3	43.3	0.004	8.4	0.680	49.8	0.001	9.9	0.537	27.1	0.133
Winterswijk A8	27.9	0.180	9.3 (9.2)	0.593 (0.603)	35.6	0.033	16.9 (15.4)	0.110 (0.167)	25.5 (23.6)	0.185 (0.260)
Hengelo A8	28.0	0.174	8.0	0.714	34.4	0.045	9.8	0.545	22.2	0.332
Hoofddorp A8	104.6	0.000	36.5 (36.5)	0.000 (0.000)	170.0	0.000	22.5 (21.5)	0.020 (0.028)	32.7 (31.5)	0.036 (0.050)
Winterswijk B3	50.6	0.000	10.7	0.471	45.2	0.003	7.1	0.793	23.0	0.290
Winterswijk B8	37.6	0.021	8.8	0.639	31.3	0.090	15.4	0.163	21.9	0.344
Hengelo B8	32.8	0.064	4.2	0.963	32.2	0.074	8.6	0.660	21.8	0.352
Hoofddorp B8	91.8	0.000	23.9	0.013	155.7	0.000	16.6	0.121	21.9	0.347

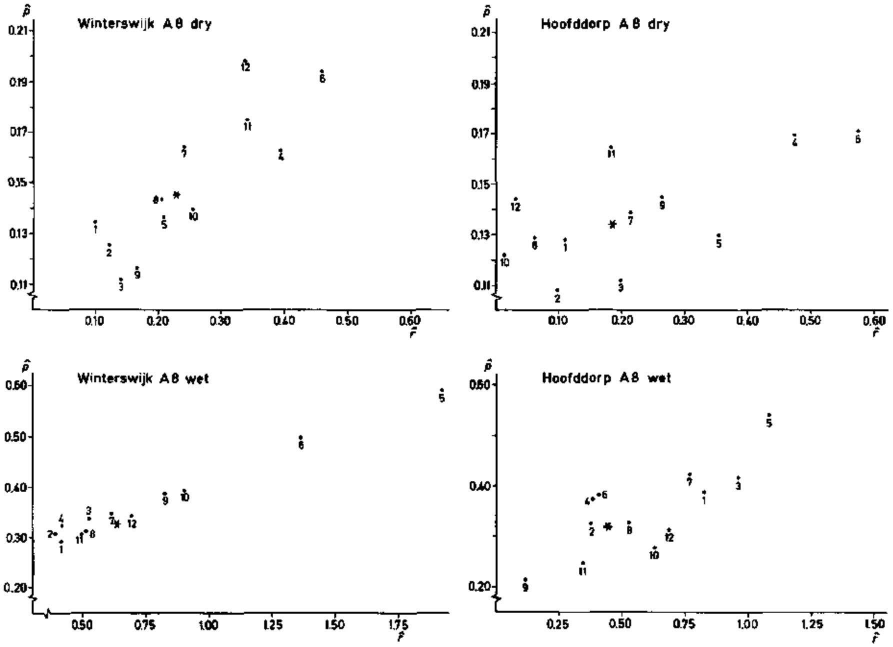


FIG. 3.1. ML estimates of the parameters p and r of the TNBD fitted to lengths of weather spells of Winterswijk A8 and Hoofddorp A8. The number attached to each point is the serial number of the month (1 corresponds to January). The annual estimates of the parameters p and r are denoted by an asterisk.

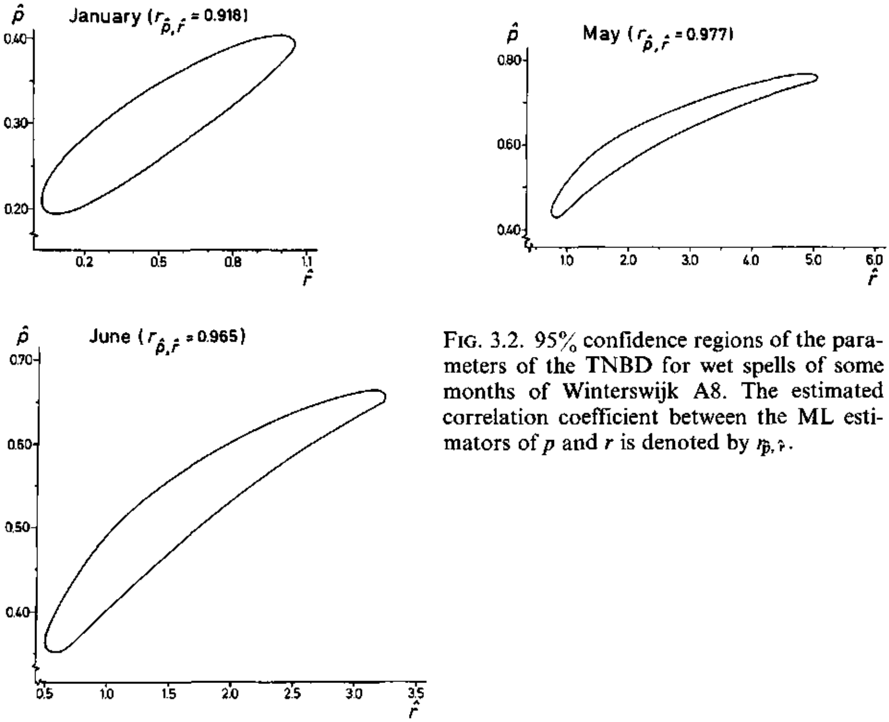


FIG. 3.2. 95% confidence regions of the parameters of the TNBD for wet spells of some months of Winterswijk A8. The estimated correlation coefficient between the ML estimators of p and r is denoted by $r_{\hat p,\hat r}$.

the log likelihood differs less than 3.0 from its maximum. The value 3.0 ($\approx \frac{1}{2}\chi_2^2$ (0.95)) is based on the asymptotic distribution of the LR statistic. This asymptotic approximation can be doubtful, especially in the months of May and June. From Figure 3.2 it is seen that for a particular month the range of acceptable p and r values is quite large.

To reduce the number of parameters, the parameter r is assumed to be constant throughout the year. For r the average $\bar r$ of the twelve monthly ML estimates can be taken. The parameter p can then be estimated by the ML method for each month separately. The likelihood equation for this case reduces to the expression for the mean in Table 3.3 with μ_1' replaced by the average interval length. This equation can be solved iteratively with the Newton-Raphson method. An initial estimate of p can be based on the identity:

$$(3.10) \qquad p^{r+1} = P(\underset{\sim}{y} = 1)/\mu_1'$$

which follows from (3.3) and Table 3.3. Estimates of $P(\underset{\sim}{y} = 1)$ and μ_1' can be based on the previously obtained ML estimates of p and r. The estimates will be denoted as approximate maximum likelihood (AML) estimates since the likelihood is not fully maximized over the parameter space.

Better estimates for the parameters can be obtained by maximizing the likelihood with respect to all parameters (ML estimates). If p_m is the parameter p in the mth month ($m = 1$ corresponds to January) and $L(r, p_1, \ldots, p_{12})$ is the log likelihood in some point $(r,p_1,\dots,p_{12})$ then:

$$(3.11) \qquad L_{max} = \max_{r,p_1,\dots,p_{12}} L(r,p_1,\dots,p_{12}) =$$

$$= \max_r(\max_{p_1,\dots,p_{12}} L(r,p_1,\dots,p_{12})) = \max_r L(r)$$

with $L(r) = \max_{p_1,\dots,p_{12}} L(r,p_1,\dots,p_{12})$.

For fixed r, the log likelihood with respect to $p_1,\dots,p_{12}$ can be maximized for each month separately as described above for the AML estimates. Values for $L_{max}-L(r)$ are given in Figure 3.3 for Winterswijk A8 and Hoofddorp A8. The maximization of $L(r)$ with respect to r can be done graphically (using Figure 3.3) or with a numerical maximization procedure.

A confidence interval for r can be based on the asymptotic behaviour of the logarithm of the LR. Note from Figure 3.3 that $\bar{r}$ always lies within the 95% confidence region. On the other hand the values $r = 0$ (LSD) and $r = 1$ (GD) always lie outside this confidence region, which supports the use of the TNBD.

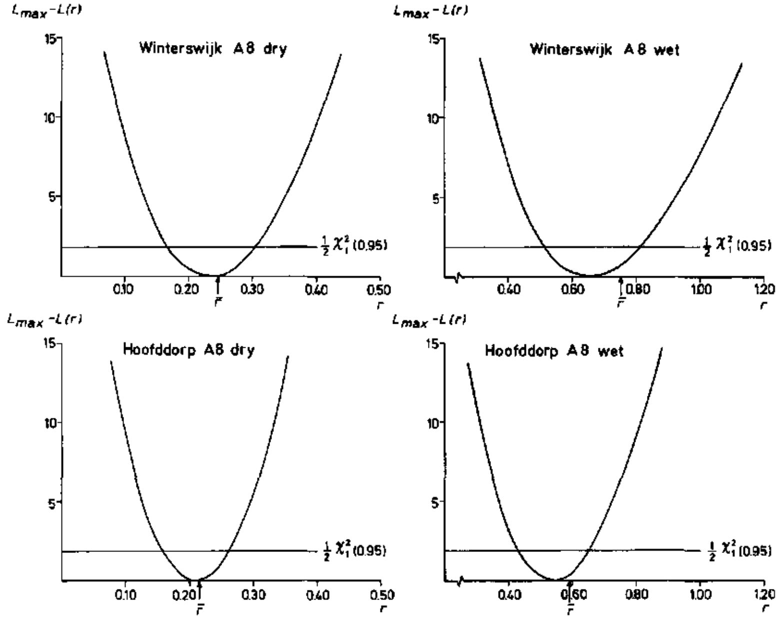


FIG. 3.3. $L_{max} - L(r)$ (for definition see Equation (3.11)) versus r for weather spells of Hoofddorp A8 and Winterswijk A8. The monthly mean of the ML estimates of r is denoted by $\bar{r}$.

This fact also follows from Figure 3.1 where nearly all values of $\hat{r}$ lie between 0 and 1.

The assumption of a constant r can be tested with the LR test. Then the asymptotic distribution of $\underline{l}$ is χ^2_{11}. The LR test is progressive if AML estimates are used because the numerator of (3.8) is too small in such cases. Realizations of the LR statistic (l_2) and their corresponding critical levels (C.L.) are given in Table 3.5. The hypothesis, r is constant during the year, is not rejected at the 5 per cent level for the Winterswijk series; for the Hoofddorp series, however, this hypothesis is nearly always rejected.

The AML estimates for the Winterswijk and Hoofddorp series (method A) are given in Figure 3.4. This figure shows that wet spells always have high values for $\hat{p}$ during the summer months (a high value for $\hat{p}$ means a small average interval length). For dry spells there are peaks during summer and

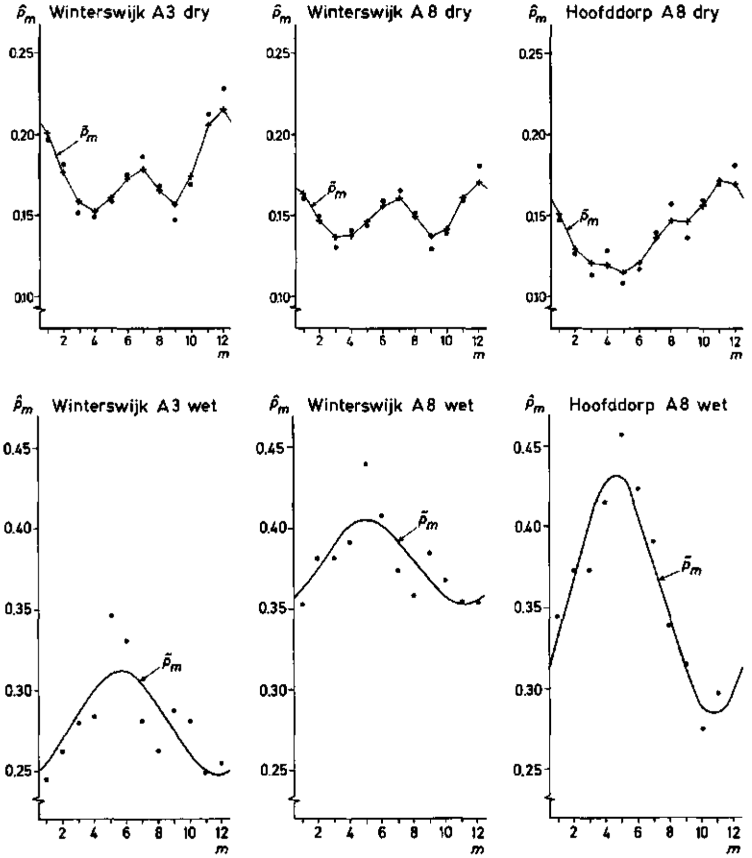


FIG. 3.4. Rough and smoothed AML estimates of the parameter p of the TNBD of Winterswijk (A3 and A8) and Hoofddorp A8. The rough estimates are denoted by $\hat{p}_m$ and the smoothed estimates by $\tilde{p}_m$.

winter for the Winterswijk series; for the Hoofddorp series there is only a clear peak during winter.

One can try to approximate the annual course of the parameter p with a Fourier series:

$$(3.12)\qquad p_m = A_0 + \sum_{k=1}^{n_0} [A_k \cos(km^*) + B_k \sin(km^*)] \qquad n_0 \leqslant 5$$

with n_0 is the number of harmonics,
$m^* = 30(m-\frac{1}{2})$ degrees $= \pi(m-\frac{1}{2})/6$ radians.
So January corresponds to 15 degrees, February to 45 degrees, and so on.

On the basis of estimates of p, one can get OLS estimates for fixed n_0 of the Fourier coefficients (cf. VAN MONTFORT (1966) and JENKINS and WATTS (1969)):

$$(3.13a)\qquad \hat{A}_0 = \frac{1}{12} \sum_{m=1}^{12} \hat{p}_m$$

$$(3.13b)\qquad \hat{A}_k = \frac{1}{6} \sum_{m=1}^{12} \hat{p}_m \cos(km^*) \qquad k = 1, \ldots, n_0$$

$$(3.13c)\qquad \hat{B}_k = \frac{1}{6} \sum_{m=1}^{12} \hat{p}_m \sin(km^*) \qquad k = 1, \ldots, n_0.$$

Equation (3.12) can also be written as

$$(3.14)\qquad p_m = R_0 + \sum_{k=1}^{n_0} R_k \cos(km^* + \phi_k)$$

where the amplitudes and phase angles are:

$$(3.15a)\qquad R_0 = A_0$$

$$(3.15b)\qquad R_k = \sqrt{A_k^2 + B_k^2} \qquad k = 1, \ldots, n_0$$

$$(3.15c)\qquad \phi_k = \arg(A_k - iB_k) \text{ radians} \qquad k = 1, \ldots, n_0 .$$

Estimates of R_k and ϕ_k can be obtained by substituting the OLS estimates of the Fourier coefficients in the right side of the above equations.

A problem still is a suitable choice of n_0. In the first instance n_0 is chosen so large that no harmonics are expected after the n_0th. Then the significance of the n_0th harmonic is tested using the statistic:

$$(3.16)\qquad T_{n_0} = 3R_{n_0}^2/\hat{\sigma}^2(n_0)$$

where $\hat{\sigma}^2(n_0)$ is the estimate of the variance of the $\underline{\hat{p}}_m$s. Under the assumption of independent and normally distributed error terms, the null distribution of the statistic $\underline{T}_{n_0}$is $\underline{F}(2,11-2n_0)$. If the n_0th harmonic is not significant, then the

TABLE 3.6. Harmonic analysis of the AML estimate of the parameter p of the TNBD for wet intervals of the Winterswijk and Hoofddorp series (method A).

	k	$\hat{R}_k$	$\hat{\phi}$ (degrees)	T_k	C.L.
Winterswijk A3	1	0.032	200	6.45	0.018
	2	0.017	98	2.41	0.160
	3	0.017	268	4.63	0.073
Winterswijk A8	1	0.027	220	6.66	0.017
	2	0.012	116	1.53	0.281
	3	0.015	278	4.63	0.073
Hoofddorp A8	1	0.074	232	61.88	0.000
	2	0.013	27	2.30	0.171
	3	0.010	298	1.64	0.284

value of n_0 is lowered by one and the test is repeated. This procedure goes on until a significant harmonic is found. The successive tests are independent at the null hypothesis (cf. HOGG (1961)).

Table 3.6 gives the estimated amplitudes, phase angles and results of the test for wet spells of the series of Winterswijk and Hoofddorp, analysed by method A. Starting with n_0 is 3 and testing at the 5 per cent level, one finds the smallest acceptable value for n_0 to be 1. The fitted harmonic series are given in Figure 3.4. For dry spells a harmonic analysis of the $\hat{p}_m$s is less successful as could be expected from Figure 3.4. At the 5 per cent level one finds 2 significant harmonics for Winterswijk A8 and 3 significant harmonics for Winterswijk A3 and Hoofddorp A8. Therefore the $\hat{p}_m$s were smoothed according to the moving average:

$$(3.17) \qquad \tilde{p}_m = \tfrac{1}{4}\hat{p}_{m-1} + \tfrac{1}{2}\hat{p}_m + \tfrac{1}{4}\hat{p}_{m+1} \qquad m = 1, \ldots, 12$$

with $\hat{p}_0 = \hat{p}_{12}$ and $\hat{p}_{13} = \hat{p}_1$.

The smoothed series are also given in Figure 3.4.

For wet spells a LR test was done to test the hypothesis that the parameter r is the same for all months and the p_ms can be approximated by a Fourier series with 1 harmonic component. Here the LR test is progressive because the likelihood is not fully maximized under the null hypothesis. The realizations of the test ststistic (l_3) and their critical levels (C.L.) are given in Table 3.5. Though l_3 is larger than l_2, as a result of the null hypothesis being more restrictive, the critical level corresponding to l_3 can be larger, because there are more degrees of freedom. The difference between l_3 and l_2 nearly always turns out to be less than 16.9 ($= \chi^2_9$ (0.95)) which confirms more or less the results of the F-test of the harmonic analysis (for AML estimates this conclusion is a bit dangerous because both the numerator and the denominator in (3.8) are too small). Table 3.5 also shows that the differences between realizations of LR statistics of AML and ML estimates are small compared with the magnitude of l_2 (influence of assuming r constant) or the difference between corresponding l_2 and l_3 values (influence of smoothing). Therefore only the simple AML estimates will be considered further.

So with respect to seasonal variation of the parameters of the TNBD it seems reasonable to take a constant value for the parameter r throughout the year. An estimate of this parameter can be the mean of the twelve monthly ML estimates. At a fixed value of r, monthly estimates of the parameter p can be obtained by the ML method. These estimates can be smoothed by a Fourier series with 1 harmonic component for the length of wet spells and by a moving average scheme for the length of dry spells.

4. The dependence of the distribution of rainfall amounts on the process for wet and dry days

In this section questions like 'Does the rainfall amount of a certain wet spell depend on the length of the preceding dry spell?' or 'Does the intensity depend on the length of wet spells?' are investigated.

Though the models by Quélennec (1973) and Dumont and Boyce (1974) do not assume any relation between rainfall depth and day of occurrence, the generation scheme by Cole and Sherriff (1972) distinguishes three different types of rainfall amounts, namely rainfall amounts on solitary wet days, rainfall amounts on the first day of other wet spells, and rainfall amounts on days preceded by a wet day. Smith and Schreiber (1974) compared empirical distribution functions of wet days preceded by wet days and preceded by dry days and found a small, but significant difference. They also found by the Wilcoxon test a significant difference between the rainfall intensities of 1-day and 2-day spells.

In Section 4.1 the correlation between the length of a dry period and the rainfall amount on the day following that period is discussed. Section 4.2 treats the problem of homogeneity of rainfall intensities of various wet spells.

4.1. *The correlation between the rainfall amount on the first day of a wet spell and the length of its preceding dry spell*

This section deals with whether the rainfall amount on the first day of a particular wet spell depends on the length of the previous dry spell. Correlation coefficients were estimated for each season separately; for that purpose the rainfall amount on the first day of a wet spell and the dry interval preceding

TABLE 4.1. Estimated correlation coefficients between the rainfall amount on the first day of a wet spell and the length of its preceding dry spell.

	Winterswijk		Hoofddorp	
Period	A3	A8	A8	B8
Winter	–0.038	–0.035	–0.085	–0.080
Spring	–0.016	–0.031	–0.028	–0.076
Summer	0.004	–0.019	0.000	–0.015
Autumn	–0.057	–0.006	–0.024	0.018

that wet spell were assigned to the period to which the dry interval belongs. Estimated correlation coefficients for Winterswijk (A3 and A8) and Hoofddorp are given in Table 4.1. From this table it is seen that the correlation coefficients are small and predominantly negative. A rough test based on the approximation of the standard deviations of the correlation coefficients (for these standard deviations the same holds as for the standard deviations of the correlation coefficients of successive wet and dry spells in Section 3.1) shows no evidence for a real correlation.

The same conclusion was reached by SMITH and SCHREIBER (1974) who found a correlation coefficient of -0.0132 based on 1172 pairs of observations.

4.2. *The relation between rainfall intensity and length of wet spells*

A possible dependence between intensities and lengths of wet spells is investigated with the regression model:

$$(4.1) \qquad \underline{y}_{ij} = \mu_i + i^{-1/2}\, \underline{e}_{ij}$$

with y_{ij}: intensity of the jth spell with length i. The length will be expressed in days and the intensity in mm/day,
μ_i: mean intensity of spells with length i.

It will be assumed that the error terms $\underline{e}_{ij}$ are independent variates with mean zero and standard deviation σ. Therefore in model (4.1) the variance of the intensity of a particular spell is assumed to be inversely proportional to its length. This assumption is true when the rainfall amounts within a wet spell are uncorrelated; when there is a positive correlation, the variance of the intensity is larger than σ^2/i if $i > 1$. In Section 5.1 it will be seen that successive rainfall amounts are nearly uncorrelated and thus the choice of the factor $\sqrt{i}$ in (4.1) looks reasonable.

For identically distributed rainfall amounts the μ_is should be mutually equal. This hypothesis can be tested with an F-test provided the error terms are independently and normally distributed. Especially the normality condition is not fulfilled here as daily rainfall amounts have very skew distributions. The monthly mean of the coefficient of skewness of rainfall amounts on wet days is 2.69 for Winterswijk A3 and 2.41 for Winterswijk A8. To examine the influence of non-normality on the F-test, the test was repeated in some cases with a normalizing transformation on the data. As a normalizing transformation the cube root was taken, which reduces the monthly mean of the coefficient of skewness to 0.65 for Winterswijk A3 and to 0.81 for Winterswijk A8.

The critical levels of the F-test are given in Table 4.2 for Winterswijk (A3 and A8) and Hoofddorp. At the 5 per cent level the hypothesis of constant intensity is rejected for all seasons. During the summer months, however, there is less evidence for different intensities at different lengths of spells. Both transformed and untransformed data lead to the same conclusions.

In Figure 4.1 the mean intensities at different lengths of wet spells are given for Winterswijk A8. From this figure it is seen that the intensity of rainfall amounts is smaller during short spells, especially in autumn and winter.

TABLE 4.2. Critical levels of the *F*-test for equality of mean intensities at different lengths of wet spells.

	Winterswijk		Winterswijk (cube root)		Hoofddorp	
Period	A3	A8	A3	A8	A8	B8
January	0.000	0.000	0.000	0.000	0.212	0.025
February	0.000	0.000	0.000	0.000	0.006	0.001
March	0.000	0.003	0.000	0.001	0.002	0.006
April	0.000	0.108	0.000	0.015	0.086	0.202
May	0.442	0.504	0.467	0.712	0.048	0.057
June	0.085	0.140	0.041	0.100	0.677	0.807
July	0.878	0.849	0.793	0.871	0.888	0.925
August	0.043	0.249	0.027	0.085	0.000	0.000
September	0.002	0.088	0.000	0.066	0.000	0.000
October	0.001	0.006	0.000	0.004	0.000	0.000
November	0.000	0.000	0.000	0.000	0.000	0.000
December	0.000	0.000	0.000	0.000	0.000	0.000
Winter	0.000	0.000	0.000	0.000	0.000	0.000
Spring	0.000	0.007	0.000	0.000	0.004	0.031
Summer	0.023	0.154	0.001	0.054	0.001	0.001
Autumn	0.000	0.000	0.000	0.000	0.000	0.000

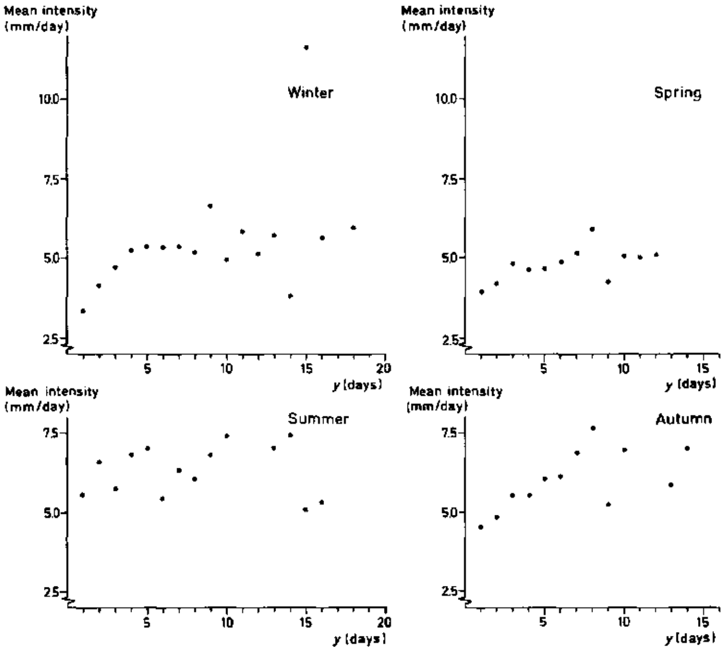


FIG. 4.1. Mean intensity versus length of wet spells (*y*) for Winterswijk A8.

TABLE 4.3. Critical levels corresponding to realizations of different test statistics for testing equality of means of type 0, 1 and 2 amounts. The critical level of Snedecor's F-test with 2 degrees of freedom in the numerator is denoted by C_F and the one-sided critical level of Student's t-test for differences between type i and j amounts is denoted by C_{ij}. A blank indicates less than 0.0005.

	Winterswijk A3				Winterswijk A3 (cube root)				Winterswijk A8				Hoofddorp A8			
Period	C_F	C_{10}	C_{20}	C_{21}	C_F	C_{10}	C_{20}	C_{21}	C_F	C_{10}	C_{20}	C_{21}	C	C_{10}	C_{20}	C_{21}
January		0.001								0.001		0.002	0.004	0.025	0.001	0.034
February		0.073				0.022				0.014				0.011		0.002
March		0.006				0.013						0.027	0.026	0.095	0.004	0.036
April		0.003		0.007		0.003			0.017	0.063	0.003	0.044	0.071	0.266	0.019	0.032
May	0.407	0.459	0.188	0.098	0.115	0.083	0.023	0.141	0.712	0.600	0.336	0.196		0.190		
June	0.009	0.036	0.001	0.043		0.008		0.006	0.241	0.108	0.045	0.245	0.508	0.412	0.138	0.163
July	0.707	0.299	0.203	0.347	0.073	0.143	0.014	0.069	0.676	0.210	0.350	0.714	0.424	0.227	0.098	0.224
August	0.003	0.007		0.086		0.002		0.027	0.122	0.056	0.020	0.259				
September	0.001	0.007		0.035		0.004		0.002	0.085	0.290	0.026	0.034		0.002		
October		0.058				0.037				0.079		0.001				
November		0.159				0.027				0.076				0.008		
December		0.002								0.001		0.002				
Winter																
Spring		0.003								0.011		0.007		0.077		
Summer		0.005		0.023					0.056	0.021	0.010	0.324		0.002		
Autumn		0.002								0.026						

A model which may explain this phenomenon is one which distinguishes three different types of wet days, namely solitary wet days, wet days at one side bounded by a wet day and at the other side bounded by a dry day, and wet days at each side bounded by a wet day. The rainfall amounts on these different wet days are denoted by type 0, 1 and 2 amounts, respectively; so a type i amount denotes a rainfall amount on a wet day with i adjacent wet days. These three types of rainfall amounts are discriminated in this way as the first and last day of a wet interval contain a part of the preceding or following dry period because of the day unit of measure. The method given here differs from the method by COLE and SHERRIFF (1972), who allocated rainfall amounts at the end of a spell longer than one day to amounts called type 2 here.

To show that there are differences between type 0, 1 and 2 amounts, tests for equality of means were done, namely an F-test (for all three types of rainfall amounts simultaneously) and a two-sample Student test (for all pairs of different types of rainfall amounts). As for the test for homogeneity of intensities the normality condition is not fulfilled here. Therefore the test was also sometimes done for the cube root of rainfall amounts. Critical levels of the test statistics are given in Table 4.3 for Winterswijk (A3 and A8) and Hoofddorp. The critical levels of the Student tests are based on one-sided tests because it is assumed on intuitive grounds that the expected rainfall amount on a particular wet day increases with the number of adjacent wet days. From Table 4.3 it is seen that the hypothesis of equal means is rejected at the 5 per cent level in nearly all cases, so there is evidence for different distributions of the three types of rainfall amounts. Tests based on transformed rainfall amounts

TABLE 4.4. Realizations of Student's t-statistic for testing a difference in mean between rainfall amounts on the first and last day of a wet spell, only for spells with a length of at least three days.

Period	Winterswijk A3	Winterswijk A8	Hoofddorp A8	Hoofddorp B8
January	-0.62	-0.34	0.20	0.66
February	0.71	-1.72	-0.09	0.00
March	1.42	0.20	0.21	-0.09
April	1.58	1.56	0.49	0.20
May	0.77	0.91	0.30	0.56
June	1.12	0.26	1.13	0.65
July	-0.21	0.21	0.63	0.45
August	2.19	1.95	1.00	1.09
September	1.31	1.81	0.54	1.13
October	-1.65	-0.23	0.42	0.17
November	1.69	1.68	1.47	1.49
December	1.98	1.80	-0.42	-0.56
Winter	1.37	0.31	-0.20	0.05
Spring	2.14	1.59	0.58	0.36
Summer	1.83	1.44	1.54	1.29
Autumn	0.66	1.75	1.34	1.54

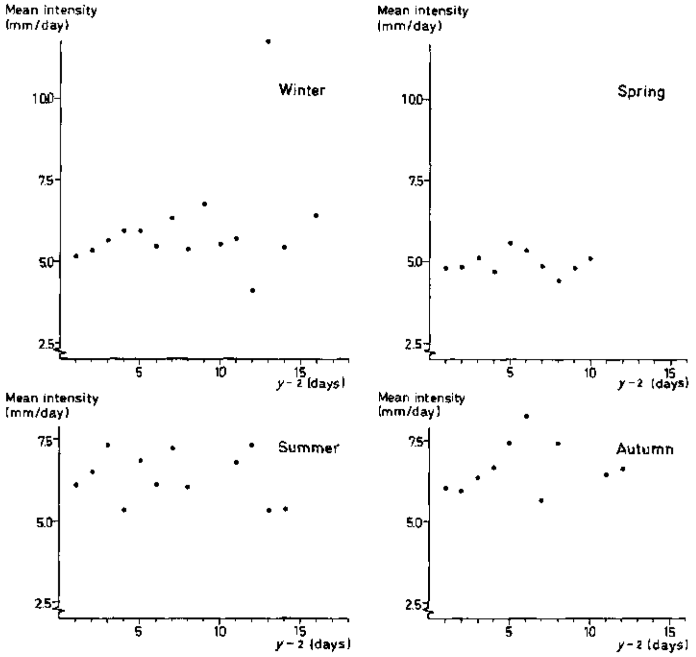


FIG. 4.2. Mean intensity versus length of spells of consecutive type 2 amounts (y-2) for Winterswijk A8.

give similar results to those based on the original data though the critical levels are in most cases somewhat smaller.

To investigate any difference between the beginning and the end of a wet spell, a two-sample Student test was done for a difference in mean between rainfall amounts on the first day of a wet spell and rainfall amounts on the last day of a wet spell, only for spells with a length of more than two days. Realizations of the test statistic are given in Table 4.4 for Winterswijk (A3 and A8) and Hoofddorp. The number of paired observations in a particular month is about 110 for Winterswijk A8, 130 for Winterswijk A3 and 160 for Hoofddorp, so it can be concluded from the tabulated realizations of the test statistic that there is no evidence for differences (the critical values at the 5 per cent level are –2.0 and 2.0). The fact that most realizations are positive is a slight indication for a larger mean at the beginning of a wet interval.

In Figure 4.2 mean intensities of spells containing only type 2 amounts are given for Winterswijk A8. At first glance there is no relation between the mean intensity and the length of spells.

To test this assumption the regression model of Equation (4.1) was repeated for wet intervals without the first and last wet day. Critical levels of the F-test for equality of mean intensities are given in Table 4.5. The F-test is less powerful here because the number of data are decreased considerably. (The fraction of type 2 amounts is about 0.45 if $\delta = 0.3$ mm and about 0.35 if $\delta = 0.8$ mm.)

TABLE 4.5. Critical levels of the *F*-test for equality of mean intensities of type 2 amounts at different lengths of wet spells, omitting the first and last day.

Period	Winterswijk		Winterswijk (cube root)		Hoofddorp	
	A3	A8	A3	A8	A8	B8
January	0.111	0.013	0.367	0.717	0.541	0.214
February	0.172	0.661	0.251	0.776	0.839	0.710
March	0.264	0.994	0.339	0.976	0.328	0.664
April	0.161	0.948	0.028	0.853	0.082	0.847
May	0.198	0.462	0.296	0.883	0.181	0.075
June	0.394	0.423	0.500	0.630	0.451	0.672
July	0.965	0.908	0.992	0.965	0.831	0.851
August	0.440	0.790	0.711	0.681	0.207	0.026
September	0.167	0.865	0.108	0.816	0.737	0.756
October	0.784	0.501	0.673	0.371	0.875	0.333
November	0.268	0.352	0.196	0.399	0.082	0.032
December	0.252	0.784	0.055	0.528	0.280	0.071
Winter	0.034	0.054	0.010	0.724	0.726	0.283
Spring	0.280	0.987	0.173	0.974	0.486	0.881
Summer	0.390	0.621	0.553	0.758	0.123	0.044
Autumn	0.621	0.261	0.228	0.161	0.417	0.497

Mostly the hypothesis of no differences is not rejected at the 5 per cent level. The small critical level for the winter data of the Winterswijk series is caused by a single rainy period of long duration with very high intensity (see Figures 4.1 and 4.2). The large standard deviation of the February monthly total (see II, 3) is also due to this rainy period.

5. THE DISTRIBUTION OF RAINFALL AMOUNTS ON WET DAYS

There are several methods in literature for modelling the behaviour of rainfall amounts on wet days. Many authors (cf. WOOLHISER et al. (1972), QUÉLENNEC (1973), DUMONT and BOYCE (1974) and SMITH and SCHREIBER (1974)) fit some theoretical distribution to these amounts and assume independence. On the other hand, COLE and SHERRIFF (1972) sampled from empirical distribution functions and used a first order Markov chain to describe the dependence between successive rainfall amounts. Quite another method, which is often used for generating rainfall amounts within time-increments shorter than one day, consists of generating a total rainfall amount for a wet spell and then splitting up this rainfall amount (cf. GRACE and EAGLESON (1966), HIEMSTRA and CREESE (1970) and LECLERC and SCHAAKE (1973)). The method of subdivision differs considerably from author to author.

In Section 5.1 the correlation between successive rainfall amounts is investigated. The marginal distribution of non-zero rainfall amounts is discussed in Section 5.2.

TABLE 5.1. Estimated first sccs of rainfall amounts within a wet spell.

Month	Winterswijk A3	Winterswijk A8	Winterswijk A8 (weighted average)	Hoofddorp A8	Hoofddorp B8
January	0.161	0.124	0.092	0.131	0.091
February	0.156	0.112	0.098	0.144	0.146
March	0.121	0.097	0.100	0.078	0.099
April	−0.010	0.000	−0.001	0.066	0.019
May	0.076	0.102	0.121	0.029	0.034
June	−0.007	−0.008	−0.005	0.059	0.065
July	0.025	0.014	0.012	0.077	0.098
August	0.055	0.058	0.053	0.064	0.082
September	0.089	0.082	0.049	0.179	0.137
October	0.141	0.119	0.112	0.137	0.155
November	0.166	0.164	0.166	0.096	0.090
December	0.176	0.148	0.130	0.027	0.049

5.1. *Correlation of rainfall amounts within a wet spell*

In this chapter no correlation between rainfall amounts of different wet spells is assumed. Therefore only pairs of observations in the same wet period are used here for the estimation of correlation coefficients.

The most simple way to estimate the first serial correlation coefficient (scc) for rainfall amounts within a wet spell is to take all pairs of successive observations, ignoring the fact that rainfall amounts at the beginning and the end of a wet spell have different distributions. Another method which takes into consideration the fact that not all rainfall amounts have the same distribution is to take a weighted average of four estimated correlation coefficients, namely between successive type 1 amounts (this case occurs only in 2-day intervals), between successive type 2 amounts, between a type 1 amount and its successive type 2 amount, and a type 2 amount and its successive type 1 amount. The weights can be chosen inversely proportional to the number of observations on which the estimated correlation coefficients are based.

Estimated lag one sccs are given in Table 5.1 for Winterswijk (A3 and A8) and Hoofddorp when there is no discrimination between different types of rainfall amounts. The estimated first scc obtained as a weighted average is given in this table for Winterswijk A8 only.

Under the assumption of independence the standard deviation of the correlation estimator is about $1/\sqrt{N}$, where N is the number of paired observations.

With this approximation, the standard deviation of the correlation estimate is about 0.040 for Winterswijk A3 and Hoofddorp and about 0.045 for Winterswijk A8. So from Table 5.1 it can be concluded that a small but significant correlation exists in the winter months. Seasonal dependence is less obvious for the Hoofddorp series than for the Winterswijk series. The estimated correlation coefficients are mostly larger at a smaller threshold. Finally Table 5.1 shows that there is only a slight difference between the results of the two estimation techniques.

In general sccs with a lag greater than 1 do not differ significantly from zero. For instance the estimates of the lag 2 scc of Winterswijk A3 are -0.018, -0.056, 0.001 and 0.048 for winter, spring, summer and autumn, respectively. These correlation estimates are based on type 2 amounts only, which gives about 600 pairs of observations for each season.

5.2 *The marginal distribution of rainfall amounts*

The frequency distribution of rainfall amounts on wet days is usually J-shaped. Many common distribution functions with this form are usually defined on the interval $[0, \infty)$, but when dealing with rainfall amounts on wet days, where a wet day is defined as a day with a rainfall amount above some threshold, the lower limit of the carrier is not zero. If the threshold is δ mm then the lower limit of the carrier is $\tilde{\delta} = \delta - 0.05$ when rainfall is recorded in tenths of millimeters. For rainfall amounts on wet days one has to work with truncated or shifted distributions. Though for physical reasons a truncated distribution might be better, shifted distributions are often preferable because of their mathematical convenience. Rainfall amounts will be called shifted rainfall amounts when they are reduced by the value $\tilde{\delta}$. For these shifted rainfall amounts the monthly mean of the ratio $\hat{\gamma}/\hat{C}$ (γ is the coefficient of skewness, defined by II, (2.1) and C is the coefficient of variation, defined by II, (3.10e)) is 2.19 for Winterswijk A3, 2.23 for Winterswijk A8 and 2.13 for Hoofddorp A8. These values are close to the theoretical value 2 of the gamma distribution, which was defined in II, 3.2. The shifted gamma distribution (SGD) might therefore be a suitable choice for fitting the frequency distributions of rainfall amounts on wet days.

A convenient property of the SGD is that the sum of, say n, iid variables is again a shifted gamma variable with an unchanged scale parameter and with both the shape parameter and the shift multiplied by n. Notice that the truncated gamma distribution does not have this property. The fact that the distribution of sums of independent variables is not complicated facilitates the numerical computation of cumulative distribution functions of k-day totals ($k = 2,3,\ldots$).

Since in most cases the frequency distribution of (shifted) rainfall amounts is J-shaped, it could be expected that the shape parameter of the fitted SGD is less than 1. Therefore ML estimates were used, because the method of moments gives very inefficient estimates, when the shape parameter is small (see II, Table 3.2). The likelihood equations, given in Chapter II (Equation (3.13)), have to be applied to shifted rainfall amounts. A direct application of these equations, however, is not advisable, because the solution is very sensitive to the accuracy of small observations. Therefore a modification of the ML method was used, which ignores the actual values of shifted rainfall amounts smaller than some value ε and merely uses the number of these observations.

From the density of the gamma distribution II, (3.7) and on the basis of independent data one gets the likelihood, $L^*(\lambda,\nu)$:

$$(5.1) \qquad L^*(\lambda,\nu) = \frac{\lambda^{n\nu}}{(\Gamma(\nu))^n}\left\{\int_0^{\varepsilon}\exp(-\lambda x)x^{\nu-1}dx\right\}^n \frac{\lambda^{m\nu}}{(\Gamma(\nu))^m}\times$$

$$\times\exp(-\lambda\sum_{i=1}^{m}x_i)\prod_{i=1}^{m}x_i^{\nu-1}$$

with $x_1, \ldots, x_m$: shifted rainfall amounts larger than ε,
n: number of shifted rainfall amounts on $(0, \varepsilon)$
(cf. DAS (1955)). So rainfall amounts on $[\delta, \delta+\varepsilon)$ are only counted.

Now for small ε the following approximately holds:

$$(5.2) \qquad \int_0^{\varepsilon}\exp(-\lambda x)x^{\nu-1}dx \approx \int_0^{\varepsilon}(1-\lambda x)x^{\nu-1}dx = \frac{\varepsilon^{\nu}}{\nu}\left(1-\frac{\lambda\nu\varepsilon}{\nu+1}\right).$$

DAS (1955) only used the first term (ε^{ν}/ν), but the approximation with two terms has the advantage that larger values for ε are admissible. For Dutch rainfall series ε was taken equal to 0.5 mm. With this value for ε, the relative error in the integral is about 0.1 per cent if the approximation (5.2) is used and about 3 per cent for the approximation with only one term.

The solution of the likelihood equations is given in Appendix A1.

Monthly ML estimates of the shape parameter and the mean (ν/λ) are given in Figures 5.1 and 5.2 for type 0, 1 and 2 amounts of Winterswijk (A3 and A8) and Hoofddorp A8. The ML estimate of the mean is given instead of the ML estimate of the scale parameter, because the ML estimates of the mean and of the shape parameter are nearly uncorrelated. For the usual ML estimation, discussed in II, 3.2, the ML estimates are even independent (cf. 6.2 of COX and LEWIS (1966) and SHENTON and BOWMAN (1970)).

Figures 5.1 and 5.2 also give monthly means of the standard deviations of

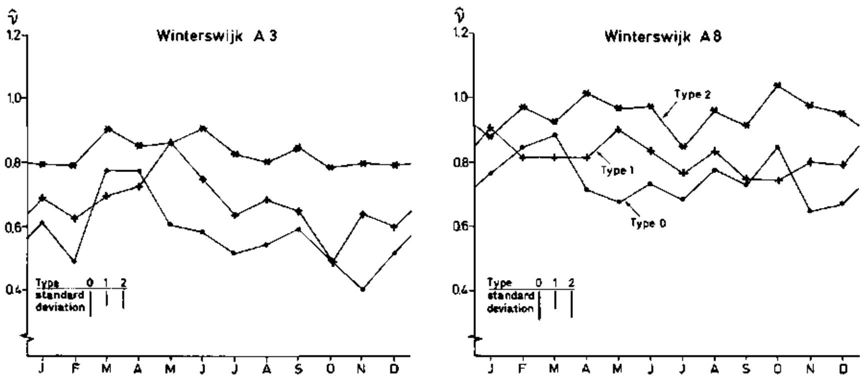


FIG. 5.1. ML estimates of the parameter ν of the SGD fitted to rainfall amounts of different types. Monthly means of the standard deviation of the ML estimators are denoted by vertical limes.

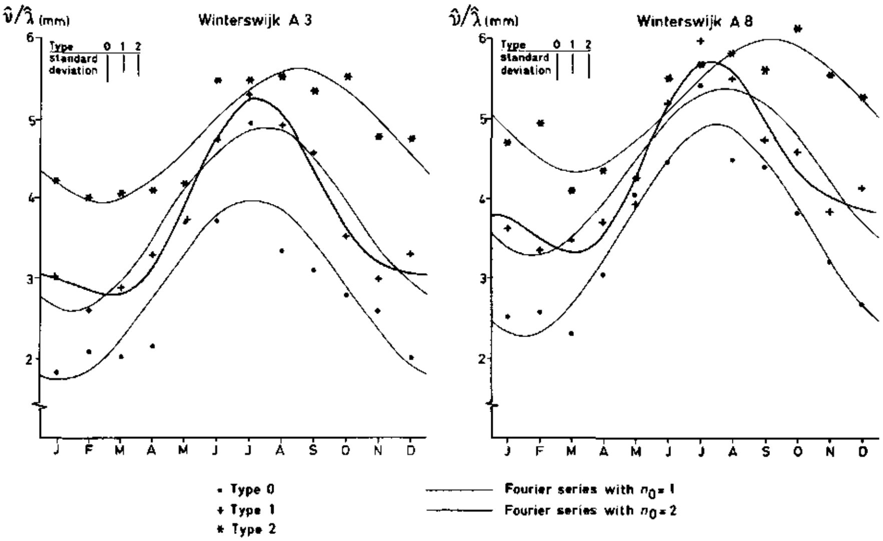


FIG. 5.2. ML estimates of v/λ (shifted mean) of the SGD fitted to rainfall amounts of different types and their Fourier series approximations. Monthly means of the standard deviations of the ML estimators are denoted by vertical lines.

the ML estimators. The standard deviation of $\hat{\underline{v}}$ was obtained from the matrix of second derivatives of the log likelihood (cf. KENDALL and STUART (1973), 18.26). Though the standard deviation of the ML estimator of the mean can also be obtained from second derivatives of the log likelihood, it is easier to use the fact that the non-modified ML estimate of the mean is the sample mean because of II, (13b). Thus the standard deviation of $\hat{\underline{v}}/\hat{\underline{\lambda}}$ can be based on the ML estimate of the variance (v/λ^2). The given standard deviations can be a bit too small due to a slightly positive serial correlation in the data (see 5.1). Figure 5.1 shows that the shape parameter of the fitted SGD is in general larger and as a result the distribution is less skew, if the number of adjacent wet days is larger. A lower value of the threshold δ gives lower values for the shape parameter. Figure 5.2 shows that differences in the mean of type 0,1 and 2 amounts are more prominent during winter. Further a seasonal change in the mean is obvious.

The goodness of fit of the SGD was tested with the X^2-test.

For the application of this test frequency distributions were constructed for the rainfall amounts with class boundaries at 0.75 (0.50) 5.25 (1.00) 10.25 (2.50) 25.25 mm in the case that $\delta = 0.3$ mm; if the threshold was at 0.8 mm the boundary 0.75 mm was omitted. Notice that the data belonging to the first class are just the data whose numerical value was ignored in the application of the modified ML procedure.

Expected cell frequencies were based on ML estimates which were obtained from the maximization of (5.1) with respect to v and λ. Because actual values were used instead of cell frequencies the X^2-test is a bit progressive here (see II, 3.2).

For the computation of the expected cell frequencies use was made of a series expansion

(ABRAMOWITZ and STEGUN (1970), Equation 6.5.29, non-alternating version) of the incomplete gamma function. If the expected number of rainfall amounts in some class was less than 5, classes were joined together in the same way as was done for frequency distributions of lengths of weather spells (see 3.2).

For Winterswijk A3 and A8 critical levels of the X^2-test are given in Table 5.2 when there is no distinction between different types of rainfall amounts and for type 0, 1 and 2 amounts separately. From the tabulated critical levels it can be concluded that the SGD fits the data well. Other rainfall series give similar results.

Special cases of the SGD show lack of fit. For instance, application of the shifted exponential distribution (a SGD with $\nu = 1$, see II, 3.2) to all rainfall amounts irrespective of their type gives for Winterswijk A3 X^2-values which are all significant at the 5 per cent level. For Winterswijk A8 the fit is somewhat better, but still 8 out of 12 monthly critical levels are less than 0.05.

One can try to smooth the monthly estimates of the SGD parameters with a Fourier series. The results of the harmonic analysis are only given for the Winterswijk and Hoofddorp series, analysed by method A.

Using the procedure of Section 3.2 and making no discrimination between different types of rainfall amounts, one does not find any significant harmonic in the monthly ML estimator of ν (except for Winterswijk A3 where 2 significant harmonics are found). Two significant harmonics are found in the ML estimate of the mean, when testing at the 5 per cent level with starting value 3 for n_0.

If different types of rainfall amounts are distinguished there are usually no significant harmonics in $\hat{\nu}$ as was to be expected from Figure 5.1. Exceptions

TABLE 5.2. Critical levels of the X^2-test of goodness of fit for the SGD for rainfall amounts.

	Winterswijk A3				Winterswijk A8			
Month	all rainfall amounts	type 0 amounts	type 1 amounts	type 2 amounts	all rainfall amounts	type 0 amounts	type 1 amounts	type 2 amounts
January	0.845	0.664	0.532	0.977	0.418	0.213	0.868	0.702
February	0.950	0.950	0.052	0.931	0.631	0.317	0.072	0.576
March	0.853	0.682	0.648	0.874	0.901	0.499	0.698	0.947
April	0.474	0.498	0.025	0.926	0.221	0.122	0.012	0.983
May	0.088	0.304	0.351	0.788	0.071	0.341	0.444	0.595
June	0.969	0.553	0.313	0.680	0.915	0.446	0.453	0.329
July	0.104	0.314	0.073	0.945	0.036	0.145	0.024	0.574
August	0.211	0.262	0.816	0.309	0.097	0.350	0.604	0.730
September	0.439	0.859	0.766	0.770	0.559	0.709	0.727	0.562
October	0.088	0.614	0.491	0.589	0.194	0.449	0.413	0.820
November	0.630	0.220	0.622	0.746	0.390	0.202	0.170	0.778
December	0.703	0.162	0.085	0.827	0.344	0.504	0.457	0.380

are 2 significant harmonics for type 2 amounts of Hoofddorp A8 and 1 significant harmonic for type 1 and 2 amounts of Winterswijk A3.

The harmonic analysis of the mean of different types of rainfall amounts gives 1 significant harmonic for type 0 and 2 amounts and 2 significant harmonics for type 1 amounts. Fitted Fourier series are given in Figure 5.2. The use of the fitted Fourier series is a bit doubtful, because 1 harmonic component for all types of rainfall amounts can lead to an underestimation of the mean during most summer months (especially for Winterswijk A8). On the other hand the use of two harmonic components for type 1 amounts can give the unrealistic situation that the mean of type 1 amounts is larger than the mean of type 2 amounts.

6. Persistence

In the previous sections a stochastic model for daily rainfall sequences was developed. The model was characterized by a wet-dry process and a probability distribution for the rainfall amounts on wet days. In this section it is discussed whether such a model can describe the persistence of the daily rainfall process.

The persistence of daily rainfall sequences is extensively studied with correlograms and variance-time curves. Section 6.1 deals with the correlogram, which shows sccs as a function of the lag. The variance-time curve, which is introduced in Section 6.2, gives variances of rainfall amounts in a period with length t as a function of t. Because rainfall is recorded in 1-day intervals, correlograms and variance-time curves can only be evaluated at discrete time points.

6.1. *Analysis with correlograms*

Estimated sccs of both the wet-dry process and the entire rainfall process were obtained from II, (3.2) for each season separately. The rainfall series were analysed by wet-dry cycles (WD) or dry-wet cycles (DW) for reasons explained in Section 2. When the historic series is analysed in this manner the series cannot be stationary, since it always begins with a wet day after a dry spell (WD) or a dry day after a wet spell (DW). However because of the large number of observations in a season (about 6000 in the case of the Winterswijk series), the influence of this initial transient is negligible.

Estimated correlograms are given in Figures 6.1 and 6.2 for Winterswijk WD3. There are apparently no seasonal differences between the correlograms of the wet-dry process, but for the entire rainfall process there are obvious differences between the correlograms of the different seasons. The largest values of the estimated sccs are found during winter. Another feature of the estimated correlograms is their slow decay, which might be an indication for a non-stationary series (cf. Box and Jenkins (1970), 6.2.1). There can be departures from homogeneity or a period of a quarter may be too long to give a satisfactory reduction of seasonality. In II, A4 it was shown that jumps in

TABLE 6.2. Estimated and theoretical first sccs. The theoretical sccs are based on models with TNBDs for wet and dry intervals and SGDs for rainfall amounts on wet days.

		Wet-dry process		Rainfall process				
Station and	season	Estimated	Model	Estimated	Model I	Model II	Model III	Model IV
Winterswijk	Winter	0.345	0.341	0.270	0.090	0.177	0.178	0.274
($\delta = 0.3$ mm)	Spring	0.359	0.357	0.194	0.119	0.146	0.179	0.209
	Summer	0.334	0.334	0.140	0.099	0.110	0.135	0.146
	Autumn	0.360	0.358	0.236	0.100	0.159	0.175	0.238
Winterswijk	Winter	0.315	0.310	0.270	0.125	0.174	0.217	0.274
($\delta = 0.8$ mm)	Spring	0.309	0.308	0.193	0.135	0.155	0.180	0.201
	Summer	0.299	0.299	0.141	0.118	0.126	0.140	0.148
	Autumn	0.336	0.336	0.235	0.134	0.173	0.197	0.239
Hoofddorp	Winter	0.318	0.311	0.236	0.129	0.163	0.194	0.232
($\delta = 0.8$ mm)	Spring	0.292	0.288	0.187	0.130	0.146	0.172	0.188
	Summer	0.290	0.286	0.190	0.114	0.139	0.159	0.185
	Autumn	0.359	0.357	0.303	0.138	0.190	0.232	0.290

first sccs with calculated values for the wet-dry process and models I, II, III and IV with TNBDs for lengths of weather spells.

The estimated sccs are averages of the WD and DW analysis. The calculated sccs underly the assumption that for a particular season the rainfall process is stationary. Quarterly values for the parameters are obtained here by averaging monthly A and B estimates of the rainfall series. The parameter values of the TNBD are based on AML estimates.

For the wet-dry process the calculated sccs resemble the estimated ones, but for the entire rainfall process only the most sophisticated model (IV) gives reasonable sccs. From the tabulated values of Winterswijk with different δs it follows that the height of the threshold has no remarkable influence on the results. Another fact is that many different distributions for lengths of wet and dry spells can give approximately the same first scc both for the wet-dry process and the rainfall process. So for the winter season of Winterswijk ($\delta = 0.8$ mm) the first scc of a GD-LSD process (wet-dry process with a GD for wet intervals and a LSD for dry intervals) is 0.312, and the first scc of a GD-GD process (two-state first order Markov chain) is 0.311. It will be shown in the next section that these processes give rise to different sccs at higher lags.

Calculated sccs at higher lags, based on a TNBD-TNBD process, are given in Figure 6.1 for the wet-dry process and in Figure 6.2 for model IV of the entire rainfall process. For lags greater than 1 day calculated sccs are in general too small, especially during winter and autumn.

6.2. *Analysis with variance-time curves*

In Section 6.1 it was noticed that in general theoretical correlograms fall off much faster than estimated correlograms. However, the significance of these differences might be questionable, because it is well known that correlograms of different realizations of a certain stochastic process can vary widely (see, e.g. Figure 5.13 of JENKINS and WATTS (1969) and WALLIS and MATALAS (1971)). Comparisons between tails of estimated and theoretical correlograms are complicated, because estimates of the scc at neighbouring lags can be strongly correlated. In fact only for an uncorrelated stochastic process is the simultaneous distribution of estimated sccs at different lags comparatively simple, because these estimates are approximately uncorrelated (see II, 3.1).

If the correlogram dies out too rapidly, the model underestimates the variances of k-day totals for large k. However for large k the sample variation of estimates of variances can be quite large. For N independent observations the variance of the variance estimator $\underline{s}^2$ is approximately (cf. KENDALL and STUART (1969), 10.9):

$$(6.1) \qquad \operatorname{var} \underline{s}^2 \approx \frac{\mu_4 - \sigma^4}{N} = \frac{\sigma^4}{N}\left(\frac{\varkappa_4}{\sigma^4} + 2\right)$$

where $\varkappa_4$ denotes the fourth cumulant. For the normal distribution one gets:

$$(6.2a) \qquad \operatorname{var} \underline{s}^2 \approx 2\sigma^4/N$$

because $\varkappa_4 = 0$; and for the 'loi des fuites' (LDF), which was introduced in II, 3.2 and provided a good fit to monthly totals, one gets:

$$(6.2b) \qquad \operatorname{var} \underline{s}^2 \approx \frac{\sigma^4}{N}\left(\frac{6}{\theta} + 2\right)$$

making use of II, (3.17).

For 30-day totals it follows from the estimated values of θ in II, Table 3.3, that the standard deviation of $\underline{s}^2$ is at most $\sigma^2 \sqrt{3/N}$, when a LDF is assumed, which is approximately $0.13\sigma^2$ for $N = 180$ and exactly $0.10\sigma^2$ for $N = 300$. When values of 30-day totals are estimated for each season, these values for N correspond to series with a length of 60 and 100 years, respectively. The variances can be estimated better by taking into account all possible k-day consecutive rainfall amounts. For instance, in a season with a length of 90 days (winter) there are 61 different consecutive periods of 30 days, but because of the large correlation between rainfall amounts of adjacent periods estimates based on these overlapping data are only slightly better.

For $k = 5(5)50$ estimates of variances of k-day sums are given in Figure 6.3 for the wet-dry process and the entire rainfall process of Winterswijk ($\delta = 0.3$ mm). The estimates are slightly biased, because the number of k-day periods was used in the denominator. The points lie approximately on a straight line, which intersects the vertical axis at the negative side. The negative intercept can be caused by the large skewness of the distribution of lengths of weather spells. This will be explained further in IV, 5.2. Notice also in Figure 6.3 that variance-time curves of the rainfall process can differ considerably from season to season.

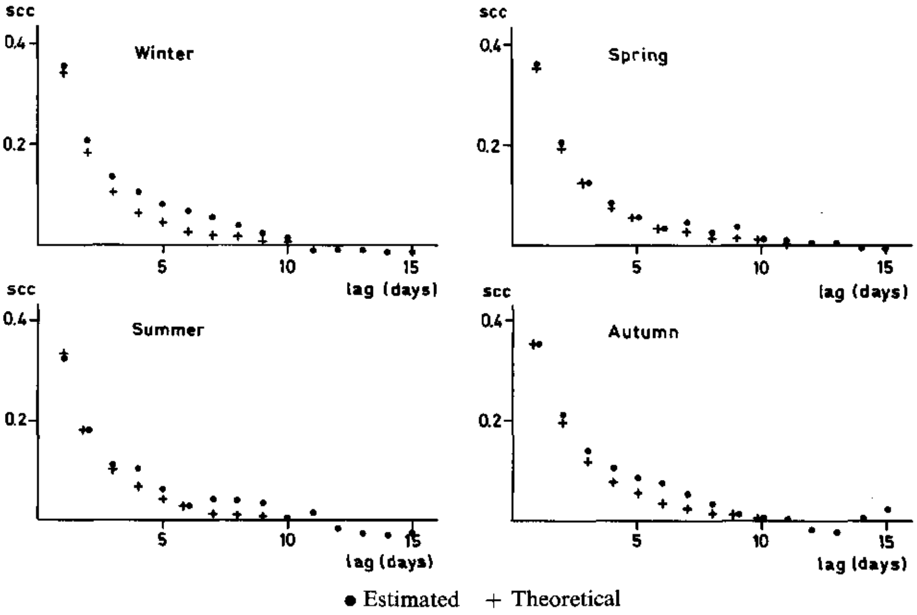


FIG. 6.1. Estimated (WD) and theoretical correlograms of wet-dry processes for Winterswijk ($\delta = 0.3$ mm). The theoretical sccs are based on TNBDs for lengths of weather spells.

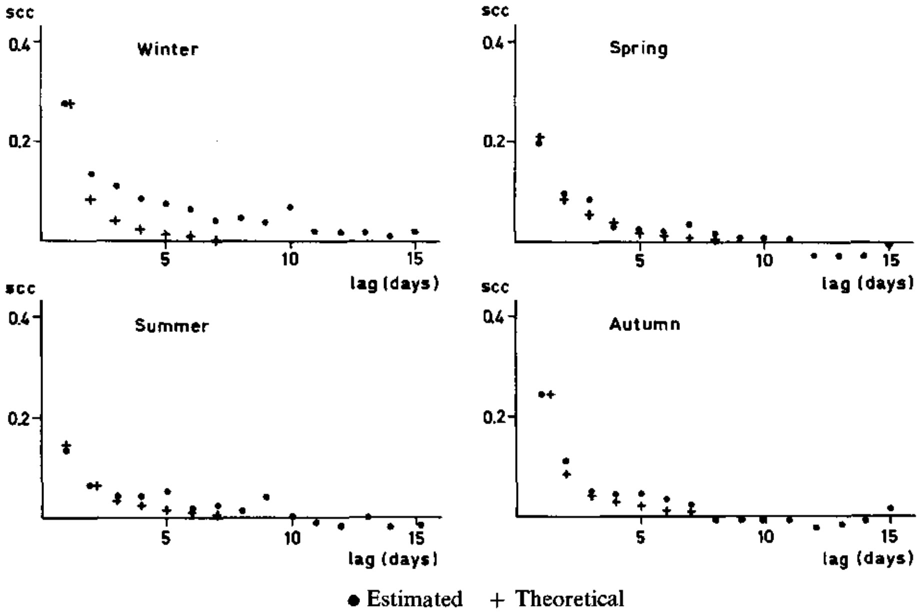


FIG. 6.2. Estimated (WD) and theoretical correlograms for Winterswijk ($\delta = 0.3$ mm). The theoretical sccs are based on model IV with TNBDs for lengths of weather spells and SGDs for rainfall amounts on wet days.

TABLE 6.1. Means of estimates of sccs of Winterswijk ($\delta = 0.8$ mm). The quarterly mean is the mean of WD and DW estimates for winter, spring, summer and autumn. Likewise the monthly mean is the mean of 12 monthly WD estimates and 12 monthly DW estimates.

	Wet-dry process		Rainfall process	
lag (days)	Quarterly mean	Monthly mean	Quarterly mean	Monthly mean
1	0.315	0.314	0.210	0.212
2	0.175	0.174	0.104	0.102
3	0.109	0.105	0.074	0.070
4	0.084	0.080	0.047	0.043
5	0.061	0.053	0.047	0.041

the mean hardly influence estimates of the variance and cross correlation coefficients for rainfall amounts within small time-increments. This can also be shown for the serial correlation coefficient with Equation (26) or (47) of YEVJEVICH and JENG (1969). The influence of the seasonal variation can be verified by comparing means of quarterly estimates of sccs with means of monthly estimates, which is done in Table 6.1 for the Winterswijk series ($\delta = 0.8$ mm). From the tabulated sccs it can be concluded that seasonal variation cannot explain the slow decay of the estimated correlograms in Figures 6.1 and 6.2, which can also be verified roughly by analytical methods.

Estimated sccs can be compared with sccs of the model. The method of calculation of the theoretical sccs will be given in IV, 4. Theoretical sccs can be calculated for models with different distributions for weather spells and rainfall amounts. With respect to the behaviour of rainfall amounts on wet days, four models can be distinguished.

Model I assumes no discrimination between different types of rainfall amounts and no serial correlation between rainfall amounts within a wet spell.

Model II assumes no discrimination between different types of rainfall amounts. Rainfall amounts within a wet spell are assumed to follow a first order moving average process which has the property that sccs of lags greater than 1 are zero. The use of a first order moving average process is based on the results of Section 5.1.

Model III distinguishes type 0, 1 and 2 amounts, but assumes no serial correlation between consecutive rainfall amounts within a wet spell.

Model IV distinguishes type 0, 1 and 2 amounts, and rainfall amounts within a wet spell are assumed to follow a first order moving average process.

Model IV is the most comprehensive model and contains all features of the simultaneous distribution of the rainfall amounts on wet days, described in the Sections 4 and 5. The marginal distribution of the rainfall amount is assumed to be a SGD in Section 6.

For the Winterswijk and Hoofddorp series Table 6.2 compares estimated

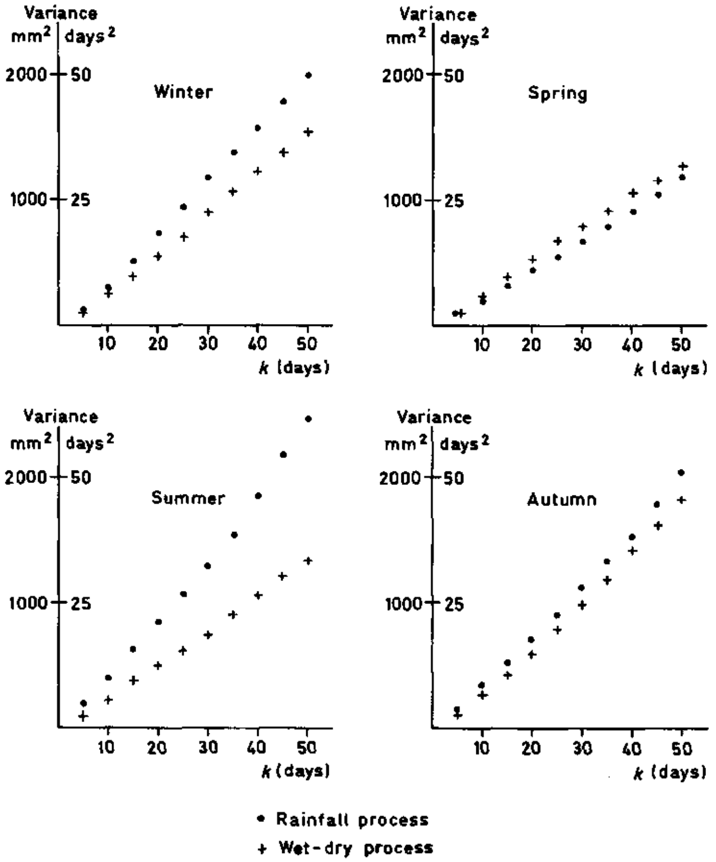


FIG. 6.3. Estimated variances of the number of wet days and of the rainfall amount for a k-day period ($k = 5(5)50$) for Winterswijk ($\delta = 0.3$ mm). Estimates are based on data with a k-1 day overlap.

Estimated and calculated variances for a 30-day period of different models with TNBDs for wet and dry intervals are given in Table 6.3. The method of calculation of theoretical variance-time curves is given in IV, 5.

Seasonal values of the parameters of the theoretical process are the same as those used for the calculation of the correlograms (see 6.1). Because of the estimation procedure by overlapping periods there is reason to give a larger weight to the parameters of the second month of a season, but this gives only small differences in most cases. A better but more laborious method for obtaining the variances of the model consists of calculating variances for each month separately and averaging the three monthly variances. However it will be seen in V, 2.4 that this method gives only small differences.

For the wet-dry process there is a reasonable correspondence between the estimated and calculated values, though the calculated values are in general a bit smaller during winter and autumn, which was to be expected from Figure

TABLE 6.3. Estimated and theoretical variances of the number of wet days in a 30-day period and of the 30-day rainfall total for models with TNBDs for wet and dry intervals and SGDs for rainfall amounts on wet days. Variances of the rainfall process are in mm^2.

		Wet-dry process		Rainfall process				
Station and	season	Estimated	Model	Estimated	Model I	Model II	Model III	Model IV
Winterswijk	Winter	22.0	18.9	1167	583	653	709	786
($\delta = 0.3$ mm)	Spring	20.0	20.0	654	508	525	564	582
	Summer	18.9	18.9	1286	1038	1053	1110	1126
	Autumn	24.6	20.0	1123	808	871	938	1005
Winterswijk	Winter	20.7	17.2	1167	639	678	759	804
($\delta = 0.8$ mm)	Spring	16.4	16.6	654	526	539	562	575
	Summer	15.2	16.6	1286	1073	1084	1114	1125
	Autumn	20.9	18.0	1123	878	919	976	1019
Hengelo	Winter	21.9	21.6	1113	746	781	931	971
($\delta = 0.8$ mm)	Spring	16.4	17.3	668	578	581	625	628
	Summer	17.2	17.3	1451	1116	1143	1208	1236
	Autumn	21.6	18.3	1195	818	846	940	971
Hoofddorp	Winter	21.4	17.9	742	551	573	618	643
($\delta = 0.8$ mm)	Spring	15.6	15.9	515	469	477	507	516
	Summer	18.0	16.4	1364	1124	1157	1196	1231
	Autumn	24.1	20.1	1817	1224	1297	1457	1538

6.1. A more complicated model for the behaviour of rainfall amounts gives larger variances of 30-day totals, but even for the most sophisticated model nearly all calculated values are smaller than the estimated values. From II, A4 it follows that these systematic differences cannot be explained by non-homogeneities due to changes of rain gauges or changes of site. For Winterswijk the seasonal means of variance estimates of monthly totals are 1196, 681, 1281 and 1249 mm^2 for winter, spring, summer and autumn, respectively which values hardly differ from those given in Table 6.3. Hence seasonal variation is sufficiently reduced. Also the height of the threshold δ has no remarkable influence on the results. Though the calculated values are systematically smaller than the estimated values, the differences are seldom larger than twice their standard deviation when model IV is assumed. The largest differences between estimated and theoretical values occur during winter and autumn, which was to be expected from Figure 6.2. In these seasons there are sometimes long rainy periods with high intensity, and the large value of the estimated variance for the winter of Winterswijk and Hengelo is mainly due to one such rainy period (see 4.2). Though differences between the estimates of corresponding seasons of the Winterswijk and Hengelo series are small, they can differ considerably from the estimates of the Hoofddorp series. The fact that differences between Dutch rainfall series can be quite large is also seen from the estimates of the standard deviation of monthly totals in II, Figure 3.1. These differences are partly due to rare rainfall events which occur very locally.

The distribution of lengths of weather spells affects the variance of 30-day

TABLE 6.4. Estimated and calculated standard deviations of the annual number of wet days.

			Calculated from estimated or theoretical 30-day variances	
Station	δ (mm)	Estimated	Estimated	TNBD-TNBD process
Winterswijk	0.3	17.2	16.0	15.3
Winterswijk	0.8	15.6	14.8	14.3
Hengelo	0.8	16.9	15.4	14.5
Hoofddorp	0.8	17.3	15.2	14.9

TABLE 6.5. Estimated and calculated standard deviations of annual rainfall totals (in mm). Calculated values are based on TNBDs for lengths of weather spells and SGDs for rainfall amounts on wet days.

			Calculated from estimated or theoretical 30-day variances				
Station	δ(mm)	Estimated	Estimated	I	II	III	IV
Winterswijk	0.3	133.3	112.6	93.8	96.5	99.8	102.5
Winterswijk	0.8	133.3	112.6	96.9	98.1	101.2	102.8
Hengelo	0.8	131.3	115.2	98.9	100.3	105.4	106.9
Hoofddorp	0.8	117.6	115.4	100.5	102.5	106.5	108.6

sums, especially when dealing with wet-dry processes. For the winter of Winterswijk ($\delta = 0.8$ mm) the 30-day variance is 19.5 days^2 for the GD–LSD process and 13.6 days^2 for the GD–GD process, and if model I is taken, the 30-day variances of the entire rainfall process are 694 and 556 mm^2, respectively.

From the seasonal variances of 30-day totals, given in Table 6.3, rough estimates of the variances or standard deviations of the annual totals can be calculated by multiplying these values by a factor 3 and adding the results. The annual standard deviations so obtained are given in the last two columns of Table 6.4 for the wet-dry process and in the last five columns of Table 6.5 for the entire rainfall process, together with the estimates based on non-overlapping annual periods in the second column of these tables. The last estimates can be somewhat higher, because:

a. they are more sensitive to departures from homogeneity (see II, A4),
b. they are based on a 365 (or 366)-day period, while the other values are based on a 360-day period,
c. they have no negative bias due to overlapping,
d. there exists a small (but not significant, see II, 3.1) serial correlation between monthly totals, which is not taken into account.

Notwithstanding these facts the large difference between the estimate based on non-overlapping annual totals and the one based on 30-day totals is surprising for the rainfall process of Winterswijk and Hengelo.

For the wet-dry process differences between estimated values and values according to the model are small, but for the entire rainfall process the differences can be quite large, depending on the type of model. It should be noticed again that the standard deviations of the estimates in Table 6.5 are quite large. For the entire rainfall process of Hoofddorp the standard deviation of the estimate based on non-overlapping annual totals was found to be about 8 mm using (6.2a) and the approximation

$$\text{var}\,\underline{s} \approx \text{var}\,\underline{s}^2/(4\sigma^2). \tag{6.3}$$

The standard deviations of the other estimates are of the same order.

7. Cumulative distribution functions of sums of daily rainfall amounts

In the previous section a study was made on the sensitivity of correlograms and variance-time curves for different features of the model. Working with correlograms or variance-time curves has the advantage that for many different types of models theoretical values can be obtained by numerical methods, but it is not easy to interpret a possible lack of fit. For instance, it is difficult to know the practical consequences of a model which underestimates 30-day variances. Another fact is that correlograms and variance-time curves only deal with second-order properties and they do not provide any information about higher order moments. Therefore the fit of different types of models is also assessed with cumulative distribution functions (cdfs) of k-day totals. Though for most types of models the cdf cannot be easily obtained by numerical methods, they may be preferred to correlograms and variance-time curves, because they give more information about the simultaneous distribution of daily rainfall amounts and besides, they are more readily interpretable in practice.

The sensitivity of the cdfs of k-day totals to the type of distribution of the lengths of weather spells is discussed in Section 7.1. Section 7.2 deals with cdfs of k-day totals for different models for the behaviour of the rainfall amounts on wet days. Also, in this section historic and synthetic sequences are compared on the basis of annual totals.

The results of Sections 7.1 and 7.2 are based on Winterswijk data with a threshold of 0.8 mm, unless another station or threshold is mentioned.

The k-day periods used for the estimation of the cdfs are the same as those on which the variance-time curves are based. As in Section 6 quarterly values for the parameters are obtained by averaging monthly estimates.

7.1. *The influence of the distribution of lengths of weather spells on the cumulative distribution function of k-day totals*

Different types of wet-dry processes can be compared by considering cdfs of the number of wet (or dry) days in a k-day period. In 6.1 nearly the same first

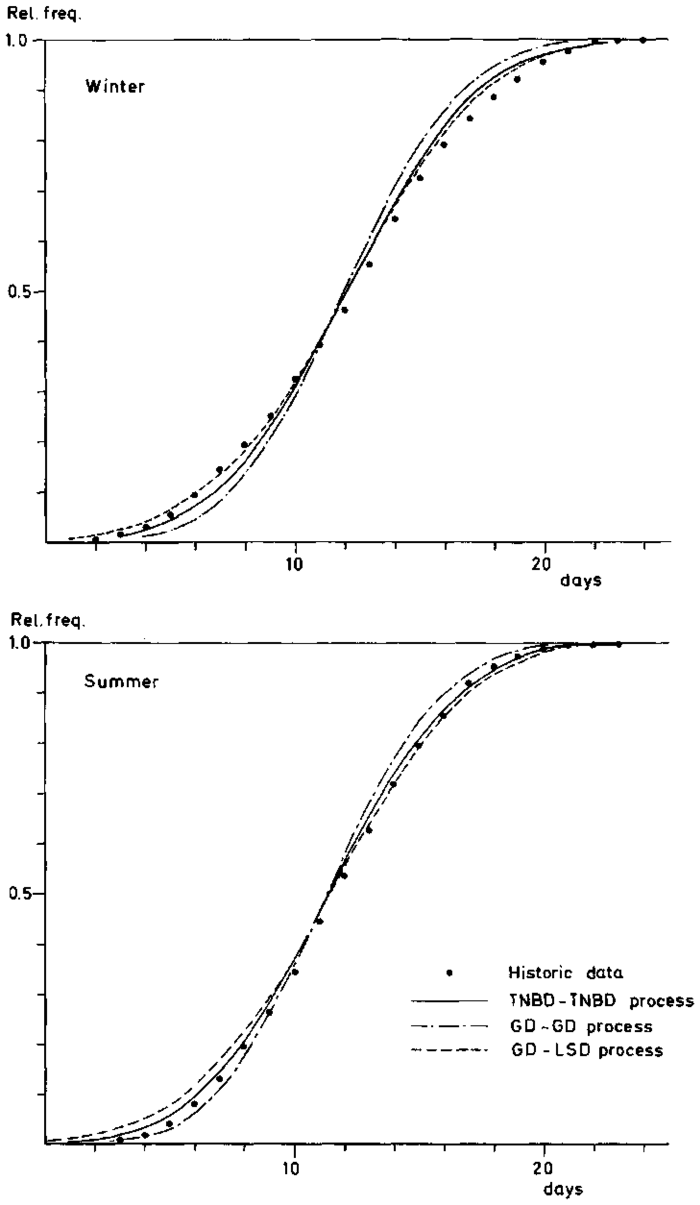


FIG. 7.1. Cumulative frequencies of the number of wet days in a 30-day period for the historic series of Winterswijk ($\delta = 0.8$ mm) and some theoretical wet-dry processes. Smooth curves are drawn through the theoretical values at the integers.

scc was found for different theoretical wet-dry processes and due to the estimation procedure there is also a correspondence in the mean. Therefore, it can be expected that for small k the cdfs of these wet-dry processes will coincide. However, in 6.2 it was shown that the 30-day variance of the wet-dry process was sensitive to the type of distribution of the lengths of weather spells. Therefore cdfs of the number of wet days in a 30-day period are compared.

Figure 7.1 compares estimated cdfs with theoretical cdfs of the wet-dry processes, considered in Section 6, for winter and summer. Formulas for the calculation of the theoretical cdfs are derived in IV,3. There are only small differences between different types of wet-dry processes and besides the theoretical cdfs fit the cdf of the historic data well. For the winter Figure 7.2 shows cdfs of TNBD–TNBD processes with and without seasonal variation. This figure also gives the cdf when the probability of a day being wet is independent of the situation on previous days (Bernoulli process). There are large differences between theoretical and estimated cdfs for the Bernoulli process and the TNBD–TNBD process without seasonal variation. For the Hoofddorp series, where seasonality of the wet-dry process is more obvious (see Figure 3.4 and Table 3.6) there are differences of more than 0.20 between the cdf of the historic data and the cdf of the non-seasonal TNBD–TNBD process.

The cdfs of rainfall totals are less sensitive to the type of wet-dry process than the cdfs of the number of wet days. This is seen from Figure 7.3 where

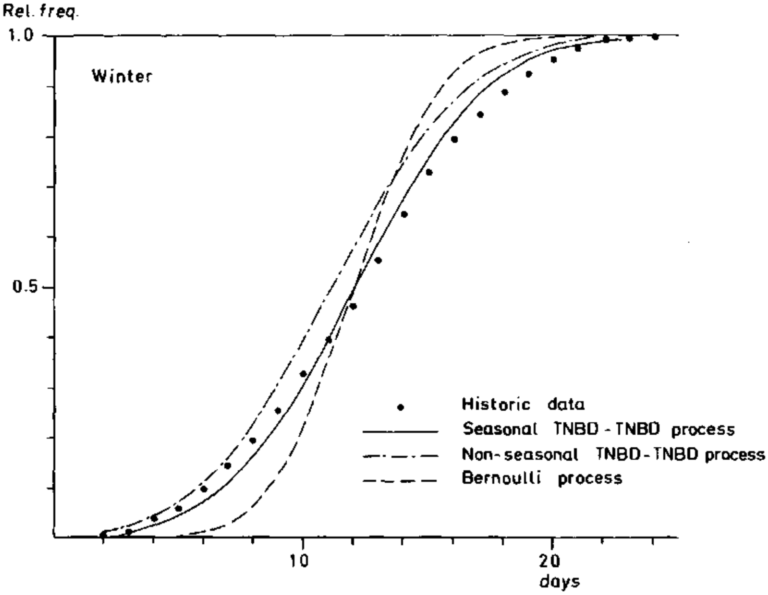


FIG. 7.2. Cumulative frequencies of the number of wet days in a 30-day period for the historic series of Winterswijk (δ = 0.8 mm) and some theoretical wet-dry processes. Smooth curves are drawn through the theoretical values at the integers.

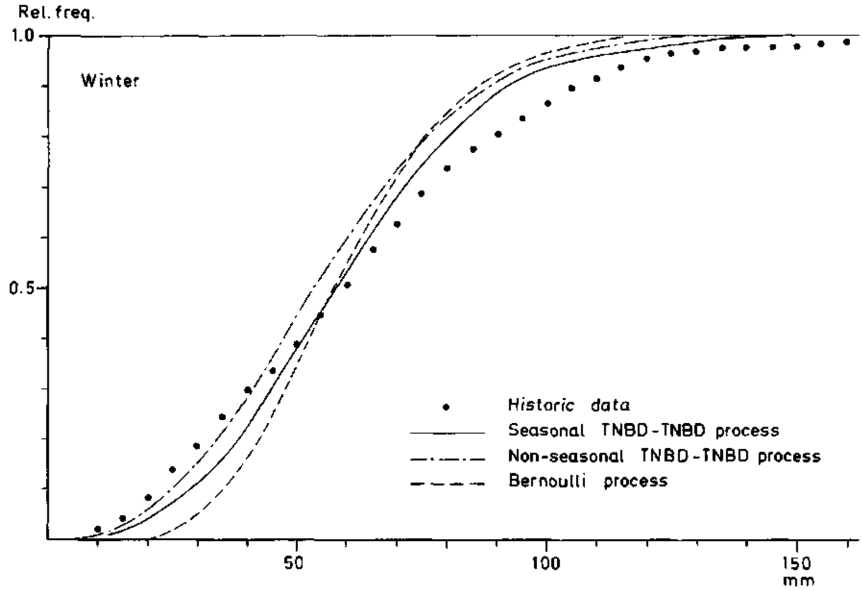


FIG. 7.3. Cumulative frequencies of 30-day totals of the historic series of Winterswijk and calculated values based on model I with a SGD for the rainfall amounts on wet days. The threshold δ is 0.8 mm.

cdfs of 30-day rainfall amounts are given for the processes considered in Figure 7.2. Calculated cdfs are based on model I with a SGD for rainfall amounts on wet days. Details about the computation of the cdfs are given in IV,3. Notice also in Figure 7.3 that the cdfs of all three models correspond poorly with the cdf of the historic data.

7.2. *The influence of the distribution of rainfall amounts on wet days on the cumulative distribution function of k-day totals*

This section deals with the sensitivity of cdfs of k-day totals to the distribution of rainfall amounts on wet days. In contrast with the comparisons of different wet-dry processes given in 7.1, it is not always possible here to get the cdf of some model by numerical methods. In fact only when rainfall amounts on wet days are independent and identically distributed (model I) is a numerical calculation of the cdf of k-day totals not complicated (see IV, 3). Otherwise the cdf of a particular model can be obtained by Monte Carlo simulation.

The generation of synthetic sequences is discussed in Section 7.2.1 and cdfs of different models are compared with empirical distribution functions in Section 7.2.2.

7.2.1. The generation of synthetic sequences

The generation of synthetic sequences consists of two parts, namely the generation of a wet-dry process and the generation of rainfall amounts on wet days.

Wet-dry sequences were generated by sampling alternately from TNBDs for wet and dry intervals. For the generation of variables with a TNBD, use was made of a table of cumulative frequencies and a geometric approximation in the tail in the following way. Let $\underline{y}$ denote the random length of a weather spell and let:

(7.1a) $$p_k = P(\underline{y} = k)$$

(7.1b) $$F_k = P(\underline{y} \leqslant k)$$

(7.1c) $$S_k = 1 - F_k = P(\underline{y} > k)$$

where S_k corresponds to the survivor function in Chapter IV.

For each month and type of weather spell a table of the F_ks was made for $k = 1(1)n$. The probabilities needed for this table were obtained from (3.3) with the recurrence relation:

(7.2) $$p_k = \frac{k + r - 1}{k} q p_{k-1} \qquad k = 1(1)n$$

starting with $p_0 = p^r/(1 - p^r)$.

From the recurrence relation it is seen that the TNBD tails off geometrically and therefore from n onwards use was made of the approximation:

(7.3) $$p_{n+l} = p_n \tilde{q}^l \qquad l = 1, 2, \ldots$$

where the parameter $\tilde{q}$ was chosen so that the p_ks add to unity.

To fulfil this requirement one must have:

(7.4) $$S_n = \sum_{l=1}^{\infty} p_{n+l} = p_n \tilde{q}/(1 - \tilde{q})$$

or

(7.5) $$\tilde{q} = S_n/(S_n + p_n) = S_n/S_{n-1}.$$

The parameter $\tilde{q}$ is in general somewhat smaller than the parameter q which is seen from Table 7.1 where for $n = 16$, values of these parameters are compared for lengths of wet and dry spells of Winterswijk A8. The parameters of the TNBDs on which this table is based are the smoothed AML estimates given in Figure 3.4. From (7.3) and (7.4) it follows:

(7.6) $$S_x = S_n \tilde{q}^{x-n} \qquad x = n + 1, n + 2, \ldots$$

Now, starting from a pseudo-random variate u, an approximate TNBD variate was obtained from the algorithm:

- if $u \leqslant F_n$ take the value k for which holds: $F_{k-1} < u \leqslant F_k$,
- if $u > F_n$ take the smallest integer larger than: $n + \log\,[(1 - u)/S_n]/\log \tilde{q}$,

which follows from (7.6). Pseudo-random standard uniform variates were

TABLE 7.1. Comparison of q and $\tilde{q}$, when $n = 16$ for lengths of weather spells of Winterswijk A8.

	Dry spells		Wet spells	
Month	q	$\tilde{q}$	q	$\tilde{q}$
January	0.837	0.806	0.637	0.628
February	0.853	0.822	0.624	0.616
March	0.864	0.833	0.610	0.602
April	0.862	0.832	0.600	0.592
May	0.854	0.824	0.595	0.587
June	0.844	0.813	0.598	0.590
July	0.840	0.809	0.607	0.598
August	0.851	0.820	0.620	0.611
September	0.863	0.832	0.633	0.625
October	0.859	0.828	0.644	0.635
November	0.841	0.810	0.648	0.640
December	0.830	0.799	0.646	0.637

obtained from the function RAN of the DEC 10 computer, which is based on an article by PAYNE et al. (1969).

Procedures for generating independent gamma variates were given by JÖHNK (1964) and GREENWOOD (1974). The first procedure is suitable for small v and the last procedure, which is based on the Wilson-Hilferty transform (see II, 3.2) is suitable for large v. For rainfall amounts on wet days, where most values of v lie in the range 0.5 – 1.0 (see Figure 5.1) it does not matter very much which procedure is taken. Here Jöhnk's procedure was chosen.

Generating dependent gamma variates is much harder and more time consuming. However, from Table 6.3 it is seen that serial correlation of rainfall amounts on wet days only causes small differences in the variance of 30-day totals, especially when the threshold is at 0.8 mm. Therefore dependent gamma variates were not generated.

Models with an alternating renewal process for the occurrence of wet and dry days and mutually independent rainfall amounts on wet days are time reversible and so it does not matter whether one generates forwards or backwards in time. Here, synthetic sequences were generated forwards in time and therefore the parameter estimates of method A were used. The beginning date of a weather spell determines the month to which it belongs and for a wet spell it also determines the parameters of the SGD of its rainfall amount(s).

7.2.2. Comparison of cumulative frequencies of k-day totals of the historic sequence with cumulative distribution functions of different types of models for the behaviour of the rainfall amounts on wet days

The following topics are considered in this section:

a. The cdfs of k-day totals ($k = 1, 3, 10, 30$) of synthetic sequences, based on model III, are compared with those of the historic series. In some cases the cdfs of the synthetic sequences are also compared with calculated cdfs based on model I.

b. The effect of replacing a shifted exponential distribution by a SGD on the cdf of 30-day totals is discussed.

c. Normal probability plots of the annual number of wet days and of the annual totals of the synthetic sequences mentioned above are compared with those of the historic series.

For Winterswijk ($\delta = 0.8$ mm) five independent synthetic sequences were generated. The synthetic sequences were based on model III with TNBDs for lengths of weather spells and SGDs for rainfall amounts on wet days.

The parameters of the TNBDs were based on AML estimates. For the length of wet spells a Fourier series with 1 harmonic component was used to describe the seasonal variation of the parameter p (see 3.12); for the length of dry spells estimates of the parameter p were smoothed according to the moving average scheme (3.17).

For rainfall amounts of a particular type the shape parameter v of the SGD was assumed to be constant throughout the year, whereas the mean was seasonally changing. Monthly values of the mean were obtained by smoothing monthly ML estimates according to the moving average scheme (3.17) and these values were assigned to the 15th (February) or the 16th day of the month. For other days the mean of the SGD was obtained by linear interpolation.

Each synthetic sequence has a length of 67 years, but because it always starts with a complete weather spell, the first year is excluded to reduce the influence of this initial transient. In contrast with the historic sequence each February month has a length of 28 days.

For winter and summer, cdfs of 1, 3, 10 and 30-day totals of the historic sequence and of the synthetic sequences are given in Figure 7.4. At $k = 30$ also the cdfs of the other seasons are given. Just like the cdfs of the historic sequence, the cdfs of the synthetic sequences are based on overlapping periods. There are differences in the mean between the historic sequence and generated sequences because in the model all rainfall amounts less than δ millimeters are assumed to be zero and consequently there may be some problems in comparing the cdfs, especially for large k. Therefore some points of the cdf of the historic series are also given when all rainfall amounts less than δ millimeters are set to zero (modified historic data). The influence of this modification, however, is negligible. For small values of k there is a good correspondence between cdfs of the historic sequence and those of synthetic sequences, but for larger values of k the model may provide a poor fit, especially in the upper tail. For the winter season differences between the cdfs of the historic series and those of the synthetic sequences could be expected for large values of k since there exists a considerable difference for the 30-day variances (see Table 6.3), but for the summer season the poor fit is much harder to explain.

Smoothing of monthly estimates can affect the results. From Figure 7.4 it is seen that the model gives rise to a smaller mean in summer. To investigate the influence of smoothing in

FIG. 7.4. Cumulative frequencies of k-day totals of the historic series of Winterswijk and 5 synthetic series, based on model III. In the modified historic series all recorded positive rainfall amounts less than 0.8 mm are assumed to be zero. For the synthetic sequences the smallest and the largest value are only given when there are visible differences between cumulative relative frequencies at a certain rainfall depth. Points of the synthetic sequences are omitted when they coincide with values of the historic series.

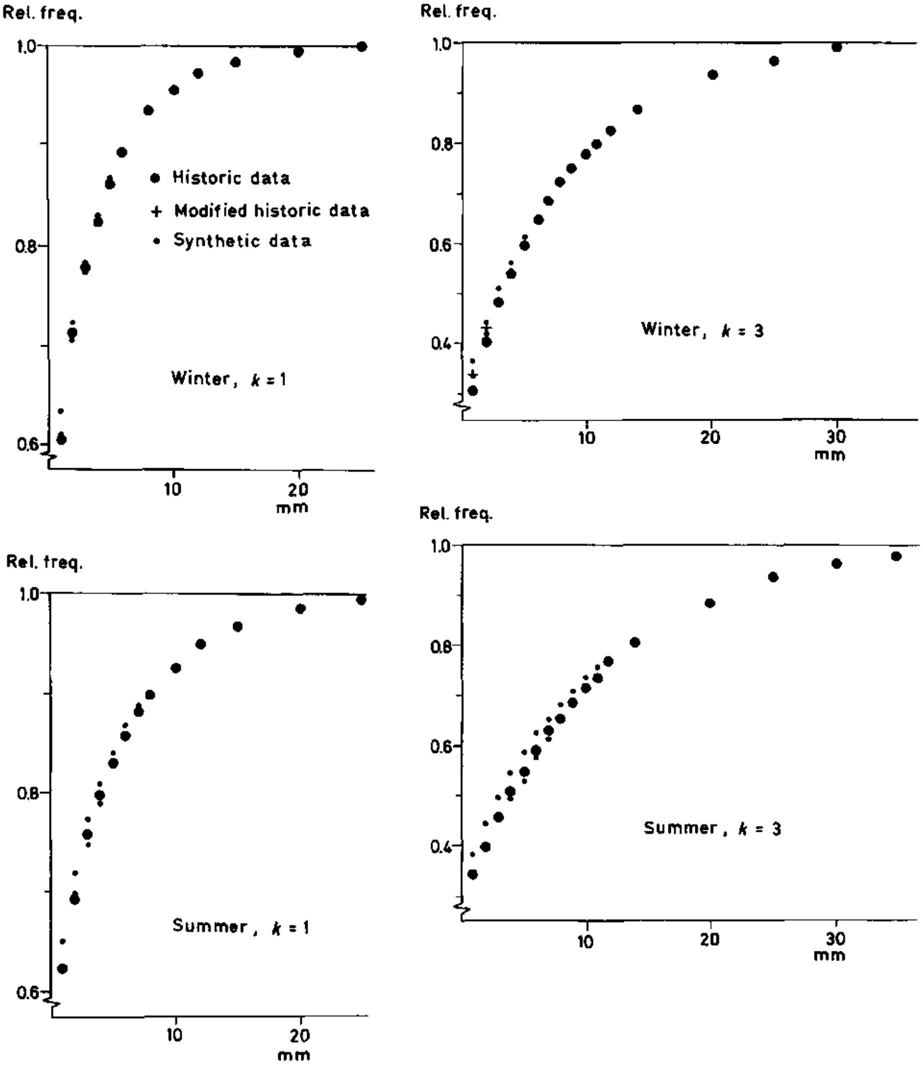
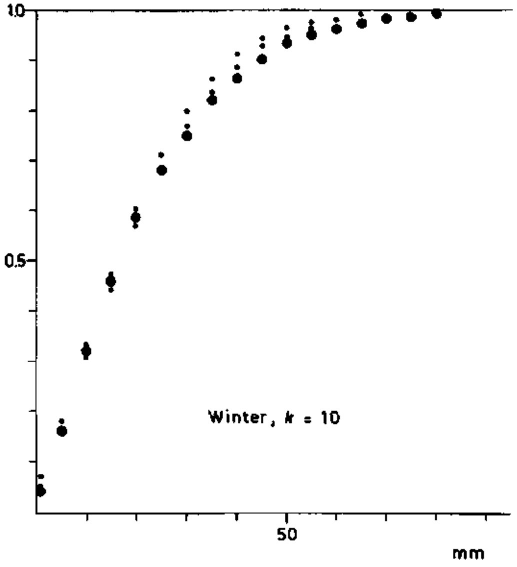


FIG. 7.4. (continued)



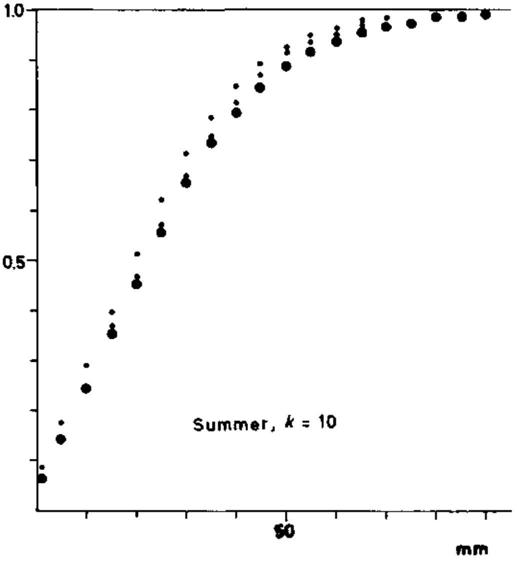


FIG. 7.4. (continued)

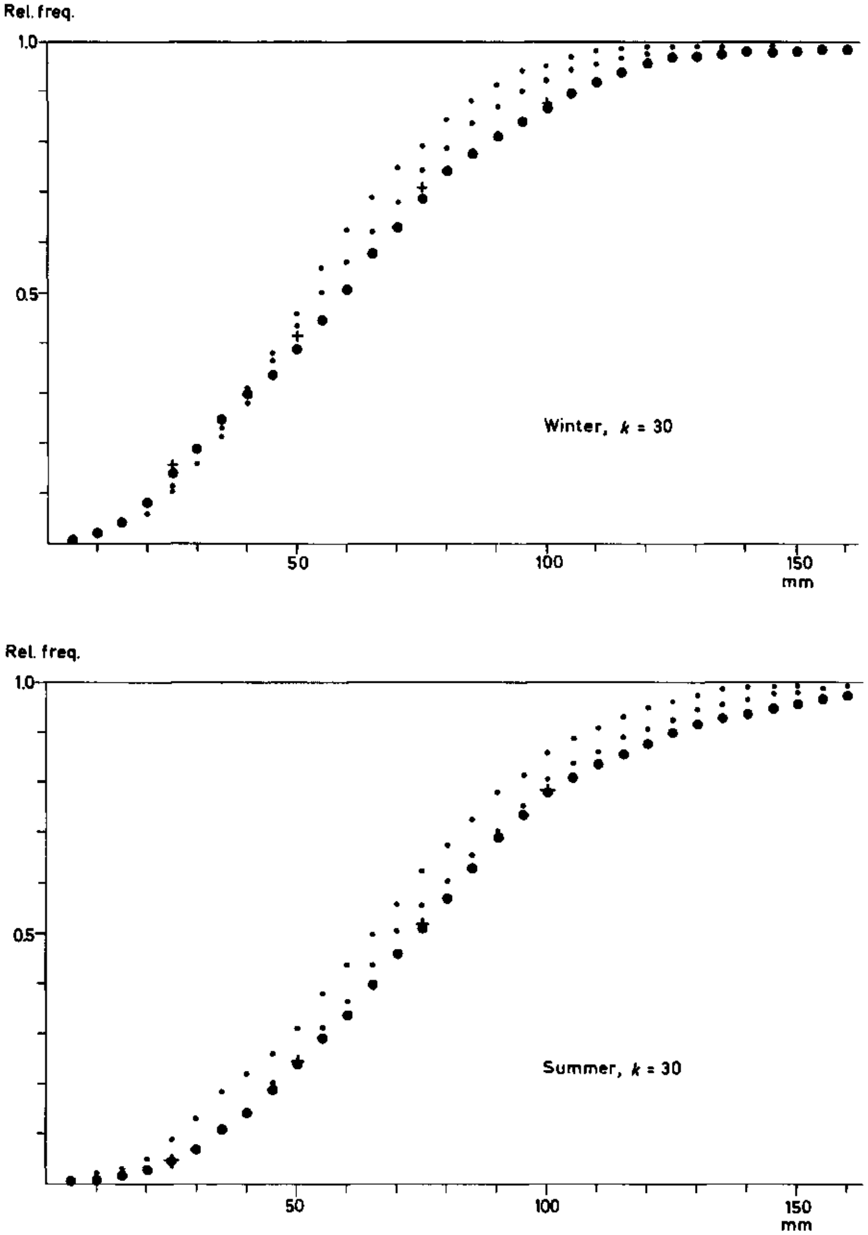
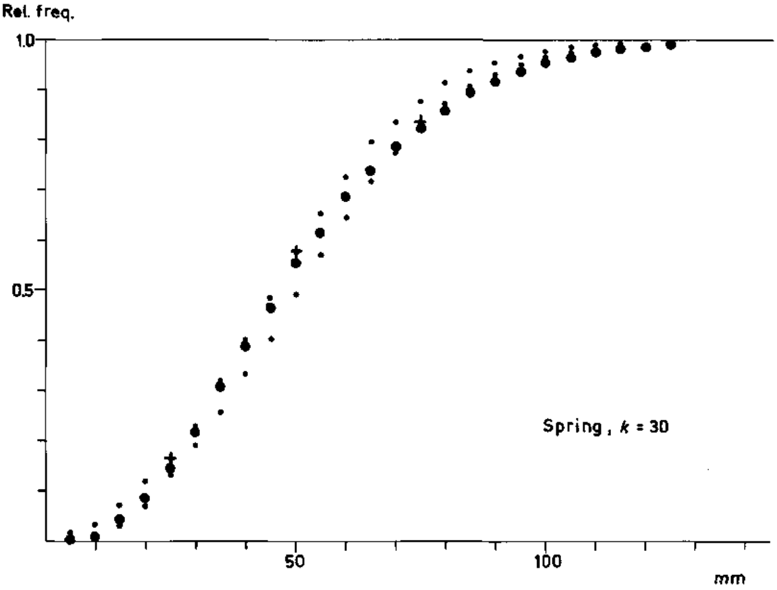
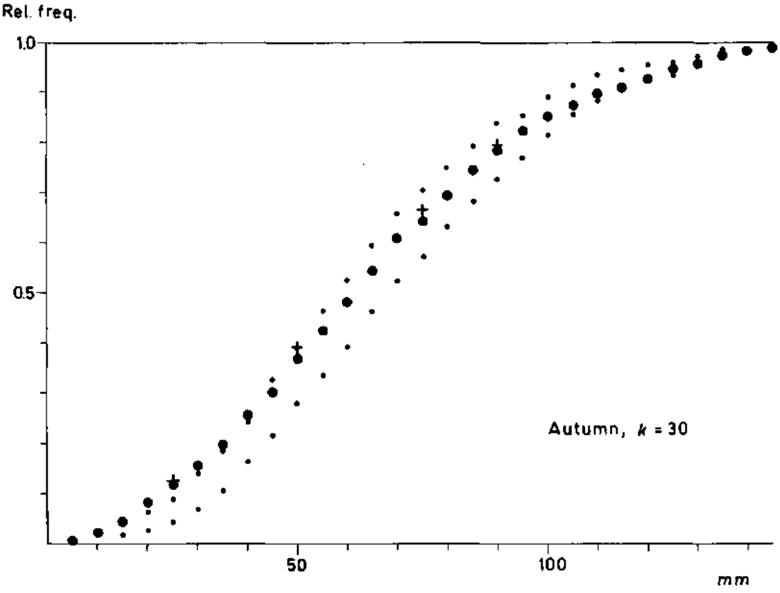


FIG. 7.4. (continued)





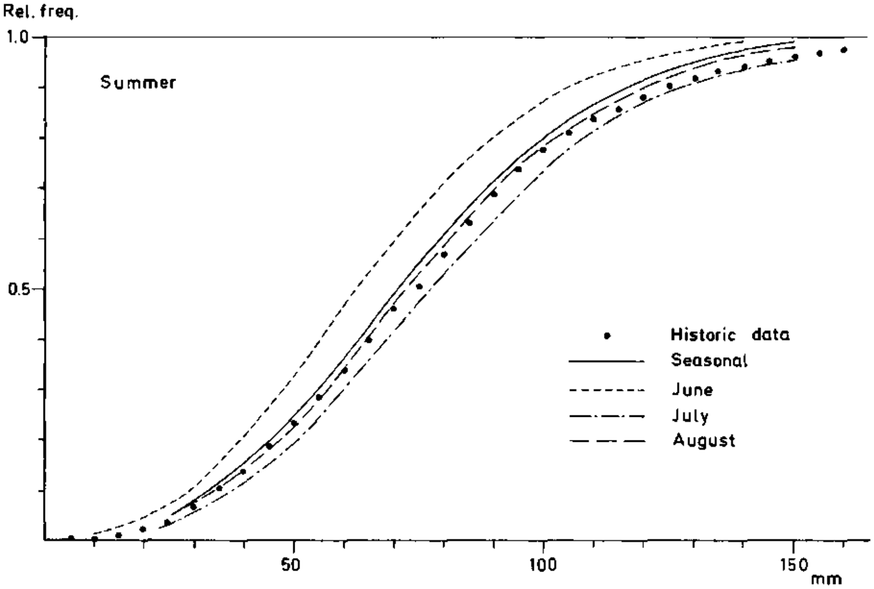


FIG. 7.5. Cumulative frequencies of 30-day totals of the historic series of Winterswijk and calculated cdfs. Theoretical curves are calculated for the whole summer season and for each month separately. The theoretical values are based on a TNBD–TNBD process and a SGD for rainfall amounts on wet days with a threshold of 0.8 mm.

more detail Figure 7.5 shows the theoretical cdf of 30-day totals for the whole summer season but also for each month separately.

The theoretical cdfs are based on model I with TNBDs for lengths of weather spells and a SGD for the rainfall amounts on wet days. Monthly and quarterly values of the parameters of the model are obtained here by averaging A and B estimates.

The curve based on seasonal means is very close to the curves of the synthetic series in Figure 7.4; this was to be expected since differences between type 0,1 and 2 amounts are small in summer.

The cdf of June differs strongly from the cdfs of July and August, which was to be expected on the basis of Figure 3.1 of Chapter II. The arithmetic means of the three monthly frequencies correspond quite well to the seasonal frequencies. For a comparison with the cdf of the historic sequence the cdf of the month of July is the most important one, because for the summer season 60 out of the 63 30-day periods have one or more days in this month. If a weighted average of the monthly cdfs is taken, the seasonal cdf corresponds better to the cdf of the historic data.

Though in winter there are considerable differences between different types of rainfall amounts the cdf of 30-day totals based on model III is nearly the same as the one based on model I (see Figure 7.3, seasonal TNBD–TNBD process and Figure 7.4).

Simulation of Hoofddorp data gives similar results, though the fit is usually slightly better.

In Section 5.2 it was noticed that the shifted exponential distribution did not

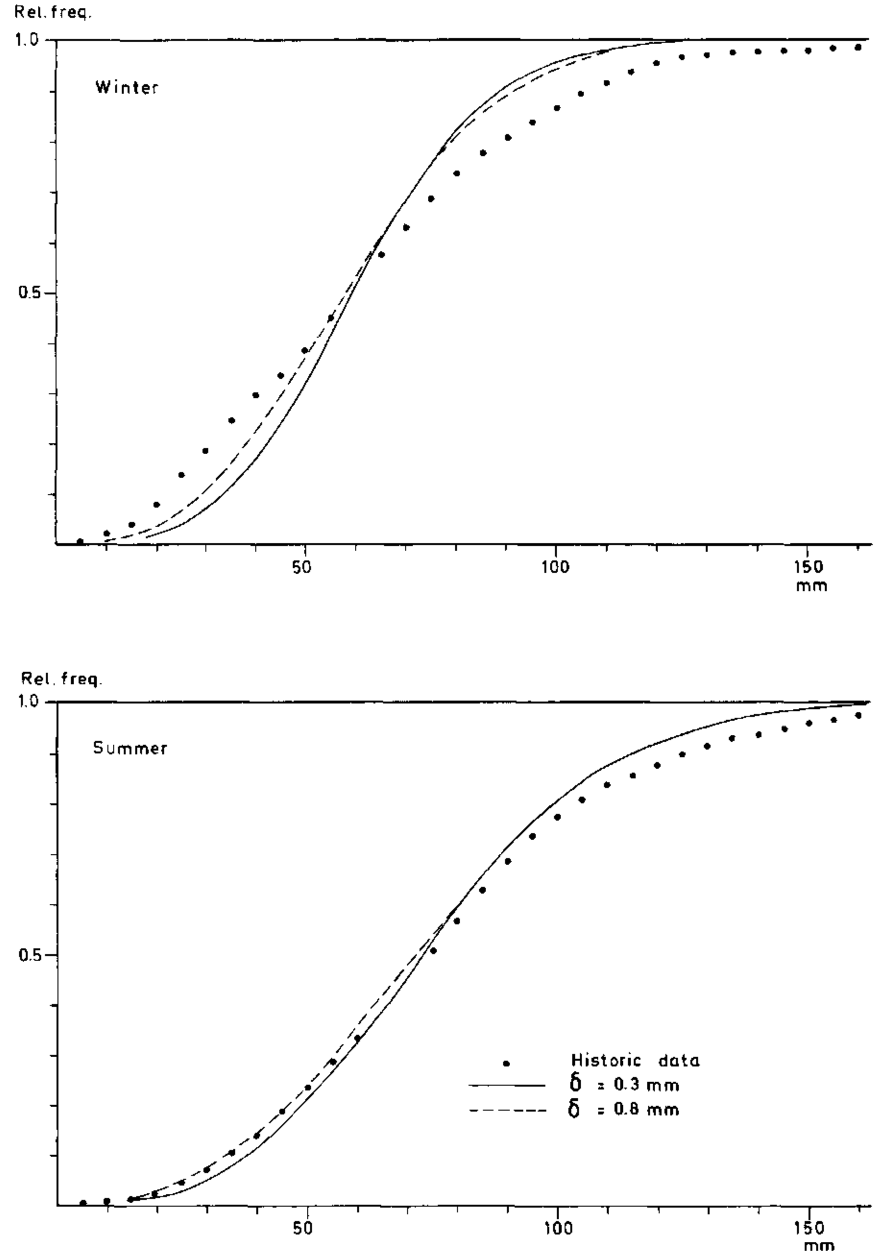


FIG. 7.6. Cumulative frequencies of 30-day totals for the historic series of Winterswijk and calculated values, based on model I, with a shifted exponential distribution for the rainfall amounts on wet days and TNBDs for lengths of weather spells.

provide a good fit to rainfall amounts on wet days, especially when the threshold was at 0.3 mm, because the distribution of rainfall amounts is more skew at a lower threshold. For model I with a TNBD–TNBD wet-dry process differences between 30-day cdfs based on rainfall amounts with a shifted exponential distribution and those based on a SGD are in general small when a threshold of 0.8 mm is used. These differences can be seen by comparing the 30-day cdfs based on rainfall amounts with a shifted exponential distribution, which are given in Figure 7.6, with the corresponding 30-day cdfs based on rainfall amounts with a SGD, given in Figure 7.3 and in Figure 7.5. When the threshold is at 0.3 mm the fit becomes poorer during winter.

Figure 7.7 shows the cdfs of the annual number of wet days of the historic series and of the generated sequences on normal probability paper. The cdfs of annual totals of these series are given in Figure 7.8. This figure also gives the cdf of a modified historic series, which is better comparable with the

FIG. 7.7.

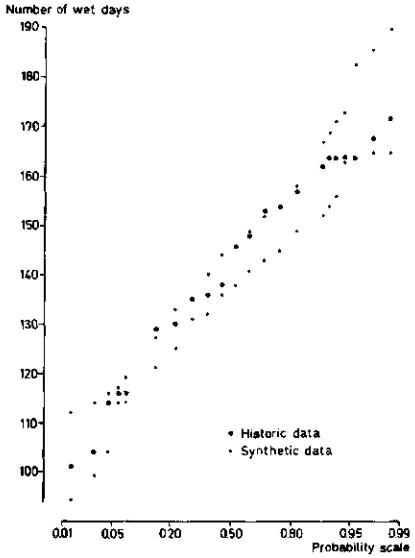


FIG. 7.8.

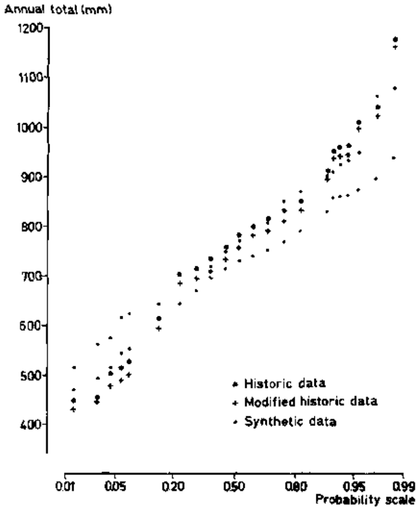


FIG. 7.7. Cumulative frequencies of the annual number of wet days for the historic series of Winterswijk ($\delta = 0.8$ mm) and 5 synthetic series based on model III. For the synthetic sequences only the smallest and the largest value of the cdf at fixed plotting positions are given.

FIG. 7.8. Cumulative frequencies of annual totals for the historic series of Winterswijk and 5 synthetic series, based on model III. In the modified historic series all recorded positive rainfall amounts less than 0.8 mm are assumed to be zero and the 29th of February is discarded. For the synthetic sequences only the smallest and the largest value of the cdf at fixed plotting positions are given.

synthetic series. The modification consists of replacing all positive values less than δ millimeters by zero and of omitting the 29th of February rainfall. There is a reasonable correspondence between the historic series and the synthetic sequences with respect to the annual number of wet days, but for the annual totals there are considerable differences. The synthetic series do not have extreme low and high values and if a straight line is fitted through the points, the slope is smaller for the synthetic series, which indicates that the variance is too small. The standard deviations of the synthetic series are 116.2, 97.3, 89.0, 114.5 and 90.6 mm, respectively; these correspond reasonably well with the theoretical value given in Table 6.5, but are much smaller than the estimated standard deviation of the historic series.

Generating synthetic data for Hoofddorp provides a better fit for annual totals as was to be expected from the variances given in Table 6.5.

8. Summary and concluding remarks

In this chapter daily rainfall sequences were analysed, first by investigating the sequence of wet and dry days and second by investigating the behaviour of rainfall amounts on wet days.

The wet-dry process can be described by an alternating renewal process, that is by a succession of wet and dry intervals, which are mutually independent. Modifications of the negative binomial distribution were fitted to the lengths of wet and dry spells. There is a seasonal change in the parameters of the distribution of the lengths of wet and dry spells.

The distribution of the rainfall amount on a particular wet day depends on the number of adjacent wet days. During winter and autumn there is a small, but significant correlation between rainfall amounts on successive wet days. The marginal distribution of rainfall amounts on wet days can be described by a shifted gamma distribution, with a seasonal changing scale parameter. Seasonal variation of the shape parameter is less obvious.

The goodness of fit of the model was tested by correlograms, variance-time curves and cumulative distribution functions of k-day ($k = 1, 2, \ldots$) totals. For large values of k the model underestimates the variance, because of lack of long-term persistence in the model. Very large differences between estimated and theoretical k-day variances, which sometimes occur during winter and autumn, are partly due to long wet periods of very high intensity in the historic record.

The model fits the cumulative distribution function of k-day totals poorly for large values of k (e.g. $k = 30$). The distribution of the lengths of wet and dry spells and the model for the behaviour of rainfall amounts on wet days (marginal distribution, discrimination of different types of rainfall amounts) have only a small influence on the cumulative distribution function of 30-day totals.

Though only a few results of the analysis of the Hengelo series have been given in this chapter, they do not differ very much from those of the nearby station of Winterswijk despite the large number of supplements and corrections in the series of Hengelo (see Chapter VI).

APPENDIX

A1. Estimation of the parameters of the gamma distribution by a modified maximum likelihood method

In this appendix the solution of the likelihood equations is given for the ML procedure suggested in Section 5.2.

From (5.1) the log likelihood $L(\lambda,\nu)$ reads:

$$\text{(A1.1)} \qquad L(\lambda,\nu) = N\nu \log \lambda - N \log \Gamma(\nu) + nh(\lambda,\nu) - \lambda \sum_{i=1}^{m} x_i + {} + (\nu-1) \sum_{i=1}^{m} \log x_i$$

where:

$$\text{(A1.2)} \qquad h(\lambda,\nu) = \log \left\{ \int_0^{\varepsilon} \exp(-\lambda x) x^{\nu-1} dx \right\}$$

and $N = n+m$, is the total number of observations. The maximization of $L(\lambda,\nu)$ with respect to λ and ν can be done iteratively by the Newton-Raphson method. The iteration formula is:

$$\text{(A1.3)} \qquad \begin{pmatrix} \lambda \\ \nu \end{pmatrix}_l = \begin{pmatrix} \lambda \\ \nu \end{pmatrix}_{l-1} - (1-\omega^l) \begin{pmatrix} L_{\lambda\lambda} & L_{\lambda\nu} \\ L_{\lambda\nu} & L_{\nu\nu} \end{pmatrix}^{-1}_{l-1} \begin{pmatrix} L_\lambda \\ L_\nu \end{pmatrix}_{l-1} \qquad l = 1,2,\ldots$$

For the relaxation factor ω the value 0.9 was chosen. Moment estimates were used as starting values λ_0 and ν_0.

The first and second derivatives of $L(\lambda,\nu)$ are obtained by differentiation of (A1.1):

$$\text{(A1.4a)} \qquad L_\lambda = \frac{N\nu}{\lambda} + nh_\lambda - \sum_{i=1}^{m} x_i$$

$$\text{(A1.4b)} \qquad L_\nu = N \log \lambda - N\psi(\nu) + nh_\nu + \sum_{i=1}^{m} \log x_i$$

$$\text{(A1.4c)} \qquad L_{\lambda\lambda} = -\frac{N\nu}{\lambda^2} + nh_{\lambda\lambda}$$

$$\text{(A1.4d)} \qquad L_{\lambda\nu} = \frac{N}{\lambda} + nh_{\lambda\nu}$$

$$\text{(A1.4e)} \qquad L_{\nu\nu} = -N\psi'(\nu) + nh_{\nu\nu}$$

where ψ and ψ' denote the digamma and trigamma functions (first and second derivatives of the logarithm of the gamma function). A numerical method to obtain these functions was given by CHOI and WETTE (1969).

For small ε it follows from (5.2) and (A1.2):

(A1.5) $$h(\lambda,\nu) \approx \nu \log \varepsilon - \log \nu + \log (1-\lambda\varepsilon + \frac{\lambda\varepsilon}{\nu+1})$$

and so the derivatives of $h(\lambda,\nu)$ are approximately:

(A1.6a) $$h_{\lambda} \approx \frac{-\nu\varepsilon}{1+\nu-\lambda\nu\varepsilon}$$

(A1.6b) $$h_{\nu} \approx \log \varepsilon - \frac{1}{\nu} - \frac{\lambda\varepsilon}{(1+\nu)\,(1+\nu-\lambda\nu\varepsilon)}$$

(A1.6c) $$h_{\lambda\lambda} \approx \frac{-(\nu\varepsilon)^2}{(1+\nu-\lambda\nu\varepsilon)^2}$$

(A1.6d) $$h_{\lambda\nu} \approx \frac{-\varepsilon}{(1+\nu-\lambda\nu\varepsilon)^2}$$

(A1.6e) $$h_{\nu\nu} \approx \frac{1}{\nu^2} + \frac{\lambda\varepsilon(2+2-2\lambda\nu\varepsilon-\lambda\varepsilon)}{(\nu+1)^2\,(1+\nu-\lambda\nu\varepsilon)^2}.$$

IV. THEORETICAL CONSIDERATIONS ABOUT THE DAILY RAINFALL MODEL

1. Introduction

In the previous chapter the adequacy of the daily rainfall model was tested by correlograms, variance-time curves and the cumulative distribution function (cdf) of k-day totals. Theoretical correlograms, variance-time curves and, for some special cases the cdf of k-day totals could be obtained by numerical methods. In this chapter the formulas underlying these computations are derived. These formulas are based on a rainfall process in which the rainfall amounts on wet days are at least $\tilde{\delta}$ mm ($\tilde{\delta}$ is the height of the threshold minus half the unit of measurement, see III, 5.2) and the rainfall amounts on other days are assumed to be zero. This modified rainfall process will be denoted by $\{\underline{x}_t\}$ and the wet-dry process by $\{\underline{n}_t\}$; $\underline{n}_t$ takes the value 0 if the tth day is dry and the value 1 if the tth day is wet.

The formulas derived in this chapter, underly the assumption of independence of successive wet and dry intervals. This assumption looks reasonable in view of the results in III, 3.1. Another assumption is that the processes $\{\underline{n}_t\}$ and $\{\underline{x}_t\}$ are stationary, which is approximately true if one considers the rainfall process for a particular month or season.

In the cases considered here the main difficulty is the derivation of features of the wet-dry process. For instance, if one considers a rainfall process with iid rainfall amounts on wet days (model I, see III, 6.1) with a SGD (shifted gamma distribution), the only problem for the derivation of the cdf of k-day totals is the derivation of the probability distribution of the number of wet days in a k-day period. If one wants an expression for the lag k serial correlation coefficient of this model, simultaneous probabilities like $P(\underline{n}_t = 1, \underline{n}_{t+k} = 1)$ must be derived from the distribution of wet and dry intervals. Because features of the rainfall process depend on the wet-dry process, some concepts about this process are derived in the beginning of this chapter (Section 2). Expressions for the calculation of the cdf of k-day totals are given in Section 3. Section 4 deals with the serial correlation coefficients (sccs) for models I, II, III and IV, which are defined in III, 6.1. The behaviour of variance-time curves is discussed in Section 5.

For reading this chapter it is assumed that the reader has some knowledge of conditional probability and conditional expectation. The required level is that of the contents of Chapter 2 of Parzen (1962). Also familiarity with generating functions is a prerequisite; the main results on these, however, are given in Appendix A1.

2. Renewal theory

This section deals with properties of wet-dry processes for which the lengths of successive wet and dry spells are independent. Further the wet-dry process is assumed to be a process in discrete time.

In Section 2.1 the renewal process is introduced. The concepts discussed in this section are mainly adopted from Chapter XIII of Feller (1968). Further, it is shown in Section 2.1 that for a renewal process at least one type of interval (wet or dry) has a geometric distribution. A generalization of the renewal process is the alternating renewal process for which lengths of both wet and dry spells can have arbitrary distributions. Alternating renewal processes are discussed in Chapter 7 of Cox (1962). However, this author dealt with processes in continuous time only. Concepts of alternating renewal processes in discrete time are given in Section 2.2 and it will be seen that most relations are nearly equivalent to the corresponding relations in continuous time. A short review of the concepts of renewal theory was given by Bernier (1967). This author also indicated possible applications of renewal theory in hydrology.

2.1. *Renewal processes*

2.1.1. Ordinary renewal processes

In Figure 2.1 a realization of a wet-dry process is given for $t = 0(1)16$. The waiting times between successive wet days are denoted by $R_1, R_2, \ldots$ and are called recurrence times. If the recurrence times are iid random variables the process is called a renewal process. Then the process is independent of its history whenever a wet day occurs. Events with this property are called recurrent events.

Let ε be a recurrent event and let $\{f_n\}$ denote the distribution of the recurrence times, that is:

$$\text{(2.1)} \qquad \begin{cases} f_1 = P\{\varepsilon \text{ at } t = m + 1 \mid \varepsilon \text{ at } t = m\} & \\ f_n = P\{\text{no } \varepsilon \text{ at } (m, m+n), \varepsilon \text{ at } t = m + n \mid \varepsilon \text{ at } t = m\} & n > 1 \end{cases}$$

for all $m \geqslant 0$. Sometimes it is convenient to define $f_0 = 0$.

The distribution $\{f_n\}$ is called periodic if an integer $k \geqslant 2$ exists such that f_n only takes non-zero values for multiples of k. Otherwise the distribution $\{f_n\}$ is called non-periodic. When dealing with rainfall processes it is assumed that the recurrence times are non-periodic.

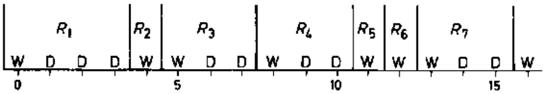


Fig. 2.1. Realization of a wet-dry process for $t = 0(1)16$. Wet days are denoted by W and dry days by D. The recurrence times $R_1, R_2, \ldots$ are the waiting times between successive wet days.

The survivor function m_n is defined by:

$$(2.2) \qquad m_n = \sum_{k=n+1}^{\infty} f_k \qquad\qquad n = 0, 1, \ldots$$

It is assumed that $\{f_n\}$ is an honest (or non-defective) probability distribution, that is:

$$(2.3) \qquad m_0 = \sum_{k=1}^{\infty} f_k = 1.$$

The following relation exists between the probability generating function (pgf), $F(s)$, of the recurrence times and the generating function (gf), $M(s)$, of the survivor function (the gf and the pgf are defined in Appendix A1):

$$(2.4) \qquad M(s) = (1-F(s))/(1-s)$$

(cf. FELLER (1968), XI.1, Theorem 1).

With reference to Figure 2.1, it might be interesting to investigate the relation between the recurrence times and the lengths of wet and dry spells. Let $\{f_n^{(w)}\}$ and $\{f_n^{(d)}\}$ denote the distributions of wet and dry intervals and denote the occurrence of a wet day by ε. It is assumed that ε is a recurrent event.

For the distribution of dry intervals one has:

$$(2.5) \qquad \begin{aligned} f_n^{(d)} &= P\{\varepsilon \text{ at } t=0, \text{no } \varepsilon \text{ at } [1,n], \varepsilon \text{ at } t=n+1 \mid \varepsilon \text{ at } t=0, \text{no } \varepsilon \text{ at } t=1\} = \\ &= \frac{P\{\varepsilon \text{ at } t=0, \text{no } \varepsilon \text{ at } [1,n], \varepsilon \text{ at } t=n+1 \mid \varepsilon \text{ at } t=0\}}{P\{\text{no } \varepsilon \text{ at } t=1 \mid \varepsilon \text{ at } t=0\}} = \\ &= f_{n+1}/m_1 \qquad\qquad n = 1, 2, \ldots \end{aligned}$$

and $\{f_n^{(d)}\}$ has pgf:

$$(2.6) \qquad F^{(d)}(s) = \sum_{n=1}^{\infty} f_{n+1}\, s^n/m_1 = (F(s)-f_1 s)/(m_1 s).$$

For the distribution of wet intervals:

$$(2.7) \qquad \begin{aligned} f_n^{(w)} &= P\{\text{no } \varepsilon \text{ at } t=0, \varepsilon \text{ for every } t \text{ at } [1,n], \text{no } \varepsilon \text{ at } t=n+1 \mid \text{no } \varepsilon \\ &\qquad \text{at } t=0, \varepsilon \text{ at } t=1\} = \\ &= \begin{cases} P\{\text{no } \varepsilon \text{ at } t=2 \mid \varepsilon \text{ at } t=1\} = 1-f_1 & n=1 \\ \prod_{k=2}^{n} P\{\varepsilon \text{ at } t=k \mid \varepsilon \text{ at } t=k-1\}\, P\{\text{no } \varepsilon \text{ at } t=n+1 \mid \varepsilon \text{ at } t=n\} \end{cases} \\ &= f_1^{n-1}(1-f_1) \qquad\qquad n>1 \end{aligned}$$

since ε is a recurrent event. So the distribution of wet intervals is geometric (see III, (3.4)) with mean $1/(1-f_1) = 1/m_1$ and pgf:

$$(2.8) \qquad F^{(w)}(s) = \frac{(1-f_1)s}{1-f_1 s}.$$

From (2.5) and (2.7) it is possible to obtain the distribution of the recurrence times from the distributions of the wet and dry intervals or the distribution of wet and dry intervals from the distribution of the recurrence times.

Relations between moments of the recurrence times and moments of wet and dry intervals can be obtained from an expansion of (2.6) in powers of $s-1$ (see Appendix A1) or alternatively as follows. Let $\underline{y}$ stand for the recurrence time and let $\underline{y}^{(w)}$ and $\underline{y}^{(d)}$ stand for the lengths of wet and dry spells. The means of $\underline{y}$, $\underline{y}^{(w)}$ and $\underline{y}^{(d)}$ are denoted by μ, $\mu^{(w)}$ and $\mu^{(d)}$, and the central moments by μ_k, $\mu_k^{(w)}$ and $\mu_k^{(d)}$ $(k = 0, 1, \ldots)$.

Moments of $\underline{y}$ and $\underline{y}^{(d)}$ are related by:

$$(2.9) \qquad \begin{aligned} E(\underline{y}^k) &= E(\underline{y}^k \mid \underline{y} > 1)P(\underline{y} > 1) + E(\underline{y}^k \mid \underline{y} = 1)P(\underline{y} = 1) = \\ &= m_1 E(\underline{y}^{(d)} + 1)^k + f_1. \end{aligned}$$

For instance, for $k = 1$:

$$(2.10) \qquad \mu = m_1(\mu^{(d)} + 1) + f_1 = m_1\mu^{(d)} + 1$$

since $m_1 = 1 - f_1$.

For the central moments one has:

$$(2.11) \qquad \begin{aligned} \mu_k = E(\underline{y}-\mu)^k &= E\{(\underline{y}-\mu)^k \mid \underline{y} > 1\}P(\underline{y} > 1) + \\ &+ E\{(\underline{y}-\mu)^k \mid \underline{y} = 1\}P(\underline{y} = 1) = \\ &= m_1 E\{(\underline{y}^{(d)} + 1 - \mu)^k\} + f_1(1-\mu)^k. \end{aligned}$$

Substitution of (2.10) gives:

$$(2.12) \qquad \begin{aligned} \mu_k &= m_1 E\{(\underline{y}^{(d)} - \mu^{(d)}) + f_1\mu^{(d)}\}^k + f_1(-m_1\mu^{(d)})^k = \\ &= m_1 \sum_{i=0}^{k} \binom{k}{i} (f_1\mu^{(d)})^{k-i}\mu_i^{(d)} + f_1(-m_1\mu^{(d)})^k. \end{aligned}$$

If the occurrence of a dry day is a recurrent event instead of the occurrence of a wet day, then dry intervals are geometrically distributed. Characteristic for the renewal process is that at least one type of interval is geometrically distributed.

The simplest case of a renewal process is the Bernoulli process. The wet-dry process is called a Bernoulli process if the probability of a day being wet or dry does not depend on the situation of previous days. Let p denote the probability of a day being wet and $q = 1 - p$ (the probability of a day being dry). Assuming that ε corresponds to a wet day, one obtains:

$$(2.13) \qquad f_n = pq^{n-1} \qquad \text{, from (2.1)}$$

$$(2.14) \qquad f_n^{(d)} = pq^{n-1} \qquad \text{, from (2.5)}$$

$$(2.15) \qquad f_n^{(w)} = qp^{n-1} \qquad \text{, from (2.7).}$$

So the recurrence times and lengths of wet and dry spells have geometric distributions. From (2.2) one gets for the survivor function:

$$(2.16) \qquad m_n = \sum_{k=n+1}^{\infty} pq^{k-1} = q^n.$$

A bit more complicated is the two-state first order Markov chain. This model has often been used to describe the sequence of wet and dry days (cf. GABRIEL and NEUMANN (1962), CASKEY (1963), WEISS (1964), TODOROVIC and WOOLHISER (1971) and SMITH and SCHREIBER (1973)).

In a two-state first order Markov chain the probability of a day being wet or dry depends on the state of the previous day. Let:

$$(2.17) \qquad p_{ij} = P\{\underline{n}_t = j \mid \underline{n}_{t-1} = i\} \qquad\qquad i = 0, 1;\ j = 0, 1.$$

From (2.1), with ε corresponding to a wet day ($\underline{n}_t = 1$), one obtains for the distribution of the recurrence times:

$$(2.18) \qquad \begin{cases} f_1 = p_{11} \\ f_n = p_{10}p_{00}^{n-2}p_{01} \qquad\qquad n > 1. \end{cases}$$

From (2.2) one obtains for the survivor function:

$$(2.19) \qquad \begin{cases} m_0 = 0 \\ m_n = p_{10}p_{01}p_{00}^{n-1}/(1-p_{00}) = p_{10}p_{00}^{n-1} \quad n > 0. \end{cases}$$

From (2.5), (2.18) and (2.19) it follows:

$$(2.20) \qquad f_n^{(d)} = p_{01}p_{00}^{n-1}$$

and from (2.7) and (2.18) one has:

$$(2.21) \qquad f_n^{(w)} = (1-p_{11})^{n-1}p_{11}^{n-1} = p_{10}p_{11}^{n-1}.$$

So for the two-state first order Markov chain, wet and dry intervals have geometric distributions.

In Figure 2.1 the process started with a wet day at $t = 0$. Renewal processes starting with an arbitrary recurrent event are called ordinary renewal processes. Let u_n denote the probability that ε occurs at $t = n$. By definition $u_0 = 1$; for $n > 0$ one can write:

$$(2.22) \qquad u_n = P\{\varepsilon \text{ at } t{=}n \mid \varepsilon \text{ at } t{=}0\} = P\{\varepsilon \text{ at } t{=}1 \text{ and } t{=}n \mid \varepsilon \text{ at } t{=}0\} + \\ + \sum_{k=2}^{n} P\{\text{no } \varepsilon \text{ at } [1,k{-}1],\ \varepsilon \text{ at } t{=}k \text{ and } t{=}n \mid \varepsilon \text{ at } t{=}0\}$$

since the events between brackets are mutually exclusive. Because ε is a recurrent event one has:

$$\begin{aligned}(2.23)\qquad & P\{\text{no } \varepsilon \text{ at } [1,k\text{–}1], \varepsilon \text{ at } t=k \text{ and } t=n \,|\, \varepsilon \text{ at } t=0\} = \\ & = P\{\text{no } \varepsilon \text{ at } [1,k\text{–}1], \varepsilon \text{ at } t=k \,|\, \varepsilon \text{ at } t=0\}P\{\varepsilon \text{ at } t=n \,|\, \varepsilon \text{ at } t=k\} = \\ & = f_k u_{n-k}\end{aligned}$$

and thus (2.22) becomes:

$$(2.24)\qquad u_n = f_1 u_{n-1} + \sum_{k=2}^{n} f_k u_{n-k} = \sum_{k=1}^{n} f_k u_{n-k} \qquad\qquad n > 0$$

(cf. FELLER (1968), XIII, (3.1)).

From (2.24) and the fact that $u_0 = 1$ it follows that for the gf $U(s)$ of $\{u_n\}$ holds:

$$(2.25)\qquad U(s) = 1/(1-F(s))$$

(cf. FELLER (1968), XIII.3).

2.1.2. Delayed and equilibrium renewal processes

In the previous section it was assumed that the process started with an arbitrary recurrent event at $t = 0$. One can also start with the condition that the waiting time till the first recurrent event has some probability distribution, say $\{b_n\}$. Then one calls the process a delayed or modified renewal process and the waiting time till the first recurrent event is called the forward recurrence time. The probabilities b_n are defined as follows:

$$(2.26)\qquad b_n = P\{\varepsilon \text{ occurs for the first time at } t = n\} \qquad n = 0, 1, \ldots$$

An example of a delayed renewal process is given in Figure 2.2. If the occurrence of a wet day is a recurrent event, the probability distribution of the waiting time R_1 differs from the probability distribution of the other waiting times R_2, R_3, ...

The probability of a recurrent event at $t = n$ is denoted by v_n. For the occurrence of ε at $t = n$ one has the following possibilities:

a. ε occurs for the first time at $t = n$ (with probability $b_n = b_n u_0$).

b. ε occurs for the first time at $t = k < n$ (with probability b_k). Starting at $t = k$ one has an ordinary renewal process and so the probability of ε at $t = n$ is u_{n-k}. Also for $t = k$ the process becomes independent of its history and so the probability that ε occurs for the first time at $t = k < n$ and ε occurs at $t = n$ is $b_k u_{n-k}$.

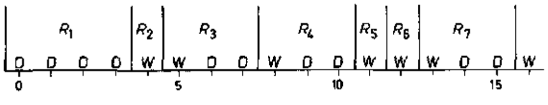


FIG. 2.2. Realization of a wet-dry process for $t = 0(1)16$. Wet days are denoted by W and dry days by D. The recurrence times R_2, R_3, ... are the waiting times between successive wet days. R_1 is the forward recurrence time.

Summing over all mutually exclusive possibilities gives:

$$(2.27) \qquad v_n = \sum_{k=0}^{n} b_k u_{n-k} \qquad\qquad n = 0, 1, \ldots$$

(cf. FELLER (1968), XIII, (5.1)), that is $\{v_n\}$ is the convolution of $\{b_n\}$ and $\{u_n\}$. Therefore, the gfs $V(s)$ of $\{v_n\}$, $B(s)$ of $\{b_n\}$ and $U(s)$ are related by (see Appendix A1):

$$(2.28) \qquad V(s) = B(s)U(s) = B(s)/(1-F(s))$$

(cf. FELLER (1968), XIII, (5.3)).

When $b_0 = 1$ and $b_n = 0$ for all $n \geqslant 1$, one has an ordinary renewal process and (2.28) reduces to (2.25).

Assume that $\{f_n\}$ is a non-periodic probability distribution with finite mean μ, then the next limit theorem holds:

$$(2.29) \qquad \lim_{n\to\infty} v_n = 1/\mu$$

(cf. FELLER (1968), XIII.10). So irrespective of the distribution of the forward recurrence time, the probability v_n tends to $1/\mu$ for large n.

For a stationary process the distribution of the forward recurrence time is such that $v_n = 1/\mu$ for all n and thus is:

$$(2.30) \qquad V(s) = \frac{1}{\mu} \sum_{n=0}^{\infty} s^n = \frac{1}{\mu(1-s)}.$$

Substituting this in (2.28) gives:

$$(2.31) \qquad B(s) = \frac{1-F(s)}{\mu(1-s)} = \frac{M(s)}{\mu}$$

or:

$$(2.32) \qquad b_n = m_n/\mu \qquad\qquad n = 0, 1, \ldots$$

The argument leading to (2.32) underlies the assumption of a renewal process. It can be shown, however, that this relation is applicable to a much wider class of processes (cf. COX (1962), 5.4).

Delayed renewal processes for which the forward recurrence time satisfies (2.32) are called equilibrium renewal processes.

2.2. *Alternating renewal processes*

For a wet-dry process the system can be in two states namely wet and dry. For theoretical considerations it is convenient to take numerical values for the possible states, for instance the values 1 and 2. Type 1 and 2 intervals can be defined in the same way as wet and dry intervals (see III, 2). A process with alternating type 1 and 2 intervals is said to be an alternating renewal process if:

a. All type 1 intervals have the same distribution, say $\{f_n^{(1)}\}$.
b. All type 2 intervals have the same distribution, say $\{f_n^{(2)}\}$.
c. All intervals are independent.

A type 1 event $\varepsilon^{(1)}$ occurs at $t = n$ if the system is in state 1 at $t = n$ and in state 2 at $t = n + 1$; a type 2 event $\varepsilon^{(2)}$ can be defined analogously. This terminology is the same as for the continuous alternating renewal process where the end of a type i interval ($i = 1, 2$) is called a type i event (cf. COX and MILLER (1965), 9.3). For instance, assume that for the rainfall process given in Figure 2.1 state 1 corresponds to a wet day and state 2 to a dry day, then $\varepsilon^{(1)}$ occurs at $t = 0, 5, 8$ and 13, and $\varepsilon^{(2)}$ occurs at $t = 3, 7, 10$ and 15. Because successive wet and dry intervals are independent, $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are recurrent events. Since the distribution of wet and dry intervals is arbitrary, it is possible that neither the occurrence of a wet day nor the occurrence of a dry day is a recurrent event.

The mean of type i intervals is denoted by $\mu^{(i)}$ and the survivor function of these intervals by $m_n^{(i)}$ (which is obtained by replacing f_n by $f_n^{(i)}$ in (2.2)). The analogue of (2.4) for the alternating renewal process is:

$$(2.33)\qquad M^{(i)}(s) = (1-F^{(i)}(s))/(1-s) \qquad i = 1, 2$$

where $M^{(i)}(s)$ and $F^{(i)}(s)$ are the gfs of $\{m_n^{(i)}\}$ and $\{f_n^{(i)}\}$, respectively.

The first probabilities of interest are:

$$(2.34)\qquad u_n^{(ij)} = P\{\varepsilon^{(j)} \text{ at } t = n \mid \varepsilon^{(i)} \text{ at } t = 0\} \quad i=1,2; j=1,2; n=0,1,\dots$$

By definition $u_0^{(ij)} = \delta_{ij}$ ($\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$). Further, if one considers type 1 events only, one gets a renewal process in which the recurrence times are the sums of the associated type 1 and 2 intervals. For this renewal process the probability u_n, defined in Section 2.1.1, corresponds to the probability $u_n^{(11)}$. Because the sequence formed by type 2 events is a renewal process with the same distribution for the recurrence times, one has the relation:

$$(2.35)\qquad u_n^{(11)} = u_n^{(22)} \qquad \text{for all } n.$$

Assume that $i \neq j$. One can write the probability $u_n^{(ij)}$ as the sum of probabilities of mutually exclusive events:

$$(2.36a)\qquad u_n^{(ij)} = \sum_{k=1}^{n} P\{\text{no}\,\varepsilon^{(j)} \text{ at } [0,k-1], \varepsilon^{(j)} \text{ at } t=k, \varepsilon^{(j)} \text{ at } t=n \mid \varepsilon^{(i)} \text{ at }$$

$$t=0\} = \sum_{k=1}^{n} f_k^{(j)} u_{n-k}^{(jj)} \qquad i \neq j, n = 1, 2, \dots$$

since the history of the process is irrelevant at $t = k$. A similar argument leads to:

(2.36b) $$u_n^{(ii)} = \sum_{k=1}^{n} f_k^{(j)} u_{n-k}^{(ji)} \qquad i \neq j,\ n = 1, 2, \ldots$$

For the derivation of correlograms of models III and IV (see III, 6.1) the following probabilities are required:

(2.37) $$w_n^{(ij)} = P\{\text{state } j \text{ at } t = n \mid \varepsilon^{(i)} \text{ at } t = 0\} \quad i=1,2; j=1,2; n=0,1,\ldots$$

Some special cases are: $w_0^{(ij)} = \delta_{ij}$ and $w_1^{(ij)} = 1 - \delta_{ij}$.

Assume that $i \neq j$, then $w_n^{(ij)}$ can be written as:

(2.38a) $$w_n^{(ij)} = P\{\text{no } \varepsilon^{(j)} \text{ at } [0,n] \mid \varepsilon^{(i)} \text{ at } t=0\} + \\ + \sum_{k=1}^{n} P\{\text{no } \varepsilon^{(j)} \text{ at } [0,k\text{–}1], \varepsilon^{(j)} \text{ at } t=k, \text{ state } j \text{ at } t=n \mid \varepsilon^{(i)} \text{ at} \\ t=0\} = m_n^{(j)} + \sum_{k=1}^{n} f_k^{(j)} w_{n-k}^{(jj)} \qquad i \neq j;\ n = 1, 2, \ldots$$

since the process becomes independent of its history at $t = k$. An analogous argument results in:

(2.38b) $$w_n^{(ii)} = \sum_{k=1}^{n} f_k^{(j)} w_{n-k}^{(ji)} \qquad i \neq j;\ n = 1, 2, \ldots$$

It is not necessary to use both (2.36a) and (2.36b) because:

(2.39) $$w_n^{(ii)} + w_n^{(ij)} = 1 \qquad i \neq j;\ n = 0, 1, \ldots$$

For obtaining the probabilities $w_n^{(ij)}$ with Equation (2.38), no use is made of the probabilities $u_n^{(ij)}$. There are methods which give the $\{w_n^{(ij)}\}$ sequences from the $\{u_n^{(ij)}\}$ sequences. For instance, let $\underline{k}_n^{(ij)}$ be the number of type j events at $[0, n\text{–}1]$, given that a type i event occurs at $t = 0$. Assume $i \neq j$ and define:

(2.40) $$\underline{g}_n^{(i)} = \underline{k}_n^{(ii)} - \underline{k}_n^{(ij)}.$$

The variable $\underline{g}_n^{(i)}$ can only take the values 0 and 1 since the two states are alternating. If $\underline{g}_n^{(i)} = 1$ the process is in state j at $t = n$; if $\underline{g}_n^{(i)} = 0$ the process is in state i at $t = n$ and thus is:

(2.41a) $$w_n^{(ij)} = P(\underline{g}_n^{(i)} = 1) = E(\underline{g}_n^{(i)}) = E(\underline{k}_n^{(ii)}) - E(\underline{k}_n^{(ij)}) = \\ = \sum_{k=0}^{n-1} u_k^{(ii)} - \sum_{k=0}^{n-1} u_k^{(ij)} \qquad i \neq j;\ n = 1, 2, \ldots$$

A similar argument gives:

(2.41b) $$w_n^{(ii)} = 1 + \sum_{k=0}^{n-1} u_k^{(ij)} - \sum_{k=0}^{n-1} u_k^{(ii)} \qquad i \neq j;\, n = 1, 2, \ldots$$

which could also be found from (2.39) and (2.41a). The relation (2.41b) is the discrete analogue of Equation (7.3.3) of Cox (1962).

So far alternating renewal processes have been considered which started with an arbitrary recurrent event ($\varepsilon^{(1)}$ or $\varepsilon^{(2)}$) at the origin. For the derivation of features of the rainfall model it is necessary to consider equilibrium processes.

For an alternating renewal process three types of equilibrium processes can be distinguished:

a. The system is in state 1 at $t = 0$ and the waiting time to the first type 1 event (forward recurrence time) has probability distribution $\{b_n^{(1)}\}$ with $b_n^{(1)} = m_n^{(1)}/\mu^{(1)}$ ($n = 0, 1, \ldots$). The process is called a type 1 equilibrium process and the state of the process at $t = 0$ is denoted briefly by type 1 eq. Notice that $\varepsilon^{(1)}$ can occur at $t = 0$ with probability $1/\mu^{(1)}$.

b. The system is in state 2 at $t = 0$ and the probability that the first type 2 event occurs at $t = n$ is $b_n^{(2)} = m_n^{(2)}/\mu^{(2)}$ (type 2 eq.).

c. The state of the system is not known at $t = 0$, but with probability $q^{(1)} = \mu^{(1)}/(\mu^{(1)} + \mu^{(2)})$ one has a type 1 equilibrium process and with probability $q^{(2)} = \mu^{(2)}/(\mu^{(1)} + \mu^{(2)})$ one has a type 2 equilibrium process. This process is called a pure equilibrium process.

Suppose that state 1 corresponds to a wet day, then the time origin is set at:

a. an arbitrary wet day for a type 1 equilibrium process,
b. an arbitrary dry day for a type 2 equilibrium process,
c. an arbitrary day for a pure equilibrium process.

For a type i equilibrium process one can define:

(2.42) $$v_n^{(ij)} = P\{\varepsilon^{(j)} \text{ at } t = n \mid \text{type } i \text{ eq.}\} \qquad i=1,2;\, j=1,2;\, n=0,1,\ldots$$

By definition one has $v_0^{(ij)} = \delta_{ij}/\mu^{(i)}$.

The events $\varepsilon^{(i)}$ in a type i equilibrium alternating renewal process form a delayed renewal process in which the distribution of the forward recurrence time is $\{b_n^{(i)}\}$ and the distribution of the recurrence times is the convolution of $\{f_n^{(1)}\}$ and $\{f_n^{(2)}\}$. The pgf of the last probability distribution is $F^{(1)}(s)\,F^{(2)}(s)$ (see Appendix A1) and for the pgf $B^{(i)}(s)$ of $\{b_n^{(i)}\}$ a similar relation as (2.31) holds, namely:

(2.43) $$B^{(i)}(s) = \frac{M^{(i)}(s)}{\mu^{(i)}} = \frac{1-F^{(i)}(s)}{\mu^{(i)}(1-s)} \qquad i = 1, 2.$$

From (2.28) one gets for the gf of $\{v_n^{(ii)}\}$

(2.44a) $$V^{(ii)}(s) = \frac{1-F^{(i)}(s)}{\mu^{(i)}(1-s)\{1-F^{(1)}(s)F^{(2)}(s)\}}\,.$$

Assume $i \neq j$, then the events $\varepsilon^{(j)}$ in a type i equilibrium process form a delayed renewal process in which the pgf of the forward recurrence time is $B^{(i)}(s)F^{(j)}(s)$ and the pgf of the recurrence times is $F^{(1)}(s)F^{(2)}(s)$. From (2.28) and (2.43) one gets for the gf of $v_n^{(ij)}$:

$$\text{(2.44b)} \qquad V^{(ij)}(s) = \frac{\{1-F^{(i)}(s)\}F^{(j)}(s)}{\mu^{(i)}(1-s)\{1-F^{(1)}(s)F^{(2)}(s)\}} \qquad \mathrm{i} \neq j$$

which is the discrete analogue of Equation (7.4.3) of Cox (1962).

Now, assume a pure equilibrium process and define:

$$\text{(2.45)} \qquad v_n^{(i)} = P\{\varepsilon^{(i)} \text{ at } t=n\} \qquad i=1,2; n=0,1,\ldots$$

For the gf $V^{(i)}(s)$ of $\{v_n^{(i)}\}$ holds:

$$\text{(2.46)} \qquad V^{(i)}(s) = q^{(i)}V^{(ii)}(s) + q^{(j)}V^{(ji)}(s) \qquad i \neq j.$$

Substitution of (2.44) results in:

$$\text{(2.47)} \qquad V^{(i)}(s) = \frac{1}{\mu^{(1)}+\mu^{(2)}}\,\frac{1}{1-s}$$

and thus for $v_n^{(i)}$ holds:

$$\text{(2.48)} \qquad v_n^{(i)} = 1/(\mu^{(1)}+\mu^{(2)}) \qquad i=1,2; n=0,1,\ldots$$

as it should be for a pure equilibrium process.

The probability $v_n^{(ii)}$ can be written as the sum of probabilities of mutually exclusive events:

$$\begin{aligned}
\text{(2.49a)} \qquad v_n^{(ii)} = {} & P\{\text{no } \varepsilon^{(i)} \text{ at } [0,n-1], \varepsilon^{(i)} \text{ at } t=n \mid \text{type } i \text{ eq.}\} + \\
& + \sum_{k=0}^{n-2} P\{\varepsilon^{(j)} \text{ at } t=k, \text{ no } \varepsilon^{(i)} \text{ at } [k+1,n-1], \varepsilon^{(i)} \\
& \text{at } t=n \mid \text{type } i \text{ eq.}\} + \\
& + P\{\varepsilon^{(j)} \text{ at } t=n-1, \varepsilon^{(i)} \text{ at } t=n \mid \text{type } i \text{ eq.}\} = \\
= {} & b_n^{(i)} + \sum_{k=0}^{n-2} v_k^{(ij)} f_{n-k}^{(i)} + v_{n-1}^{(ij)} f_1^{(i)} = \\
= {} & b_n^{(i)} + \sum_{k=0}^{n-1} v_k^{(ij)} f_{n-k}^{(i)} \qquad i \neq j, n=1,2,\ldots
\end{aligned}$$

A similar argument gives:

$$\text{(2.49b)} \qquad v_n^{(ij)} = \sum_{k=0}^{n-1} v_k^{(ii)} f_{n-k}^{(j)} \qquad i \neq j, n=1,2,\ldots$$

In the argument of (2.49) the last recurrent event before $t = n$ was taken into account. One can also pay special attention to the first recurrent event, which was done in the derivation of (2.27). This leads to a system of equations differing from (2.49), namely:

$$\text{(2.50a)} \qquad v_n^{(ii)} = \sum_{k=0}^{n} b_k^{(i)} u_{n-k}^{(ii)} \qquad n = 0, 1, \ldots$$

$$\text{(2.50b)} \qquad v_n^{(ij)} = \sum_{k=0}^{n} b_k^{(i)} u_{n-k}^{(ij)} \qquad i \neq j, n = 0, 1, \ldots$$

The relations (2.49) have the advantage that calculation of the $\{u_n^{(ij)}\}$ sequence is not necessary.

The following probabilities are important for the derivation of correlograms and variance-time curves:

$$\text{(2.51)} \qquad h_n^{(ij)} = P\{\text{state } j \text{ at } t = n \,|\, \text{type } i \text{ eq.}\} \qquad i = 1, 2; j = 1, 2; n = 0, 1, \ldots$$

By definition one has $h_0^{(ij)} = \delta_{ij}$.

Expressions for the probabilities $h_n^{(ij)}$ can be obtained in the same manner as those for the probabilities $w_n^{(ij)}$. A derivation analogous to that of (2.38) results in:

$$\text{(2.52a)} \qquad h_n^{(ij)} = \sum_{k=0}^{n} b_k^{(i)} w_{n-k}^{(ij)} \qquad i \neq j, n = 0, 1, \ldots$$

$$\text{(2.52b)} \qquad h_n^{(ii)} = \sum_{k=n+1}^{\infty} b_k^{(i)} + \sum_{k=0}^{n} b_k^{(i)} w_{n-k}^{(ii)} \qquad n = 0, 1, \ldots$$

It is not necessary to use both relations, because:

$$\text{(2.53)} \qquad h_n^{(ii)} + h_n^{(ij)} = 1 \qquad i \neq j, n = 0, 1, \ldots$$

Analogous to (2.41) one has:

$$\text{(2.54a)} \qquad h_n^{(ij)} = \sum_{k=0}^{n-1} v_k^{(ii)} - \sum_{k=0}^{n-1} v_k^{(ij)} \qquad i \neq j, n = 1, 2, \ldots$$

$$\text{(2.54b)} \qquad h_n^{(ii)} = 1 + \sum_{k=0}^{n-1} v_k^{(ij)} - \sum_{k=0}^{n-1} v_k^{(ii)} \qquad i \neq j, n = 1, 2, \ldots$$

From (2.54a) one obtains for the gf $H^{(ij)}(s)$ of $\{h_n^{(ij)}\}$:

$$(2.55) \qquad H^{(ij)}(s) = \sum_{n=1}^{\infty} h_n^{(ij)} s^n = \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} (v_k^{(ii)} - v_k^{(ij)}) s^n =$$

$$= \sum_{k=0}^{\infty} \sum_{n=k+1}^{\infty} (v_k^{(ii)} - v_k^{(ij)}) s^n =$$

$$= \left\{ \sum_{k=0}^{\infty} (v_k^{(ii)} - v_k^{(ij)}) s^k \right\} \times \left\{ \sum_{n=k+1}^{\infty} s^{n-k} \right\} =$$

$$= s\{V^{(ii)}(s) - V^{(ij)}(s)\}/(1-s) \qquad i \neq j.$$

Finally, substitution of (2.44) results in:

$$(2.56a) \qquad H^{(ij)}(s) = \frac{s\{1-F^{(i)}(s)\}\{1-F^{(j)}(s)\}}{\mu^{(i)}(1-s)^2\{1-F^{(i)}(s)F^{(j)}(s)\}} \qquad i \neq j.$$

For the gf $H^{(ii)}(s)$ of $\{h_n^{(ii)}\}$ it follows from (2.53):

$$(2.56b) \qquad H^{(ii)}(s) = \frac{1}{1-s} - H^{(ij)}(s) =$$

$$= \frac{1}{1-s} - \frac{s\{1-F^{(i)}(s)\}\{1-F^{(j)}(s)\}}{\mu^{(i)}(1-s)^2\{1-F^{(i)}(s)F^{(j)}(s)\}} \qquad i \neq j$$

which is the discrete analogue of Equation (7.4.4) of Cox (1962).

From (2.56a) it follows:

$$(2.57) \qquad \mu^{(i)} H^{(ij)}(s) = \mu^{(j)} H^{(ji)}(s).$$

This equation is obtained because a pure equilibrium alternating renewal process is reversible in time. For a stationary time-reversible process:

$$(2.58) \qquad P\{\text{state } i \text{ at } t = 0, \text{ state } j \text{ and } t = n\} =$$

$$= P\{\text{state } j \text{ at } t = 0, \text{ state } i \text{ at } t = n\} \qquad \text{for all } n.$$

Since for a pure equilibrium alternating renewal process:

$$(2.59) \qquad P\{\text{state } i \text{ at } t = 0, \text{ state } j \text{ at } t = n\} =$$

$$= P\{\text{state } j \text{ at } t = n \mid \text{type } i \text{ eq.}\} P\{\text{type } i \text{ eq.}\} =$$

$$= h_n^{(ij)} \mu^{(i)}/(\mu^{(i)} + \mu^{(j)}) \qquad i \neq j$$

expression (2.58) leads to (2.57).

Let $h_n^{(i)}$ be the probability that the system is in state i at $t = n$. For a pure equilibrium process one obtains for the gf $H^{(i)}(s)$ of $\{h_n^{(i)}\}$:

$$(2.60) \qquad H^{(i)}(s) = q^{(i)} H^{(ii)}(s) + q^{(j)} H^{(ji)}(s) \qquad i \neq j.$$

Substitution of (2.56) gives:

$$(2.61) \qquad H^{(i)}(s) = q^{(i)}/(1-s)$$

and thus is $h_n^{(i)} = q^{(i)}$ for all n, as was to be expected for a pure equilibrium process.

Assume that for a wet-dry process state 1 corresponds to a wet day and state 2 to a dry day. If the occurrence of a wet day is a recurrent event, then a type 1 equilibrium process is an ordinary renewal process (starting with a wet day at $t = 0$). From (2.6), (2.7) and (2.8) one obtains for the terms on the right side of (2.56b), with $i = 1$ and $j = 2$:

$$\text{(2.62a)} \qquad \{1-F^{(1)}(s)\}\{1-F^{(2)}(s)\} = \{1-F^{(w)}(s)\}\{1-F^{(d)}(s)\} = \frac{\{s-F(s)\}(1-s)}{m_1 s(1-f_1 s)}$$

$$\text{(2.62b)} \qquad 1-F^{(1)}(s)F^{(2)}(s) = 1-F^{(w)}(s)F^{(d)}(s) = \frac{1-F(s)}{1-f_1 s}$$

$$\text{(2.62c)} \qquad \mu^{(1)} = \mu^{(w)} = 1/m_1 .$$

So for this alternating renewal process one obtains:

$$\text{(2.63)} \qquad H^{(11)}(s) = H^{(ww)}(s) = 1/\{1-F(s)\}.$$

Thus (2.56b) reduces to (2.25) and one gets:

$$\text{(2.64)} \qquad u_n = h_n{}^{(11)} = h_n{}^{(ww)}$$

which was to be expected since both u_n and $h_n{}^{(ww)}$ denote the probability of a day being wet at $t = n$, given that a wet day occurs at $t = 0$.

3. The cumulative distribution function of k-day totals

In III,7 the cdf of k-day totals was obtained by numerical methods for iid rainfall amounts on wet days (model I). For this model the number of wet days in a k-day period is denoted by $S_n(k)$ and the k-day rainfall total by $S_x(k)$. Let $\underline{z}_i$ be the rainfall amount on the ith wet day. The carrier of $\underline{z}_i$ is $[\tilde{\delta},\infty)$ where $\tilde{\delta}$ is the height of the threshold minus half the unit of measurement (see III, 5.2). One has:

$$\text{(3.1)} \qquad S_x(k) = \begin{cases} 0 & \text{if } S_n(k) = 0 \\ \sum\limits_{i=1}^{S_n(k)} z_i & \text{if } S_n(k) > 0. \end{cases}$$

Thus is:

(3.2) $$P\{\underline{S}_x(k) \leqslant s\} = \sum_{j=0}^{k} P\{\underline{S}_x(k) \leqslant s \mid \underline{S}_n(k) = j\} P\{\underline{S}_n(k) = j\} =$$

$$= \begin{cases} P\{\underline{S}_n(k) = 0\} & \text{if } s < \tilde{\delta} \\ P\{\underline{S}_n(k) = 0\} + \sum\limits_{j=1}^{k} P\left\{\sum\limits_{i=1}^{j} \underline{z}_i \leqslant s\right\} P\{\underline{S}_n(k) = j\} & \text{if } s \geqslant \tilde{\delta}. \end{cases}$$

If $\tilde{\underline{z}}_i$ denotes the shifted rainfall amount on the ith wet day ($= \underline{z}_i - \tilde{\delta}$), then:

(3.3) $$P\left\{\sum_{i=1}^{j} \underline{z}_i \leqslant s\right\} = \begin{cases} 0 & \text{if } s < j\tilde{\delta} \\ P\left\{\sum\limits_{i=1}^{j} \tilde{\underline{z}}_i \leqslant s - j\tilde{\delta}\right\} & \text{if } s \geqslant j\tilde{\delta}. \end{cases}$$

In the previous chapter it was assumed that $\tilde{\underline{z}}_i$ was gamma distributed with shape parameter ν and scale parameter λ. Then $\sum\limits_{i=1}^{j} \tilde{\underline{z}}_i$ is gamma distributed with shape parameter $\tilde{\nu} = j\nu$ and scale parameter λ, and the probabilities on the right side of (3.3) follow from a numerical evaluation of the incomplete gamma function.

The method of evaluation of the incomplete gamma function depends on $\tilde{\nu}$ and $\tilde{s} = \lambda(s - j\tilde{\delta})$. In III, 7.2 use was made of an asymptotic expansion (Equation 6.5.32 of ABRAMOWITZ and STEGUN (1970)) of the incomplete gamma function when $\tilde{s} > \tilde{\nu}$ and $\tilde{s} > 15$. In other cases a series expansion of the incomplete gamma function was used (Equation 6.5.29, non-alternating version of ABRAMOWITZ and STEGUN (1970)).

The calculation of the cdf of $\underline{S}_x(k)$ with Equation (3.2) also requires knowledge of the probability distribution of $\underline{S}_n(k)$. There are some special cases of the wet-dry process for which this probability distribution is well known:

a. The Bernoulli process (see 2.1.1). For this process $\underline{S}_n(k)$ has a binomial distribution. Calculated cdfs of k-day totals under the assumption of a Bernoulli wet-dry process were given by QUÉLENNEC (1973).

b. The two-state first order Markov chain (see 2.1.1). An expression for the probability distribution of $\underline{S}_n(k)$ was given by GABRIEL (1959).

A method for the calculation of the probability distribution of $\underline{S}_n(k)$ for a renewal process was given by ELLIOT (1965). His method is applicable to both ordinary and delayed renewal processes and is discussed in Section 3.1. In Section 3.2 Elliot's method is extended for application to the alternating renewal process.

The calculation of the cdf of k-day totals becomes very complicated when the rainfall amounts on wet days are not iid (models II, III and IV). For these models the cdfs of k-day totals were based on Monte Carlo simulations (see III, 7.2).

3.1. *The distribution of the number of events in a sequence of length n for a renewal process*

For a renewal process one can define:

$$R(m,n) = P\{\varepsilon_m \text{ at } [0,n-1]\,|\,\varepsilon \text{ at } t = 0\} \qquad n \geqslant 1 \tag{3.4}$$

where ε_m stands for 'the recurrent event ε occurs m times'. Thus $R(m,n)$ is the probability of m events in a sequence of length n, starting at $t = 0$, for an ordinary renewal process. Some special cases are:

$$R(m,n) = 0 \qquad \text{if } m > n \tag{3.5a}$$

$$R(0,n) = 0 \tag{3.5b}$$

$$R(1,n) = m_{n-1} \text{ (and thus is } R(1,1) = 1) \tag{3.5c}$$

$$R(n,n) = f_1^{n-1}. \tag{3.5d}$$

For $2 \leqslant m \leqslant n$ and $n \geqslant 2$ holds:

$$\begin{aligned} R(m,n) = {} & P\{\varepsilon \text{ at } t = 1, \varepsilon_{m-1} \text{ at } [1,n-1]\,|\,\varepsilon \text{ at } t=0\} + \\ & + \sum_{k=2}^{n-m+1} P\{\text{no } \varepsilon \text{ at } [1,k-1], \varepsilon \text{ at } t=k, \varepsilon_{m-1} \text{ at} \\ & [k,n-1]\,|\,\varepsilon \text{ at } t = 0\} = \\ = {} & P\{\varepsilon \text{ at } t=1\,|\,\varepsilon \text{ at } t=0\}P\{\varepsilon_{m-1} \text{ at } [1,n-1]\,|\,\varepsilon \text{ at } t=1\} + \\ & + \sum_{k=2}^{n-m+1} P\{\text{no } \varepsilon \text{ at } [1, k-1], \varepsilon \text{ at } t=k\,|\,\varepsilon \text{ at } t=0\} \times \\ & \times P\{\varepsilon_{m-1} \text{ at } [k,n-1]\,|\,\varepsilon \text{ at } t=k\} \end{aligned} \tag{3.6}$$

since the process becomes independent of its history at $t = k$. From (2.1) and (3.4), relation (3.6) results in:

$$\begin{aligned} R(m,n) & = f_1 R(m-1,n-1) + \sum_{k=2}^{n-m+1} f_k R(m-1,n-k) = \\ & = \sum_{k=1}^{n-m+1} f_k R(m-1,n-k) \qquad 2 \leqslant m \leqslant n;\ n \geqslant 2. \end{aligned} \tag{3.7}$$

Let $Q(m,n)$ be the probability of m events in a sequence of length n for an equilibrium renewal process. Some special cases are:

$$(3.8a)\qquad Q(m,n) = 0 \qquad\qquad \text{if } m > n$$

$$(3.8b)\qquad Q(0,n) = \sum_{k=n}^{\infty} b_k$$

$$(3.8c)\qquad Q(n,n) = \frac{1}{\mu} f_1^{n-1} \text{ (and thus } Q(1,1) = \frac{1}{\mu}\text{)}$$

where b_k is given by (2.32).

The probabilities $Q(m,n)$ can be written as:

$$\begin{aligned}(3.9)\qquad Q(m,n) &= P\{\varepsilon \text{ at } t = 0, \varepsilon_{m-1} \text{ at } [1,n-1] \mid \text{equilibrium process}\} + \\ &\quad + \sum_{k=1}^{n-m} P\{\text{no } \varepsilon \text{ at } [0,k-1], \varepsilon \text{ at } t = k, \varepsilon_m \text{ at } \\ &\quad [k,n-1] \mid \text{equilibrium process}\} = \\ &= b_0 R(m,n) + \sum_{k=1}^{n-m} b_k R(m,n-k) = \\ &= \sum_{k=0}^{n-m} b_k R(m,n-k) \qquad\qquad 1 \leqslant m \leqslant n\end{aligned}$$

since the past is irrelevant at $t = k$. Substitution of (2.32) results in:

$$(3.10)\qquad Q(m,n) = \frac{1}{\mu} \sum_{k=0}^{n-m} m_k R(m,n-k) \qquad\qquad 1 \leqslant m \leqslant n.$$

This relation can be modified as follows:

$$(3.11)\qquad Q(m,n) = \frac{1}{\mu} \{ \sum_{k=0}^{n-m} R(m,n-k) - \sum_{k=0}^{n-m} (1-m_k) R(m,n-k)\} \qquad 1 \leqslant m \leqslant n.$$

For the second summation within brackets one can write:

$$(3.12)\qquad \sum_{k=0}^{n-m} (1-m_k) R(m,n-k) = \sum_{k=1}^{n-m} \sum_{j=1}^{k} f_j R(m,n-k) \qquad 1 \leqslant m \leqslant n.$$

The area of summation is given in Figure 3.1. The contribution to (3.12) of the n–m points on the line $k = j$ is $\sum_{j=1}^{n-m} f_j R(m,n-j)$; the n–m–1 points on the

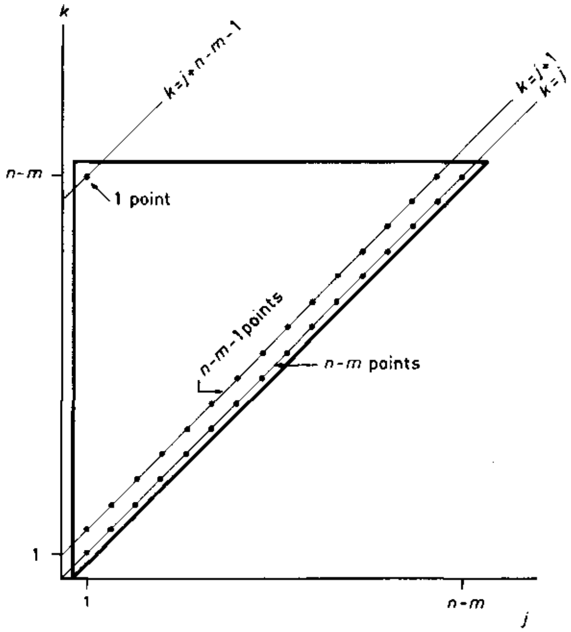


FIG. 3.1. Area of summation in Equation (3.12).

line $k=j+1$ give a contribution of $\sum\limits_{j=1}^{n-m-1} f_j R(m,\, n\text{–}j\text{–}1)$ and so on. So (3.12) results in:

$$(3.13) \qquad \sum_{k=0}^{n-m} (1\text{–}m_k) R(m,n\text{–}k) = \sum_{j=1}^{n-m} f_j R(m,n\text{–}j) +$$

$$+ \sum_{j=1}^{n-m-1} f_j R(m,n\text{–}j\text{–}1) + \ldots + f_1 R(m,m) \qquad 1 \leqslant m \leqslant n.$$

Application of (3.7) for each term separately gives:

$$(3.14) \qquad \sum_{k=0}^{n-m} (1\text{–}m_k) R(m,n\text{–}k) = R(m+1,n) + R(m+1,n\text{–}1) + \ldots +$$

$$+ R(m+1,m+1) \qquad 1 \leqslant m \leqslant n$$

and thus (3.11) becomes:

$$(3.15) \qquad Q(m,n) = \frac{1}{\mu} \left\{ \sum_{k=m}^{n} R(m,k) - \sum_{k=m+1}^{n} R(m+1,k) \right\} \quad 1 \leqslant m \leqslant n$$

which is the discrete analogue of Equation (3.2.9) of Cox (1962).

The term $\sum_{k=m}^{n} R(m,k)/\mu$ represents the probability that there are at least m events at $[0,n-1]$; $\sum_{k=m+1}^{n} R(m+1,k)/\mu$ stands for the probability of at least $m+1$ events at $[0,n-1]$.

For computer calculations, (3.15) is a bit faster than (3.10) because it does not contain multiplications.

From the relations given above it is possible to calculate the probability distribution of the number of wet days in a k-day period in the following situations:

a. Wet intervals have a geometric distribution. Then the occurrence of a wet day is a recurrent event and the required probability is $P\{S_n(k)=j\} =$ $= Q(j,k)$.

b. Only dry intervals have a geometric distribution. Here the occurrence of a dry day is a recurrent event and the required probability is $P\{S_n(k)=j\} =$ $= Q(k-j,k)$.

If both wet and dry intervals are not geometrically distributed, it is not possible to calculate the probability distribution of the number of wet days with the formulas given in this section. An extension of the theory for this purpose is given in the next section.

3.2. *The distribution of the number of times that the process is in a certain state in a sequence of length n for an alternating renewal process*

For an alternating renewal process one can define:

(3.16) $R^{(ij)}(m,n) = P\{S_m^{(j)} \text{ at } [0,n-1] \mid \varepsilon^{(i)} \text{ at } t=0\}$ $i=1,2;\ j=1,2;\ n=1,2,\ldots$

where $S_m^{(j)}$ stands for 'state j occurs m times'. For $n=1$ holds:

(3.17a) $R^{(ij)}(0,1) = 1-\delta_{ij}$

(3.17b) $R^{(ij)}(1,1) = \delta_{ij}$.

If $i \neq j$, then:

(3.18) $R^{(ij)}(m,n) = R^{(ii)}(n-m,n)$

and for $n > 1$ one has the following special cases:

(3.19a) $R^{(ij)}(m,n) = R^{(ii)}(m,n) = 0$ if $m > n$

(3.19b) $R^{(ij)}(0,n) = R^{(ii)}(n,n) = 0$

(3.19c) $R^{(ij)}(1,n) = R^{(ii)}(n-1,n) = \begin{cases} 1 & \text{if } n = 2 \\ f_1^{(j)} m_{n-2}^{(i)} & \text{if } n > 2 \end{cases}$

(3.19d) $$R^{(ij)}(n,n) = R^{(ii)}(0,n) = 0$$

(3.19e) $$R^{(ij)}(n-1,n) = R^{(ii)}(1,n) = m^{(j)}_{n-2}.$$

For $2 \leqslant m \leqslant n-1$; $n \geqslant 3$ and $i \neq j$ holds:

(3.20) $$R^{(ii)}(m,n) = \sum_{k=1}^{n-m} P\{\text{no } \varepsilon^{(j)} \text{ at } [0,k-1],\ \varepsilon^{(j)} \text{ at } t=k,\ S^{(i)}_{m-1}$$

$$\text{at } [k,n-1] \mid \varepsilon^{(i)} \text{ at } t=0\} =$$

$$= \sum_{k=1}^{n-m} P\{\text{no } \varepsilon^{(j)} \text{ at } [0,\ k-1],\ \varepsilon^{(j)} \text{ at } t=k \mid \varepsilon^{(i)} \text{ at } t=0\} \times$$

$$\times P\{S^{(i)}_{m-1} \text{ at } [k,n-1] \mid \varepsilon^{(j)} \text{ at } t=k\}$$

since the process becomes independent of its history at $t=k$. Hence:

(3.21) $$R^{(ii)}(m,n) = \sum_{k=1}^{n-m} f^{(j)}_k R^{(ji)}(m-1,n-k) =$$

$$= \sum_{k=1}^{n-m} f^{(j)}_k R^{(jj)}(n-k-m+1,n-k)\ i \neq j; 2 \leqslant m \leqslant n-1; n \geqslant 3.$$

From (3.18) and (3.21) it follows:

(3.22) $$R^{(ij)}(m,n) = \sum_{k=1}^{m} f^{(j)}_k R^{(ji)}(n-m-1,n-k) = \sum_{k=1}^{m} f^{(j)}_k R^{(jj)}(m-k+1,n-k)$$

$$i \neq j; 2 \leqslant m \leqslant n-1; n \geqslant 3.$$

To make the reader more familiar with this material some formulas are worked out for a Bernoulli wet-dry process. It is assumed that a wet day corresponds to state 1 and that the probability of a wet day equals p. For this process the probability distributions of dry and wet spells are given by (2.14) and (2.15). Further, a type 1 event occurs at $t=0$ if a wet day occurs at $t=0$ and a dry day at $t=1$.

For $n \geqslant 2$ the number of events does not depend on the initial condition, since the process has no memory. Thus:

(3.23) $$R^{(11)}(m,n) = R^{(21)}(m,n) = R^{(12)}(n-m,n) = R^{(22)}(n-m,n) \quad n \geqslant 2.$$

By definition:

(3.24) $$R^{(11)}(m,n) = P\{m \text{ wet days at } [0,n-1] \mid \text{wet at } t=0, \text{ dry at } t=1\} = \\ = P\{m-1 \text{ wet days at } [2,n-1]\} = \\ = \binom{n-2}{m-1} p^{m-1} q^{n-m-1} \qquad 1 \leqslant m \leqslant n-1; n \geqslant 3.$$

The relation (3.21) has the following form:

(3.25) $$R^{(11)}(m,n) = \sum_{k=1}^{n-m} f_k^{(2)} R^{(21)}(m-1,n-k) = \\ = \sum_{k=1}^{n-m} pq^{k-1} \binom{n-k-2}{m-2} p^{m-2} q^{n-k-m} = \\ = p^{m-1} q^{n-m-1} \sum_{k=1}^{n-m} \binom{n-k-2}{m-2} = \\ = \binom{n-2}{m-1} p^{m-1} q^{n-m-1} \qquad 2 \leqslant m \leqslant n-1; n \geqslant 3.$$

The last equality follows from II, (12.6) or II, (12.8) of FELLER (1968).

Let

(3.26) $$Q^{(ij)}(m,n) = P\{S_m^{(j)} \text{ at } [0,n-1] \mid \text{type } i \text{ eq.}\} \quad i=1,2; j=1,2; n=1,2,\ldots$$

that is $Q^{(ij)}(m,n)$ denotes the probability that state j occurs m times in a sequence of length n, starting at $t=0$, for a type i equilibrium alternating renewal process. For $n=1$ holds:

(3.27a) $$Q^{(ij)}(0,1) = 1-\delta_{ij}$$

(3.27b) $$Q^{(ij)}(1,1) = \delta_{ij}.$$

If $i \neq j$, then:

(3.28) $$Q^{(ij)}(m,n) = Q^{(ii)}(n-m,n)$$

and for $n > 1$ one has the following special cases:

(3.29a) $$Q^{(ij)}(m,n) = Q^{(ii)}(m,n) = 0 \qquad m > n$$

(3.29b) $$Q^{(ij)}(0,n) = Q^{(ii)}(n,n) = \sum_{k=n-1}^{\infty} b_k^{(i)}$$

(3.29c) $$Q^{(ij)}(n,n) = Q^{(ii)}(0,n) = 0$$

(3.29d) $$Q^{(ij)}(n-1,n) = Q^{(ii)}(1,n) = m^{(j)}_{n-2}/\mu^{(i)}.$$

For $1 \leqslant m \leqslant n-1$, $n \geqslant 2$ and $i \neq j$ holds:

(3.30) $$\begin{aligned} Q^{(ij)}(m,n) &= P\{\varepsilon^{(i)} \text{ at } t=0,\ S^{(j)}_m \text{ at } [0,n-1] \mid \text{type } i \text{ eq.}\} + \\ &\quad + \sum_{k=1}^{n-m-1} P\{\text{no } \varepsilon^{(i)} \text{ at } [0,k-1],\ \varepsilon^{(i)} \text{ at } t=k,\ S^{(j)}_m \\ &\quad \text{at } [k,n-1] \mid \text{type } i \text{ eq.}\} = \\ &= b_0{}^{(i)}\, R^{(ij)}(m,n) + \sum_{k=1}^{n-m-1} b_k^{(i)} R^{(ij)}(m,n-k) \end{aligned}$$

since the process becomes independent of its history at $t=k$. For a type i equilibrium process $b_k{}^{(i)} = m_k{}^{(i)}/\mu^{(i)}$ (see Section 2.2) and therefore (3.30) results in:

(3.31) $$Q^{(ij)}(m,n) = \left\{ \sum_{k=0}^{n-m-1} m^{(i)}_k R^{(ij)}(m,n-k) \right\} / \mu^{(i)} \quad i \neq j;\ 1 \leqslant m \leqslant n-1;\ n \geqslant 2.$$

From (3.18), (3.28) and (3.31) it follows:

(3.32) $$\begin{aligned} Q^{(ii)}(m,n) &= \left\{ \sum_{k=0}^{m-1} m^{(i)}_k R^{(ij)}(n-m,n-k) \right\} / \mu^{(i)} = \\ &= \left\{ \sum_{k=0}^{m-1} m^{(i)}_k R^{(ii)}(m-k,n-k) \right\} / \mu^{(i)} \quad i \neq j; 1 \leqslant m \leqslant n-1; n \geqslant 2. \end{aligned}$$

Let:

(3.33) $$Q^{(i)}(m,n) = P\{S^{(i)}_m \text{ at } [0,n-1]\}$$

then, for a pure equilibrium process holds:

(3.34) $$Q^{(i)}(m,n) = q^{(i)} Q^{(ii)}(m,n) + q^{(j)} Q^{(ji)}(m,n) \qquad i \neq j.$$

For the Bernoulli process, which was considered earlier in this section, a type 1 equilibrium process is a process starting with a wet day and a type 2 equilibrium process is a process starting with a dry day. The means $\mu^{(1)}$ and $\mu^{(2)}$ are $1/q$ and $1/p$ respectively. For the survivor function one has similar relations as (2.16), namely:

(3.35a) $$m^{(1)}_n = p^n$$

$$(3.35b) \qquad m_n^{(2)} = q^n.$$

By definition one has:

$$(3.36) \qquad \begin{aligned} Q^{(11)}(m,n) &= P\{m \text{ wet days at } [0,n-1] \mid \text{wet at } t=0\} = \\ &= P\{m-1 \text{ wet days at } [1,n-1]\} = \\ &= \binom{n-1}{m-1} p^{m-1} q^{n-m} \qquad 1 \leqslant m \leqslant n;\, n \geqslant 2. \end{aligned}$$

The relation (3.32) has the following form:

$$(3.37) \qquad \begin{aligned} Q^{(11)}(m,n) &= \frac{1}{\mu^{(1)}} \sum_{k=0}^{m-1} m_k^{(1)} R^{(11)}(m-k,n-k) = \\ &= q \sum_{k=0}^{m-1} p^k \binom{n-k-2}{m-k-1} p^{m-k-1} q^{n-m-1} = \\ &= p^{m-1} q^{n-m} \sum_{k=0}^{m-1} \binom{n-k-2}{m-k-1} = \\ &= p^{m-1} q^{n-m} \sum_{k=0}^{m-1} \binom{n-k-2}{n-m-1} = \\ &= \binom{n-1}{m-1} p^{m-1} q^{n-m} \qquad 1 \leqslant m \leqslant n-1;\, n \geqslant 2. \end{aligned}$$

The last equality follows from II, (12.6) and II, (12.8) of FELLER (1968).

Since the process has no memory:

$$(3.38) \qquad \begin{aligned} Q^{(11)}(m,n) &= Q^{(21)}(m-1,n) = Q^{(12)}(n-m,n) = \\ &= Q^{(22)}(n-m+1,n) \qquad 1 \leqslant m \leqslant n. \end{aligned}$$

From (3.37) and (3.38) it follows:

$$(3.39) \qquad Q^{(21)}(m,n) = \binom{n-1}{m} p^m q^{n-m-1} \qquad 1 \leqslant m \leqslant n-1;\, n \geqslant 2$$

which is the probability of m wet days in a period of length n–1.

Since $q^{(1)} = P(\text{wet}) = p$ and $q^{(2)} = P(\text{dry}) = q$, substitution of (3.37) and (3.39) in (3.34) gives:

$$(3.40) \qquad \begin{aligned} Q^{(1)}(m,n) &= p \binom{n-1}{m-1} p^{m-1} q^{n-m} + q \binom{n-1}{m} p^m q^{n-m-1} = \\ &= \left\{ \binom{n-1}{m-1} + \binom{n-1}{m} \right\} p^m q^{n-m} = \\ &= \binom{n}{m} p^m q^{n-m} \qquad 1 \leqslant m \leqslant n-1;\, n \geqslant 2. \end{aligned}$$

It can easily be shown that (3.40) also holds for $m=0$ or $m=n$ and $n=1$.

Though the probabilities $Q^{(ij)}(m,n)$ can be obtained from the relations (3.31) and (3.32) it is more convenient, however, to use some alternative expressions for these probabilities. The relation (3.31) can be written as:

$$(3.41) \qquad Q^{(ij)}(m,n) = \left\{ \sum_{k=0}^{n-m-1} R^{(ij)}(m,n\text{-}k) + \right.$$

$$\left. - \sum_{k=0}^{n-m-1} (1\text{-}m_k{}^{(i)})\, R^{(ij)}(m,n\text{-}k) \right\} / \mu^{(i)}$$

$$i \neq j;\ 1 \leqslant m \leqslant n-1;\ n \geqslant 2.$$

Analogous to (3.14) one has:

$$(3.42) \qquad \sum_{k=0}^{n-m-1} (1\text{-}m_k{}^{(i)}) R^{(ij)}(m,n\text{-}k) = \sum_{k=m+2}^{n} R^{(jj)}(m+1,k)$$

and so (3.41) becomes:

$$(3.43) \qquad Q^{(ij)}(m,n) = \left\{ \sum_{k=m+1}^{n} R^{(ij)}(m,k) - \sum_{k=m+2}^{n} R^{(jj)}(m+1,k) \right\} / \mu^{(i)}$$

$$i \neq j;\ 1 \leqslant m \leqslant n-1;\ n \geqslant 2.$$

From this relation it follows that

$$(3.44) \qquad Q^{(ji)}(m,n) = \left\{ \sum_{k=m+1}^{n} R^{(ji)}(m,k) - \sum_{k=m+2}^{n} R^{(ii)}(m+1,k) \right\} / \mu^{(j)}$$

$$i \neq j;\ 1 \leqslant m \leqslant n\text{-}1;\ n \geqslant 2.$$

$$(3.45) \qquad Q^{(ii)}(m,n) = Q^{(ij)}(n-m,n) = \left\{ \sum_{k=n-m+1}^{n} R^{(ij)}(n\text{-}m,k) + \right.$$

$$\left. - \sum_{k=n-m+2}^{n} R^{(jj)}(n\text{-}m+1,k) \right\} / \mu^{(i)} \qquad i \neq j; 1 \leqslant m \leqslant n\text{-}1; n \geqslant 2.$$

Substituting (3.44) and (3.45) in (3.34) results in:

$$(3.46)\qquad Q^{(i)}(m,n) = \left\{ \sum_{k=m+1}^{n} R^{(ji)}(m,k) + \sum_{k=n-m+1}^{n} R^{(ij)}(n-m,k) + \right.$$

$$\left. - \sum_{k=m+2}^{n} R^{(ii)}(m+1,k) - \sum_{k=n-m+2}^{n} R^{(jj)}(n-m+1,k) \right\} /$$

$$/ \quad (\mu^{(i)} + \mu^{(j)}) \qquad\qquad i \neq j;\, 1 \leqslant m \leqslant n-1;\, n \geqslant 2.$$

From (3.46) the probabilities $Q^{(i)}(m,n)$ can be obtained by using (3.18), with (3.21) or (3.22).

It can be shown that (3.46) reduces to (3.15) for a renewal process. For instance, assume that the occurrence of state 1 is a recurrent event ε. When a type 2 event occurs at $t=0$, the system is in state 1 at $t=1$ and thus is:

$$(3.47)\qquad R^{(21)}(m,k) = R(m,k-1) \qquad\qquad k \geqslant 2.$$

Further, for $k \geqslant 2$

$$(3.48)\qquad R^{(11)}(m,k) = P\{\varepsilon_m \text{ at } [0,k-1] \mid \varepsilon \text{ at } t=0, \text{ no } \varepsilon \text{ at } t=1\} =$$

$$= \frac{P\{\varepsilon_m \text{ at } [0,k-1], \text{ no } \varepsilon \text{ at } t=1 \mid \varepsilon \text{ at } t=0\}}{1-f_1} =$$

$$= \frac{P\{\varepsilon_m \text{ at } [0,k-1] \mid \varepsilon \text{ at } t=0\} - P\{\varepsilon \text{ at } t=1, \varepsilon_{m-1} \text{ at } [1,k-1] \mid \varepsilon \text{ at } t=0\}}{1-f_1}$$

$$= \frac{R(m,k) - f_1\, R(m-1,k-1)}{1-f_1}.$$

From (3.18) it follows:

$$(3.49)\qquad R^{(12)}(n-m,k) = R^{(11)}(k-n+m,k) = \frac{1}{1-f_1}\{R(k-n+m,k) +$$

$$-f_1 R(k-n+m-1,k-1)\} \qquad\qquad k \geqslant 2$$

$$(3.50)\qquad R^{(22)}(n-m+1,k) = R^{(21)}(k-n+m-1,k) = R(k-n+m-1,k-1) \qquad k \geqslant 2.$$

And thus is:

$$(3.51)\qquad R^{(21)}(m,k) - R^{(11)}(m+1,k) = \frac{1}{1-f_1}\{R(m,k-1) - R(m+1,k)\} \qquad k \geqslant 2$$

$$(3.52)\qquad R^{(12)}(n-m,k) - R^{(22)}(n-m+1,k) = \frac{1}{1-f_1}\{R(k-n+m,k) +$$

$$- R(k-n+m-1,k-1)\} \qquad\qquad k \geqslant 2.$$

Since $\mu^{(1)} = \frac{1}{1-f_1}$ (see 2.1.1) one has:

$$(3.53)\qquad \frac{1}{1-f_1}\,\frac{1}{\mu^{(1)}+\mu^{(2)}} = \frac{\mu^{(1)}}{\mu^{(1)}+\mu^{(2)}} = \frac{1}{\mu}.$$

and thus (3.46) results in:

$$(3.54)\qquad Q^{(1)}(m,n) = \frac{1}{\mu}\sum_{k=m+1}^{n}\{R(m,k-1)-R(m+1,k)\}+\frac{1}{\mu}\sum_{k=n-m+1}^{n}\{R(k-n+m,k)+$$

$$- R(k-n+m-1,k-1)\} \qquad 1 \leqslant m \leqslant n-1; n \geqslant 2.$$

The second sum in (3.54) equals:

$$(3.55)\qquad \sum_{k=n-m+1}^{n} R(k-n+m,k) - \sum_{k=n-m}^{n-1} R(k-n+m,k) = R(m,n) - R(0,n-m) = R(m,n).$$

This result was to be expected since (3.55) represents $Q^{(11)}(m,n)$ (see Equations (3.45) and (3.52)) being the probability that there are m recurrent events in a renewal process starting with a recurrent event at $t = 0$ (ordinary renewal process).

Substituting (3.55) in (3.54) gives:

$$(3.56)\qquad Q^{(1)}(m,n) = \frac{1}{\mu}\left\{\sum_{k=m}^{n} R(m,k) - \sum_{k=m+1}^{n} R(m+1,k)\right\} = Q(m,n)$$

$$1 \leqslant m \leqslant n-1; n \geqslant 2.$$

4. The correlogram

Let $\{\underline{y}_t\}$ be a stationary process in discrete time, then its lag k autocovariance is defined by:

$$(4.1)\qquad C_{yy}(k) = \mathrm{cov}(\underline{y}_t, \underline{y}_{t+k}) = E(\underline{y}_t\underline{y}_{t+k}) - E(\underline{y}_t)E(\underline{y}_{t+k}) =$$

$$= E(\underline{y}_t\underline{y}_{t+k}) - \{E(\underline{y}_t)\}^2.$$

For $k=0$ one has:

$$(4.2)\qquad C_{yy}(0) = E(\underline{y}_t^2) - \{E(\underline{y}_t)\}^2 = \mathrm{var}\ \underline{y}.$$

The lag k serial correlation coefficient is the quotient of $C_{yy}(k)$ and $C_{yy}(0)$ and is therefore readily obtained from the autocovariances.

In the subsequent sections, expressions are derived for the autocovariances of the wet-dry process and of the rainfall models I, II, III and IV. Throughout this section the wet-dry process is assumed to be an alternating renewal process in which state 1 corresponds to a wet day. Since the autocovariance is symmetric in k, non-negative lags are considered only.

4.1. *The autocovariances of the wet-dry process and of the rainfall process with independently and identically distributed rainfall amounts on wet days*

For the wet-dry process $\{\underline{n}_t\}$ holds:

$$(4.3a)\qquad E(\underline{n}_t) = P(\underline{n}_t=1) = q^{(1)}$$

$$(4.3b)\qquad E(\underline{n}_t\underline{n}_{t+k}) = P(\underline{n}_t=1,\underline{n}_{t+k}=1) = q^{(1)}h_k^{(11)} \qquad k \geqslant 0$$

(4.4a) $$C_{nn}(k) = q^{(1)}h_k{}^{(11)} - (q^{(1)})^2 \qquad k \geqslant 0$$

which reduces to:

(4.4b) $$C_{nn}(k) = \frac{1}{\mu}\, u_k - \frac{1}{\mu^2} \qquad k \geqslant 0$$

if the occurrence of a wet day is a recurrent event (cf. Equation (2.64)).

For the rainfall process $\{\underline{x}_t\}$ with iid rainfall amounts (model I) holds:

(4.5) $$\begin{cases} \underline{x}_t = 0 & \text{if the } t\text{th day is dry} \\ \underline{x}_t \simeq \underline{z} & \text{if the } t\text{th day is wet} \end{cases}$$

where $\underline{z}$ is a random variable on the interval $[\delta,\infty)$.

For this rainfall process:

(4.6a) $$E(\underline{x}_t) = E(\underline{x}_t \mid \underline{n}_t = 0)P(\underline{n}_t = 0) + E(\underline{x}_t \mid \underline{n}_t = 1)P(\underline{n}_t = 1) = \\ = 0 \cdot q^{(2)} + E(\underline{z}) \cdot q^{(1)} = q^{(1)}E(\underline{z})$$

and analogously:

(4.6b) $$E(\underline{x}_t\underline{x}_{t+k}) = E(\underline{x}_t\underline{x}_{t+k} \mid \underline{n}_t = 1, \underline{n}_{t+k} = 1)P(\underline{n}_t = 1, \underline{n}_{t+k} = 1) = \\ = \begin{cases} q^{(1)}E(\underline{z}^2) & \text{if } k = 0 \\ q^{(1)}h_k{}^{(11)}\{E(\underline{z})\}^2 & \text{if } k > 0. \end{cases}$$

From (4.1) one obtains for the autocovariance:

(4.7) $$C_{xx}(k) = \begin{cases} q^{(1)}E(\underline{z}^2) - (q^{(1)})^2\{E(\underline{z})\}^2 & \text{if } k = 0 \\ \{E(\underline{z})\}^2\{q^{(1)}h_k{}^{(11)} - (q^{(1)})^2\} = \{E(\underline{z})\}^2 C_{nn}(k) & \text{if } k > 0. \end{cases}$$

The calculation of the autocovariances of the wet-dry process with (4.4) and of the rainfall process with (4.7) only involves the calculation of the probabilities $h_k{}^{(11)}$. These probabilities can be obtained from (2.54b), using (2.49) for the calculation of the probabilities $v_l{}^{(11)}$ and $v_l{}^{(12)}$ ($l = 0, \ldots, k$–1).

From (4.4) and (4.7) it follows that the autocovariances $C_{nn}(k)$ and $C_{xx}(k)$ are monotonically decreasing in k if the probabilities $h_k{}^{(11)}$ are monotonically decreasing in k. The reverse is also true.

For $k = 1$ holds:

(4.8a) $$h_1{}^{(11)} = P\{\text{state 1 at } t = 1 \mid \text{type 1 eq.}\} = 1 - b_0{}^{(1)}$$

and for $k = 2$ holds:

(4.8b) $$h_2{}^{(11)} = P\{\text{state 1 at } t = 1, \text{state 1 at } t = 2 \mid \text{type 1 eq.}\} + \\ + P\{\text{state 2 at } t = 1, \text{state 1 at } t = 2 \mid \text{type 1 eq.}\} = \\ = 1 - b_0^{(1)} - b_1^{(1)} + b_0^{(1)}f_1^{(2)}.$$

The difference of these probabilities is:

$$(4.9) \qquad h_1^{(11)}-h_2^{(11)} = b_1^{(1)}-b_0^{(1)}f_1^{(2)} = (1-f_1^{(1)}-f_1^{(2)})/\mu^{(1)}$$

and thus holds:

$$(4.10) \qquad \left.\begin{array}{l} C_{nn}(2) < C_{nn}(1) \\ \\ C_{xx}(2) < C_{xx}(1) \end{array}\right\} \Leftrightarrow h_1^{(11)}-h_2^{(11)} > 0 \Leftrightarrow f_1^{(1)} + f_1^{(2)} < 1.$$

For most rainfall series the distributions fitted to the length of weather spells are such that the inequalities in (4.10) are satisfied.

4.2. *The autocovariances of the rainfall process with non-identically distributed rainfall amounts on wet days*

In III, 4.2 it was demonstrated that the distribution of the rainfall amount on a particular wet day depends on the adjacent number of dry days. Rainfall amounts on wet days bounded by i wet days ($i=0,1,2$) were called type i amounts. The occurrence of a type i amount is denoted by $\eta^{(i)}$. For the rainfall process $\{\underline{x}_t\}$ with different types of rainfall amounts on wet days holds:

$$(4.11) \qquad \begin{cases} \underline{x}_t = 0 & \text{if the } t\text{th day is dry} \\ \underline{x}_t \simeq \underline{z}^{(i)} & \text{if the } t\text{th day is wet and is bounded by } i \text{ wet days.} \end{cases}$$

The random variable $\underline{z}^{(i)}$ is defined on $[\tilde{\delta},\infty)$. In this section the rainfall amounts on wet days are assumed to be uncorrelated (model III). The process with correlation between successive rainfall amounts (model IV) is discussed in the next section.

For a stationary process the probability that $\eta^{(i)}$ occurs at t does not depend on t. If this probability is denoted by $p^{(i)}$, then

$$(4.12a) \qquad E(\underline{x}_t) = \sum_{i=0}^{2} p^{(i)} E(\underline{z}^{(i)})$$

$$(4.12b) \qquad E(\underline{x}_t^2) = \sum_{i=0}^{2} p^{(i)} E(\underline{z}^{(i)})^2.$$

The variance ($C_{xx}(0)$) is obtained by substituting (4.12) in (4.2). The only problem for the calculation of this quantity is the calculation of the probabilities $p^{(i)}$. These probabilities are obtained as follows. If a wet day occurs at $t=0$ (or another arbitrary time point), one can distinguish 4 different cases (a, b, c and d) depending on the state (wet or dry) of the adjacent days (see Figure 4.1). Let $p^{(a)}$ be the probability of situation a (this is the probability of a type 0 amount); $p^{(b)}$ the probability of situation b and so on. Then:

$$(4.13a) \qquad p^{(a)} = \theta f_1^{(1)}$$

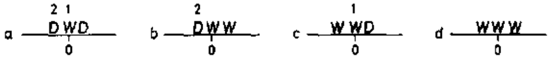


FIG. 4.1. Different possibilities for the occurrence of a wet day (W) at $t = 0$. State 1 corresponds to a wet day. At the end of a weather spell the type of event (1 or 2) is indicated.

(4.13b) $p^{(b)} = \theta m_1^{(1)}$

(4.13c) $p^{(c)} = p^{(b)}$ (since the process is time-reversible)

(4.13d) $p^{(d)} = Q^{(1)}(3,3) = q^{(1)}\,Q^{(11)}(3,3) = q^{(1)}(1-b_0^{(1)}-b_1^{(1)}) =$
$= \theta(\mu^{(1)}-1-m_1^{(1)})$

where θ stands for $1/(\mu^{(1)}+\mu^{(2)})$. This is the probability that $\varepsilon^{(1)}$ (or $\varepsilon^{(2)}$) occurs at an arbitrary time point (see Equation (2.48)). From (4.13) one gets for the probabilities $p^{(i)}$:

(4.14a) $p^{(0)} = p^{(a)} = \theta f_1^{(1)}$

(4.14b) $p^{(1)} = p^{(b)} + p^{(c)} = 2\theta m_1^{(1)}$

(4.14c) $p^{(2)} = p^{(d)} = \theta(\mu^{(1)}-1-m_1^{(1)}) = q^{(1)}-p^{(0)}-p^{(1)}.$

For $k \geqslant 1$ holds:

(4.15) $$E(\underline{x}_t\underline{x}_{t+k}) = \sum_{i=0}^{2}\sum_{j=0}^{2} E(\underline{z}^{(i)})E(\underline{z}^{(j)})p_k^{(ij)}$$

with: $p_k^{(ij)} = P\{\eta^{(i)} \text{ at } t, \eta^{(j)} \text{ at } t+k\} = P\{\eta^{(i)} \text{at } 0, \eta^{(j)} \text{ at } k\}.$

The autocovariances are obtained by substituting (4.15) and (4.12a) in (4.1) and so the problem of the calculation of the autocovariances is reduced to the calculation of the probabilities $p_k^{(ij)}$. Now if a wet day occurs at $t=0$ and at $t=k$ there are different possibilities leading to this event, depending on the states at $t=-1$, 1, k-1 and $k+1$. These possibilities are given in Figure 4.2. From this figure it can be seen that for $k=1$ and $k=2$ not all 16 situations are possible.

Thus the lag 1 and lag 2 autocovariances of a rainfall process with a Bernoulli wet-dry process are in general not equal to zero. For instance, if the probability of a day being wet equals $\frac{1}{2}$, then:

(4.16a) $$E(\underline{x}_t) = \frac{1}{8}\,\{E(\underline{z}^{(0)}) + 2E(\underline{z}^{(1)}) + E(\underline{z}^{(2)})\}$$

since each of the situations a, b, c and d can occur with probability $\frac{1}{8}$. For $k=1$ only the situations bc, bd, dc and dd are possible, each with probability $\frac{1}{16}$ and thus:

(4.16b) $$E(\underline{x}_t\underline{x}_{t+1}) = \frac{1}{16}\,\{[E(\underline{z}^{(1)})]^2 + 2E(\underline{z}^{(1)})E(\underline{z}^{(2)}) + [E(\underline{z}^{(2)})]^2\}$$

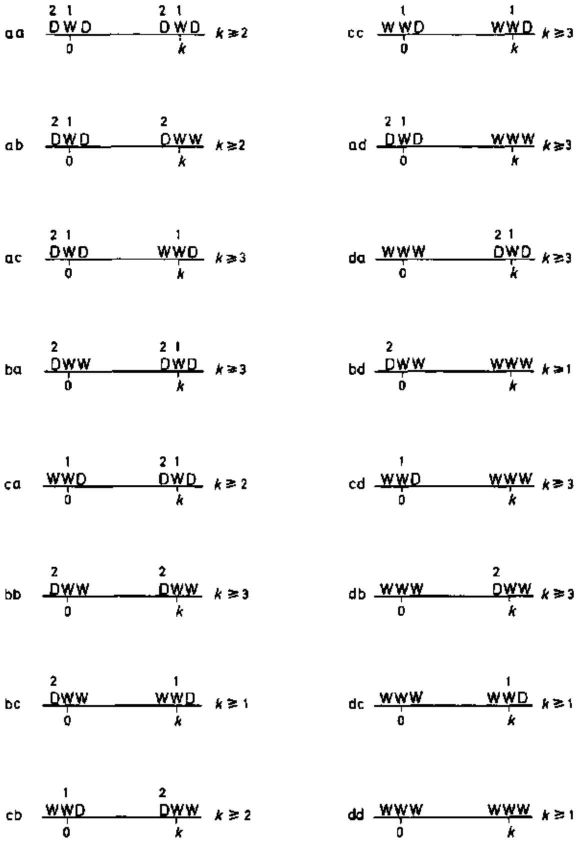


FIG. 4.2. Different possibilities for the occurrence of a wet day (W) at $t = 0$ and at $t = k$. State 1 corresponds to a wet day. At the end of a weather spell the type of event (1 or 2) is indicated.

which does not depend on $E(\underline{z}^{(0)})$.

From (4.1) and (4.16) it follows:

$$(4.17) \qquad C_{xx}(1) = \{3[E(\underline{z}^{(2)})]^2 - [E(\underline{z}^{(0)})]^2 + 4E(\underline{z}^{(1)})E(\underline{z}^{(2)}) + -4E(\underline{z}^{(0)})E(\underline{z}^{(1)}) - 2E(\underline{z}^{(0)})E(\underline{z}^{(2)})\}/64.$$

If $E(\underline{z}^{(0)}) < E(\underline{z}^{(1)}) < E(\underline{z}^{(2)})$, then $C_{xx}(1)$ is positive, since then: $3[E(\underline{z}^{(2)})]^2 > [E(\underline{z}^{(0)})]^2 + 2E(\underline{z}^{(0)})E(\underline{z}^{(2)})$ and $E(\underline{z}^{(1)})E(\underline{z}^{(2)}) > E(\underline{z}^{(0)})E(\underline{z}^{(1)})$.

When dealing with real rainfall processes a model with different types of rainfall amounts (model III) usually has a larger first scc than a model with identically distributed rainfall amounts (model I), see III, Table 6.2.

Let $p_k^{(aa)}$ be the probability that situation a occurs at $t=0$ and $t=k$ $(k \geqslant 1)$. The probabilities of the other situations can be defined analogously. From Figure 4.2 it is seen that:

$$p_k^{(00)} = p_k^{(aa)} \tag{4.18a}$$

$$p_k^{(01)} = p_k^{(ab)} + p_k^{(ac)} \tag{4.18b}$$

$$p_k^{(10)} = p_k^{(ba)} + p_k^{(ca)} \tag{4.18c}$$

$$p_k^{(11)} = p_k^{(bb)} + p_k^{(bc)} + p_k^{(cb)} + p_k^{(cc)} \tag{4.18d}$$

$$p_k^{(02)} = p_k^{(ad)} \tag{4.18e}$$

$$p_k^{(20)} = p_k^{(da)} \tag{4.18f}$$

$$p_k^{(12)} = p_k^{(bd)} + p_k^{(cd)} \tag{4.18g}$$

$$p_k^{(21)} = p_k^{(db)} + p_k^{(dc)} \tag{4.18h}$$

$$p_k^{(22)} = p_k^{(dd)}. \tag{4.18i}$$

For the probabilities $p_k^{(aa)}$; $p_k^{(ab)}$, ... one can derive the following relations:

$$p_k^{(aa)} = \begin{cases} 0 & \text{if } k=1 \\ \theta(f_1{}^{(1)})^2 u_{k-1}^{(12)} & \text{if } k \geqslant 2 \end{cases} \tag{4.19a}$$

$$p_k^{(ab)} = \begin{cases} 0 & \text{if } k=1 \\ \theta f_1^{(1)} u_{k-1}^{(12)}\, m_1{}^{(1)} & \text{if } k \geqslant 2 \end{cases} \tag{4.19b}$$

$$p_k^{(ac)} = \begin{cases} 0 & \text{if } k=1,2 \\ \theta f_1^{(1)} u_k^{(11)} - p_k^{(aa)} & \text{if } k \geqslant 3 \end{cases} \tag{4.19c}$$

$$p_k^{(ba)} = p_k^{(ac)} \tag{4.19d}$$

$$p_k^{(ca)} = p_k^{(ab)} \tag{4.19e}$$

$$p_k^{(bb)} = \begin{cases} 0 & \text{if } k=1,2 \\ \theta u_k^{(22)} m_1{}^{(1)} - p_k^{(ab)} & \text{if } k \geqslant 3 \end{cases} \tag{4.19f}$$

$$p_k^{(bc)} = \begin{cases} \theta f_2^{(1)} & \text{if } k=1 \\ \theta f_3^{(1)} & \text{if } k=2 \\ \theta u_{k+1}^{(21)} - p_k^{(aa)} - p_k^{(ac)} - p_k^{(ba)} & \text{if } k \geqslant 3 \end{cases} \tag{4.19g}$$

$$p_k^{(cb)} = \begin{cases} 0 & \text{if } k=1 \\ q^{(1)} b_1{}^{(1)} u_{k-1}^{(12)} m_1{}^{(1)} = \theta(m_1{}^{(1)})^2 u_{k-1}^{(12)} & \text{if } k \geqslant 2 \end{cases} \tag{4.19h}$$

$$p_k^{(cc)} = p_k^{(bb)} \tag{4.19i}$$

$$(4.19j) \qquad p_k^{(ad)} = \begin{cases} 0 & \text{if } k=1,2 \\ \theta f_1^{(1)} w_k^{(11)} - p_k^{(aa)} - p_k^{(ab)} - p_k^{(ac)} & \text{if } k \geqslant 3 \end{cases}$$

$$(4.19k) \qquad p_k^{(da)} = p_k^{(ad)}$$

$$(4.19l) \qquad p_k^{(bd)} = \begin{cases} \theta m_2^{(1)} & \text{if } k=1 \\ \theta m_3^{(1)} & \text{if } k=2 \\ \theta w_{k+1}^{(21)} - p_k^{(aa)} - p_k^{(ab)} - p_k^{(ac)} - p_k^{(ba)} - p_k^{(bb)} - p_k^{(bc)} - p_k^{(ad)} & \text{if } k \geqslant 3 \end{cases}$$

$$(4.19m) \qquad p_k^{(cd)} = \begin{cases} 0 & \text{if } k=1,2 \\ q^{(1)} b_1^{(1)} w_k^{(11)} - p_k^{(ca)} - p_k^{(cb)} - p_k^{(cc)} = & \\ = \theta m_1^{(1)} w_k^{(11)} - p_k^{(ca)} - p_k^{(cb)} - p_k^{(cc)} & \text{if } k \geqslant 3 \end{cases}$$

$$(4.19n) \qquad p_k^{(db)} = p_k^{(cd)}$$

$$(4.19o) \qquad p_k^{(dc)} = p_k^{(bd)}$$

$$(4.19p) \qquad p_k^{(dd)} = q^{(1)} h_k^{(11)} - p_k^{(aa)} - p_k^{(ab)} - \ldots - p_k^{(db)} - p_k^{(dc)}.$$

If one takes $k=2$ in $\theta f_1^{(1)} u_k^{(11)} - p_k^{(aa)}$ (Equation (4.19c)) one gets: $\theta f_1^{(1)} u_2^{(11)} + -\theta (f_1^{(1)})^2 u_1^{(12)} = 0$, using the fact that $u_2^{(11)} = f_1^{(1)} u_1^{(12)}$. Taking $k=1$ gives: $\theta f_1^{(1)} u_1^{(11)} = 0$ and so it can be concluded that:

$$p_k^{(ac)} = \theta f_1^{(1)} u_k^{(11)} - p_k^{(aa)} \qquad \text{for all } k \geqslant 1.$$

Something similar holds for the other relations (4.19). Thus, substitution of (4.19) in (4.18) results in:

$$(4.20a) \qquad p_k^{(00)} = \theta (f_1^{(1)})^2 u_{k-1}^{(12)}$$

$$(4.20b) \qquad p_k^{(01)} = p_k^{(10)} = \theta f_1^{(1)} \{u_k^{(11)} + m_1^{(1)} u_{k-1}^{(12)}\} - p_k^{(00)}$$

$$(4.20c) \qquad p_k^{(11)} = \theta \{2 m_1^{(1)} u_k^{(11)} + u_{k+1}^{(21)} + (m_1^{(1)})^2 u_{k-1}^{(12)}\} - p_k^{(00)} - 2 p_k^{(01)}$$

$$(4.20d) \qquad p_k^{(02)} = p_k^{(20)} = \theta f_1^{(1)} w_k^{(11)} - p_k^{(00)} - p_k^{(01)}$$

$$(4.20e) \qquad p_k^{(12)} = p_k^{(21)} = \theta \{w_{k+1}^{(21)} + m_1^{(1)} w_k^{(11)}\} - p_k^{(00)} - 2 p_k^{(01)} - p_k^{(11)} - p_k^{(02)}$$

$$(4.20f) \qquad p_k^{(22)} = q^{(1)} h_k^{(11)} - p_k^{(00)} - 2 p_k^{(01)} - p_k^{(11)} - 2 p_k^{(02)} - 2 p_k^{(12)}.$$

One can substitute (4.20a) in the right side of (4.20b); (4.20a and b) in the right side of (4.20c) and so on, giving:

(4.21a) $$p_k^{(00)} = \theta(f_1^{(1)})^2 u_{k-1}^{(12)}$$

(4.21b) $$p_k^{(01)} = p_k^{(10)} = \theta f_1^{(1)}\{(m_1^{(1)}-f_1^{(1)})u_{k-1}^{(12)} + u_k^{(11)}\}$$

(4.21c) $$p_k^{(11)} = \theta\{(m_1^{(1)}-f_1^{(1)})^2 u_{k-1}^{(12)} + 2(m_1^{(1)}-f_1^{(1)})u_k^{(11)} + u_{k+1}^{(21)}\}$$

(4.21d) $$p_k^{(02)} = p_k^{(20)} = \theta f_1^{(1)}\{w_k^{(11)}-u_k^{(11)}-m_1^{(1)}u_{k-1}^{(12)}\}$$

(4.21e) $$p_k^{(12)} = p_k^{(21)} = \theta\{(m_1^{(1)}-f_1^{(1)})w_k^{(11)} + w_{k+1}^{(21)}-m_1^{(1)}(m_1^{(1)}-f_1^{(1)}) \times \\ \times u_{k-1}^{(12)}-(2m_1^{(1)}-f_1^{(1)})u_k^{(11)}-u_{k+1}^{(21)}\}$$

(4.21f) $$p_k^{(22)} = q^{(1)}h_k^{(11)} + \theta\{(m_1^{(1)})^2 u_{k-1}^{(12)} + 2m_1^{(1)}u_k^{(11)} + u_{k+1}^{(21)} + \\ -2m_1^{(1)}w_k^{(11)}-2w_{k+1}^{(21)}\}.$$

For the calculation of the probabilities $p_k^{(ij)}$ from (4.21) one needs:

a. the probabilities $u_{k-1}^{(12)}$, $u_k^{(11)}$ and $u_{k+1}^{(21)}$. These probabilities can be obtained from (2.36).

b. the probabilities $w_k^{(11)}$ and $w_{k+1}^{(21)}$. These probabilities can be obtained from the probabilities $u_l^{(ij)}$ $(l=0,\dots,k)$, using (2.35) and (2.41).

c. the probabilities $h_k^{(11)}$. These probabilities can be obtained from the probabilities $w_l^{(11)}$ $(l=0,\dots,k)$, using (2.52).

4.3. *The autocovariances of the rainfall process with correlated rainfall amounts within a wet spell*

In III, 6.1 two models (II and IV) were described with serial correlation between rainfall amounts within a wet spell. In that section it was assumed that rainfall amounts within a wet spell were correlated according to a first order moving average process, but here no special assumptions are made about the sccs at different lags.

In (4.1) only the term $E(\underline{x}_t\underline{x}_{t+k})$ is influenced by correlation between rainfall amounts within a wet spell. The difference between the autocovariances of models with and without correlation between rainfall amounts within a wet spell is the difference between $E(\underline{x}_t\underline{x}_{t+k})$ for these models.

First, the process with iid rainfall amounts (model II) is considered. The lag k scc of rainfall amounts within a wet spell is denoted by ρ_k. To get an expression for the autocovariances of this model let W_k $(k=1,2,\dots)$ stand for the event that k consecutive wet days occur. For a pure equilibrium process the probability of this event $P(W_k)$ does not depend on the time t. If $\bar{W}_k$ denotes the complement of this event (there is at least 1 dry day), then:

$$(4.22) \qquad E(\underset{\sim}{x}_t\underset{\sim}{x}_{t+k}) = \begin{cases} E\{\underset{\sim}{x}_t\underset{\sim}{x}_{t+1} \mid W_2 \text{ at } [t,t+1]\}\, P(W_2) & \text{if } k=1 \\ E\{\underset{\sim}{x}_t\underset{\sim}{x}_{t+k} \mid W_{k+1} \text{ at } [t,t+k]\}\, P(W_{k+1}) + \\ \quad + E\{\underset{\sim}{x}_t\underset{\sim}{x}_{t+k} \mid W_1 \bar{W}_{k-1} W_1 \text{ at } [t,t+k]\} \times \\ \quad \times P(W_1 \bar{W}_{k-1} W_1) & \text{if } k \geqslant 2. \end{cases}$$

For the conditional expectations on the right side of (4.22):

$$(4.23\text{a}) \qquad \begin{aligned} E\{\underset{\sim}{x}_t\underset{\sim}{x}_{t+k} \mid W_{k+1} \text{ at } [t,t+k]\} &= \operatorname{cov}\{\underset{\sim}{x}_t, \underset{\sim}{x}_{t+k} \mid W_{k+1} \text{ at } [t,t+k]\} + \\ &+ E\{\underset{\sim}{x}_t \mid W_{k+1} \text{ at } [t,t+k]\} \times E\{\underset{\sim}{x}_{t+k} \mid W_{k+1} \text{ at } [t,t+k]\} = \\ &= \rho_k \operatorname{var} \underset{\sim}{z} + \{E(\underset{\sim}{z})\}^2 \end{aligned}$$

and similarly:

$$(4.23\text{b}) \qquad E(\underset{\sim}{x}_t\underset{\sim}{x}_{t+k} \mid W_1 \bar{W}_{k-1} W_1 \;\text{ at } [t,t+k]) = \{E(\underset{\sim}{z})\}^2 \qquad k \geqslant 2.$$

Further

$$(4.24\text{a}) \qquad P(W_k) = Q^{(1)}(k,k) = q^{(1)} Q^{(11)}(k,k)$$

(and thus is $P(W_2) = q^{(1)} h_1{}^{(11)}$)

$$(4.24\text{b}) \qquad P(W_1 \bar{W}_{k-1} W_1) + P(W_{k+1}) = q^{(1)} h_k^{(11)} \qquad k \geqslant 2.$$

The probability $Q^{(11)}(k,k)$ follows from (3.29b).
Substituting (4.23) and (4.24) in (4.22) gives:

$$(4.25) \qquad E(\underset{\sim}{x}_t\underset{\sim}{x}_{t+k}) = \rho_k Q^{(1)}(k+1,k+1) \operatorname{var} \underset{\sim}{z} + q^{(1)} h_k^{(11)} \{E(\underset{\sim}{z})\}^2 \quad k \geqslant 1.$$

From (4.6b) and (4.25) it follows that there is a difference of $\rho_k Q^{(1)}(k+1, k+1) \operatorname{var} \underset{\sim}{z}$ between the autocovariances of models I and II.

When the rainfall amounts are not iid (model IV) the situation is slightly more complicated. If there are $k+1$ consecutive wet days at $[0,k]$ one can distinguish four cases:

a. Dry at $t=-1$, dry at $t=k+1$, giving a type 1 amount at $t=0$ and at $t=k$. The probability of this event is denoted by $P(1, W_{k+1}, 1)$ and equals $\theta f_{k+1}^{(1)}$ (θ is as before $1/(\mu^{(1)} + \mu^{(2)})$).
b. Dry at $t=-1$, wet at $t=k+1$, giving a type 1 amount at $t=0$ and a type 2 amount at $t=k$. The probability of this event is $P(1, W_{k+1}, 2) = \theta m_{k+1}^{(1)}$.
c. Wet at $t=-1$, dry at $t=k+1$, giving a type 2 amount at $t=0$ and a type 1 amount at $t=k$. The probability of this event is $P(2, W_{k+1}, 1) = \theta m_{k+1}^{(1)}$.
d. Wet at $t=-1$, wet at $t=k+1$, giving a type 2 amount at $t=0$ and at $t=k$. This event occurs with probability $P(2, W_{k+1}, 2) = Q^{(1)}(k+3, k+3)$.

Analogous to (4.25) it can be shown that:

$$(4.26) \qquad E(\underline{x}_t\underline{x}_{t+k}) = \rho_k \sum_{i=1}^{2}\sum_{j=1}^{2} \{\text{var}\,\underline{z}^{(i)}\,\text{var}\,\underline{z}^{(j)}\}^{\frac{1}{2}} P(i,W_{k+1},j) + $$

$$+ \sum_{i=0}^{2}\sum_{j=0}^{2} E(\underline{z}^{(i)})E(\underline{z}^{(j)})p_k^{(ij)} \qquad k \geqslant 1$$

assuming that the lag k scc ρ_k does not depend on the different types of rainfall amounts (this assumption is also made in III, 5.1).

From (4.15) and (4.26) it is seen that there is a difference of

$$\rho_k \sum_{i=1}^{2}\sum_{j=1}^{2} \{\text{var}\,\underline{z}^{(i)}\,\text{var}\,\underline{z}^{(j)}\}^{\frac{1}{2}} P(i,W_{k+1},j)$$

between the autocovariances of models III and IV.

For the calculation of the first scc it should be noted that:

$$(4.27) \qquad P(i,W_2,j) = p_1^{(ij)} \qquad i=1,2;j=1,2$$

which simplifies Equation (4.26).

5. The variance-time curve

In this section a short discussion is given about the calculation of the variance-time curve for different types of models. The main body of this section deals with the asymptotic behaviour of the variance-time curve.

Expressions for the variance-time curve can be readily obtained from those of the autocovariances. Let $\underline{S}_n(k)$ be the number of wet days in a k-day period, then its variance is:

$$(5.1) \qquad \text{var}\,\underline{S}_n(k) = kC_{nn}(0) + 2\sum_{l=1}^{k-1}(k-l)C_{nn}(l).$$

The autocovariances $C_{nn}(l)$ follow from (4.4). An analogous relation exists between the variance of the k-day rainfall total $\underline{S}_x(k)$ and the autocovariances $C_{xx}(l)$ of the rainfall process.

When dealing with a model with iid rainfall amounts (Model I, see (4.5)) the following relation exists between var $\underline{S}_x(k)$ and var $\underline{S}_n(k)$:

$$(5.2) \qquad \text{var}\,\underline{S}_x(k) = E\{\text{var}\,[\underline{S}_x(k)\,|\,\underline{S}_n(k)]\} + \text{var}\{E\,[\underline{S}_x(k)\,|\,\underline{S}_n(k)]\} =$$

$$= E\,[\underline{S}_n(k)]\,\text{var}\,\underline{z} + [E(\underline{z})]^2\,\text{var}\,\underline{S}_n(k).$$

The wet-dry processes, considered in the previous chapter, have approximately the same mean and therefore, only the second term on the right side of (5.2) is different for these processes. The type of wet-dry process has therefore a larger influence on var $\underline{S}_n(k)$ than on var $\underline{S}_x(k)$(see also III, 6.2).

In this section special attention is paid to the asymptotic behaviour of var $\underline{S}_n(k)$. It is shown that for the wet-dry processes considered in the previous section

$$(5.3) \qquad \operatorname{var} \underline{S}_n(k) = a_2 k + b_2 + o(1)$$

when k tends to infinity. In (5.3), $o(1)$ denotes a function of k tending to zero when $k \rightarrow \infty$. The coefficient a_2 only depends on the first 2 moments of the lengths of wet and dry spells, whereas the coefficient b_2 depends on the first 3 moments. For model I the asymptotic behaviour of var $\underline{S}_x(k)$ follows immediately from (5.2) and (5.3).

Section 5.1 deals with the asymptotic behaviour of var $\underline{S}_n(k)$ when the occurrence of a wet day is a recurrent event. The asymptotic behaviour of var $\underline{S}_n(k)$ for an alternating renewal process is discussed in Section 5.2.

5.1. *The asymptotic behaviour of the variance-time curve for renewal processes*

Discussions about the asymptotic behaviour of the variance-time curve of a renewal process were given by FELLER (1949), SMITH (1954, 1959) and COX (1962). The first author dealt with a process in discrete time, while the others discussed processes in continuous time. The derivation, given here, follows the one given by COX (1962) closely.

Instead of working with var $\underline{S}_n(k)$ or $\{E[\underline{S}_n(k)]\}^2$, it is easier to consider (cf. COX (1962), Section 4.5):

$$(5.4) \qquad \psi_k = E[\underline{S}_n(k)\{\underline{S}_n(k) + 1\}] \qquad k=1,2,\dots$$

One sees immediately that for an equilibrium renewal process $\psi_1 = 2/\mu$. For this process the ψ_ks and the variances are related by:

$$(5.5) \qquad \psi_k = \operatorname{var} \underline{S}_n(k) + \{E[\underline{S}_n(k)]\}^2 + E[\underline{S}_n(k)] =$$

$$= \operatorname{var} \underline{S}_n(k) + \frac{k^2}{\mu^2} + \frac{k}{\mu} \qquad k=1,2,\dots$$

Substituting (5.1) in (5.5) gives:

$$(5.6) \qquad \psi_k = kC_{nn}(0) + 2\sum_{l=1}^{k-1}(k-l)C_{nn}(l) + \frac{k^2}{\mu^2} + \frac{k}{\mu} \qquad k=1,2,\dots.$$

From (4.4b) it follows:

$$(5.7) \qquad \psi_k = \frac{k}{\mu} - \frac{k}{\mu^2} + \frac{2}{\mu}\sum_{l=1}^{k-1}(k-l)u_l - \frac{2}{\mu^2}\,\frac{1}{2}\,k(k-1) + \frac{k^2}{\mu^2} + \frac{k}{\mu} =$$

$$= \frac{2k}{\mu} + \frac{2}{\mu}\sum_{l=1}^{k-1}(k-l)u_l \qquad k=1,2,\dots.$$

For the gf $\Psi(s)$ of $\{\psi_k\}$ holds:

$$(5.8) \qquad \Psi(s) = \sum_{k=1}^{\infty} \psi_k s^k = \frac{2}{\mu} \sum_{k=1}^{\infty} k s^k + \frac{2}{\mu} \sum_{k=1}^{\infty} \sum_{l=1}^{k-1} (k-l) u_l s^k =$$

$$= \frac{2}{\mu} \sum_{k=1}^{\infty} k s^k + \frac{2}{\mu} \sum_{l=1}^{\infty} u_l s^l \sum_{k=l+1}^{\infty} (k-l) s^{k-l} =$$

$$= \frac{2}{\mu} U(s) \sum_{k=1}^{\infty} k s^k.$$

Since $\sum_{k=1}^{\infty} k s^k = s/(1-s)^2$ and substituting (2.25), one gets:

$$(5.9) \qquad \Psi(s) = \frac{2s}{\mu(1-s)^2} \frac{1}{1-F(s)}$$

which is the discrete analogue of Equation (4.5.5) of Cox (1962).

For the term $1-F(s)$ in (5.9) it follows from (A1.2):

$$(5.10) \qquad 1-F(s) = -(s-1) \left\{ \mu_{[1]} + \frac{\mu_{[2]}}{2!}(s-1) + \frac{\mu_{[3]}}{3!}(s-1)^2 + \ldots \right\}$$

where $\mu_{[1]}, \mu_{[2]}, \mu_{[3]}, \ldots$ are the factorial moments (see Appendix A1) of the recurrence times.

Since $\mu_{[1]} = \mu > 0$ the equation $F(s) = 1$ has a simple root at $s = 1$. The following remarks can be made about the other roots of $F(s) = 1$:

a. If s is a complex root, then its conjugate $\bar{s}$ is also a root. Namely, from $F(s) = 1$ it follows $\overline{F(s)} = F(\bar{s}) = 1$.

b. There are no roots within the unit circle, since if $|s| < 1$ then: $|F(s)| \leqslant F(|s|) < F(1) = 1$.

c. If $s = e^{i\phi}$ is a root of $F(s) = 1$, then:

$$F(s) + F(\bar{s}) = \sum_{k=1}^{\infty} f_k(e^{ik\phi} + e^{-ik\phi}) = 2$$

giving $\sum_{k=1}^{\infty} f_k \cos k\phi = 1$.

But, since $\sum_{k=1}^{\infty} f_k = 1$ and $0 \leqslant f_k \leqslant 1$ this can only be true if $f_k = 0$ for values of k for which $\cos k\phi \neq 1$. That is f_k can only be non-zero if $k = 2m\pi/\phi$ (m is a positive integer).

For rainfall processes $\{f_k\}$ is a non-periodic distribution and therefore the equation $F(s) = 1$ has a simple root at $s = 1$ and the other roots lie outside the unit circle. Hence, expanding (5.9) in powers of $s-1$ results in an expression of the form:

$$\Psi(s) = -\frac{2s}{\mu}\left\{\frac{A_3}{(s-1)^3} + \frac{A_2}{(s-1)^2} + \frac{A_1}{s-1} + o(\frac{1}{s-1})\right\} \tag{5.11}$$

where $o(s-1)^h$ stands for functions $f(s)$ with the property that

$$\lim_{s\to 1} \frac{f(s)}{(s-1)^h} = 0.$$

Insight into the behaviour of the ψ_ks for large values of k can be obtained by inverting (5.11) term by term. This technique is discussed for $F(s)$ being a rational function in s. Then $\psi(s)$ also is a rational function in s and the remainder term in (5.11) can be split into partial fractions. For the inversion of (5.11), one needs:

$$\sum_{k=1}^{\infty} s^k = \frac{s}{1-s} \tag{5.12a}$$

$$\sum_{k=1}^{\infty} ks^k = \frac{s}{(1-s)^2} \tag{5.12b}$$

$$\sum_{k=2}^{\infty} k(k-1)s^k = \frac{2s^2}{(1-s)^3} \text{ or } \sum_{k=1}^{\infty} k(k+1)s^k = \frac{2s}{(1-s)^3}. \tag{5.12c}$$

The Equations (5.12b and c) follow from differentiation of (5.12a).

Therefore, Equation (5.11) can be written as:

$$\sum_{k=1}^{\infty} \psi_k s^k = -\frac{2}{\mu}\left\{-A_1 \sum_{k=1}^{\infty} s^k + A_2 \sum_{k=1}^{\infty} ks^k - \frac{1}{2} A_3 \sum_{k=1}^{\infty} k(k+1)s^k + \right. \\ \left. + \sum_{j=1}^{n} \sum_{k=1}^{\infty} \theta_{jk} s^k \right\} \tag{5.13}$$

where n is a finite number related to the roots of $F(s) = 1$. A simple real root s_j of $F(s) = 1$ gives a θ_{jk} proportional to s_j^{-k} (since $\frac{s}{s_j - s} = \sum_{k=1}^{\infty} (s/s_j)^k$). A root s_j with multiplicity $r+1$ gives θ_{jk}s proportional to $s_j^{-k}, ks_j^{-k}, \ldots, k^r s_j^{-k}$, whereas a simple complex root, $s_j = |s_j| e^{i\phi}$, and its conjugate lead to a θ_{jk} proportional to $|s_j|^{-k} \cos(k\phi)$. However, since $|s_j| > 1$, the term θ_{jk} tends to zero for large k. Therefore if $k \to \infty$ one gets:

$$\psi_k = \frac{2}{\mu}\{\frac{1}{2} A_3 k^2 + (\frac{1}{2} A_3 - A_2)k + A_1\} + o\,(1). \tag{5.14}$$

The argument leading to (5.14) underlies the assumption that $F(s)$ is a rational function in s. It can be shown, however, that (5.14) holds under very weak assumptions about the distribution of the recurrence times. A rigorous proof requires difficult Tauberian arguments, analogous to the continuous case (cf. SMITH (1959)).

From (5.9) and (5.10) it follows:

$$(5.15)\qquad \Psi(s) = \frac{-2s}{\mu^2(s-1)^3}\left\{1+\frac{\mu_{[2]}}{2\mu_{[1]}}(s-1)+\frac{\mu_{[3]}}{6\mu_{[1]}}(s-1)^2+\ldots\right\}^{-1} =$$

$$= \frac{-2s}{\mu^2(s-1)^3}\left\{1-\left[\frac{\mu_{[2]}}{2\mu_{[1]}}(s-1)+\frac{\mu_{[3]}}{6\mu_{[1]}}(s-1)^2+\ldots\right]+\right.$$

$$\left.+\left[\frac{\mu_{[2]}}{2\mu_{[1]}}(s-1)+\frac{\mu_{[3]}}{6\mu_{[1]}}(s-1)^2+\ldots\right]^2+\ldots\right\}=$$

$$= -\frac{2s}{\mu^2(s-1)^3}\left\{1-\frac{\mu_{[2]}}{2\mu_{[1]}}(s-1)+\left[\frac{\mu_{[2]}^2}{4\mu_{[1]}^2}-\frac{\mu_{[3]}}{6\mu_{[1]}}\right](s-1)^2+\right.$$

$$\left.+\,o(s-1)^2\right\}$$

when $s\rightarrow 1$.

From a comparison of (5.11) and (5.15) one gets relations between the coefficients A_1, A_2, A_3 and the first 3 factorial moments. To express these coefficients in the mean, the variance and the third central moment, use is made of the relations:

$$(5.16a)\qquad \mu_{[2]} = \mu_2 + \mu^2 - \mu$$

$$(5.16b)\qquad \mu_{[3]} = \mu_3 + 3\mu\mu_2 + \mu^3 - 3\mu_2 - 3\mu^2 + 2\mu.$$

One finally gets:

$$(5.17a)\qquad A_3 = 1/\mu$$

$$(5.17b)\qquad A_2 = \frac{-(\mu_2+\mu^2-\mu)}{2\mu^2}$$

$$(5.17c)\qquad A_1 = \frac{\mu_2^2}{4\mu^3}-\frac{\mu_3}{6\mu^2}-\frac{1}{12\mu}+\frac{\mu}{12}.$$

Substituting this in (5.14) gives:

$$(5.18)\qquad \psi_k = \frac{k^2}{\mu^2}+\frac{\mu_2+\mu^2}{\mu^3}k+\frac{\mu_2^2}{2\mu^4}-\frac{\mu_3}{3\mu^3}-\frac{1}{6\mu^2}+\frac{1}{6}+o(1).$$

From (5.5) and (5.18) the asymptotic formula for the variance becomes:

$$(5.19)\qquad \operatorname{var}\underline{S}_n(k) = \frac{\mu_2}{\mu^3}k+\frac{\mu_2^2}{2\mu^4}-\frac{\mu_3}{3\mu^3}-\frac{1}{6\mu^2}+\frac{1}{6}+o(1).$$

Problems arise when μ, μ_2 or μ_3 are infinite. For instance, assume that for large n the survivor function has the form:

(5.20) $$m_n \sim An^{-\alpha} \qquad 1 < \alpha < 2.$$

For this renewal process the distribution of the recurrence times has a finite mean, but the higher moments are infinite.

For the equilibrium renewal process it can be shown, that as $k \to \infty$:

(5.21) $$\text{var } \underline{S}_n(k) \sim \frac{2A}{(\alpha-1)\,(2-\alpha)\,(3-\alpha)\mu^3}\, k^{3-\alpha}.$$

The proof is analogous to that of FELLER (1949) for the ordinary renewal process. So the variance-time curve tends to a straight line on double logarithmic paper.

The correctness of (5.19) can easily be verified for the Bernoulli process, discussed in 2.1.1. For this process the recurrence times are geometrically distributed (see (2.13)); expressions for the mean, the variance and the third central moment of this distribution are given in Table 3.3 of Chapter III. Substituting these expressions in (5.19) gives:

(5.22) $$\text{var } \underline{S}_n(k) = kpq + \frac{1}{2}q^2 - \frac{1}{3}q(1+q) - \frac{1}{6}p^2 + \frac{1}{6} + o(1) =$$
$$= kpq + o(1).$$

which is obviously correct, since kpq is the variance of a binomial variable.

5.2. *The asymptotic behaviour of the variance-time curve for alternating renewal processes*

In this section the asymptotic formula of the variance-time curve is given for a pure equilibrium alternating renewal process. It is assumed that state 1 corresponds to a wet day and state 2 to a dry day. For this reason these events are denoted by w and d, respectively.

For ψ_k defined by (5.4) one gets the expression:

(5.23) $$\psi_k = \text{var } \underline{S}_n(k) + k^2(q^{(w)})^2 + kq^{(w)}.$$

Analogous to (5.7) it can be shown that:

(5.24) $$\psi_k = 2kq^{(w)} + 2q^{(w)} \sum_{l=1}^{k-1} (k-l)h_l^{(ww)}.$$

The analogue of (5.9) is:

(5.25) $$\Psi(s) = 2q^{(w)}\, sH^{(ww)}(s)/(1-s)^2.$$

Substitution of (2.56b) gives:

(5.26) $$\Psi(s) = \frac{2q^{(w)}s}{(1-s)^2} \left\{ \frac{1}{1-s} - \frac{s\{1-F^{(w)}(s)\}\{1-F^{(d)}(s)\}}{\mu^{(w)}(1-s)^2\{1-F^{(w)}(s)F^{(d)}(s)\}} \right\}.$$

Let $K(s)$ and $L(s)$ be functions of s, such that

$$(5.27a) \qquad \{1-F^{(w)}(s)\}\{1-F^{(d)}(s)\} = (s-1)^2 K(s)$$

$$(5.27b) \qquad 1-F^{(w)}(s)F^{(d)}(s) = -(s-1)L(s)$$

then (5.26) becomes:

$$(5.28) \qquad \Psi(s) = \frac{2q^{(w)}s}{(1-s)^3} + \frac{2q^{(w)}s^2}{\mu^{(w)}(s-1)^3}\,\frac{K(s)}{L(s)}.$$

Since both wet and dry intervals have non-zero means it follows that $K(1) \neq 0$ and $L(1) \neq 0$. Therefore expansion of (5.28) gives, as $s \to 1$:

$$(5.29) \qquad \Psi(s) = \frac{2q^{(w)}s}{(1-s)^3} + \frac{2q^{(w)}s^2}{\mu^{(w)}}\left\{\frac{A_3}{(s-1)^3} + \frac{A_2}{(s-1)^2} + \frac{A_1}{s-1} + o\left(\frac{1}{s-1}\right)\right\}.$$

Using (5.12) one gets for $k \to \infty$:

$$(5.30) \qquad \psi_k = q^{(w)}k(k+1) + 2q^{(w)}\{-\tfrac{1}{2}k(k-1)A_3 + (k-1)A_2 - A_1\}/\mu^{(w)} + o(1)$$

provided some convergence conditions are satisfied. Equation (5.30) can also be written as:

$$(5.31) \qquad \psi_k = q^{(w)}\left\{1 - \frac{A_3}{\mu^{(w)}}\right\}k^2 + q^{(w)}\left\{1 + \frac{2A_2}{\mu^{(w)}} + \frac{A_3}{\mu^{(w)}}\right\}k - \frac{2q^{(w)}}{\mu^{(w)}} \times$$
$$\times \left\{A_1 + A_2\right\} + o(1).$$

For obtaining expressions for the coefficients A_1, A_2 and A_3, the functions $K(s)$ and $L(s)$, defined by (5.27), are expanded in powers of $s-1$. Since for the pgfs $F^{(d)}(s)$ and $F^{(w)}(s)$ similar expressions hold as (5.10), one gets if $s \to 1$:

$$(5.32a) \qquad K(s) = \mu^{(w)}_{[1]}\mu^{(d)}_{[1]} + \tfrac{1}{2}\{\mu^{(w)}_{[1]}\mu^{(d)}_{[2]} + \mu^{(w)}_{[2]}\mu^{(d)}_{[1]}\}(s-1) +$$
$$+ \frac{1}{12}\{2\mu^{(w)}_{[1]}\mu^{(d)}_{[3]} + 2\mu^{(w)}_{[3]}\mu^{(d)}_{[1]} + 3\mu^{(w)}_{[2]}\mu^{(d)}_{[2]}\}(s-1)^2 + o(s-1)^2 =$$
$$= K(1) + K'(1)(s-1) + \tfrac{1}{2}K''(1)(s-1)^2 + o(s-1)^2$$

$$(5.32b) \qquad L(s) = \mu^{(w)}_{[1]} + \mu^{(d)}_{[1]} + \tfrac{1}{2}\{\mu^{(w)}_{[2]} + \mu^{(d)}_{[2]} + 2\mu^{(w)}_{[1]}\mu^{(d)}_{[1]}\}(s-1) +$$
$$+ \frac{1}{6}\{\mu^{(w)}_{[3]} + \mu^{(d)}_{[3]} + 3\mu^{(w)}_{[1]}\mu^{(d)}_{[2]} + 3\mu^{(w)}_{[2]}\mu^{(d)}_{[1]}\}(s-1)^2 + o(s-1)^2 =$$
$$= L(1) + L'(1)(s-1) + \tfrac{1}{2}L''(1)(s-1)^2 + o(s-1)^2.$$

A similar expansion as (5.15) gives, for $s \to 1$:

$$(5.33) \qquad \frac{1}{L(s)} = \frac{1}{L(1)}\left\{1 - \frac{L'(1)}{L(1)}(s-1) + \left[\frac{L'(1)^2}{L(1)^2} - \frac{L''(1)}{2L(1)}\right](s-1)^2\right\} + o(s-1)^2.$$

Multiplication of (5.32a) by (5.33) gives:

$$(5.34)\qquad \frac{K(s)}{L(s)} = \frac{K(1)}{L(1)} + \left\{\frac{K'(1)}{L(1)} - \frac{K(1)L'(1)}{L(1)^2}\right\}(s-1) +$$

$$+ \left\{\frac{K(1)L'(1)^2}{L(1)^3} - \frac{K(1)L''(1)}{2L(1)^2} - \frac{K'(1)L'(1)}{L(1)^2} + \frac{K''(1)}{2L(1)}\right\}(s-1)^2 +$$

$$+ o(s-1)^2 = A_3 + A_2(s-1) + A_1(s-1)^2 + o(s-1)^2$$

using (5.28) and (5.29).

Relations between the coefficients A_1, A_2 and A_3, and the first 3 factorial moments of lengths of wet and dry spells are readily obtained from (5.32) and (5.34). Expressing these coefficients in the mean, the variance and the third central moment of wet and dry intervals by (5.16) involves lengthly computations. One finally gets:

$$(5.35a)\qquad A_3 = \frac{\mu^{(w)}\mu^{(d)}}{\mu^{(w)}+\mu^{(d)}}$$

$$(5.35b)\qquad A_2 = \frac{\mu^{(d)}\mu_2^{(w)}+\mu^{(w)}\mu_2^{(d)}}{2(\mu^{(w)}+\mu^{(d)})} - \frac{(\mu_2^{(w)}+\mu_2^{(d)})\mu^{(w)}\mu^{(d)}}{2(\mu^{(w)}+\mu^{(d)})^2} - \frac{\mu^{(w)}\mu^{(d)}}{2(\mu^{(w)}+\mu^{(d)})}$$

$$(5.35c)\qquad A_1+A_2 = -\frac{(\mu^{(d)}\mu_2^{(w)}-\mu^{(w)}\mu_2^{(d)})^2}{4(\mu^{(w)}+\mu^{(d)})^3} + \frac{(\mu^{(d)})^2\mu_3^{(w)}+(\mu^{(w)})^2\mu_3^{(d)}}{6(\mu^{(w)}+\mu^{(d)})^2} +$$

$$- \frac{2\mu^{(w)}\mu^{(d)}+(\mu^{(w)}\mu^{(d)})^2}{12(\mu^{(w)}+\mu^{(d)})}.$$

From (5.23), (5.31) and (5.35) it follows, for $k \to \infty$:

$$(5.36)\qquad \operatorname{var} \underline{S}_n(k) = \frac{(\mu^{(d)})^2\mu_2^{(w)}+(\mu^{(w)})^2\mu_2^{(d)}}{(\mu^{(w)}+\mu^{(d)})^3}k + \frac{(\mu^{(d)}\mu_2^{(w)}-\mu^{(w)}\mu_2^{(d)})^2}{2(\mu^{(w)}+\mu^{(d)})^4} +$$

$$- \frac{(\mu^{(d)})^2\mu_3^{(w)}+(\mu^{(w)})^2\,\mu_3{}^{(d)}}{3(\mu^{(w)}+\mu^{(d)})^3} + \frac{2\mu^{(w)}\mu^{(d)}+(\mu^{(w)}\mu^{(d)})^2}{6(\mu^{(w)}+\mu^{(d)})^2} +$$

$$+\ o(1).$$

When the wet-dry process is a renewal process (wet or dry intervals are geometrically distributed) it can be shown that Equation (5.36) reduces to (5.19).

For instance, let the occurrence of a wet day be a recurrent event, then it follows from (2.7) and Table 3.3 of Chapter III:

$$(5.37a)\qquad \mu^{(w)} = 1/m_1$$

$$(5.37b)\qquad \mu_2^{(w)} = f_1/m_1^2$$

$$(5.37c)\qquad \mu_3^{(w)} = f_1(1+f_1)/m_1^3.$$

From the relations (2.10) and (2.12) it can be shown that for lengths of dry spells:

(5.38a) $\mu^{(d)} = (\mu-1)/m_1$

(5.38b) $\mu_2^{(d)} = \{m_1\mu_2 - f_1(\mu-1)^2\}/m_1^2$

(5.38c) $\mu_3^{(d)} = \{m_1^2\mu_3 - 3m_1 f_1 \mu_2(\mu-1) + f_1(1+f_1)(\mu-1)^3\}/m_1^3.$

From (5.37) and (5.38) it follows:

(5.39a) $\mu^{(w)} + \mu^{(d)} = \mu/m_1$

(5.39b) $(\mu^{(d)})^2\mu_2^{(w)} = f_1(\mu-1)^2/m_1^4$

(5.39c) $(\mu^{(w)})^2\mu_2^{(d)} = \dfrac{\mu_2}{m_1^3} - \dfrac{f_1(\mu-1)^2}{m_1^4}$

and thus the coefficient of k in (5.36) is:

(5.40) $$\frac{(\mu^{(d)})^2\mu_2^{(w)} + (\mu^{(w)})^2\mu_2^{(d)}}{(\mu^{(w)}+\mu^{(d)})^3} = \frac{\mu_2}{\mu^3}.$$

For the constant term in (5.36) it can be deduced that:

(5.41a) $$\frac{(\mu^{(d)}\mu_2^{(w)} - \mu^{(w)}\mu_2^{(d)})^2}{2(\mu^{(w)}+\mu^{(d)})^4} = \frac{\mu_2^2}{2\mu^4} - \frac{f_1(\mu-1)\mu_2}{m_1\mu^3} + \tfrac{1}{2}f_1^2\,\frac{(\mu-1)^2}{m_1^2\mu^2}$$

(5.41b) $$-\frac{(\mu^{(d)})^2\mu_3^{(w)} + (\mu^{(w)})^2\mu_3^{(d)}}{3(\mu^{(w)}+\mu^{(d)})^3} = -\frac{\mu_3}{3\mu^3} + \frac{f_1\mu_2(\mu-1)}{m_1\mu^3} - \frac{1}{3}f_1(1+f_1)\,\frac{(\mu-1)^2}{m_1^2\mu^2}$$

(5.41c) $$\frac{2\mu^{(w)}\mu^{(d)} + (\mu^{(w)}\mu^{(d)})^2}{6(\mu^{(w)}+\mu^{(d)})^2} = \frac{2(\mu-1)}{6\mu^2} + \frac{1}{6}\,\frac{(\mu-1)^2}{m_1^2\mu^2}.$$

Since:

(5.42) $$\frac{1}{2}f_1^2 - \frac{1}{3}f_1(1+f_1) + \frac{1}{6} = \frac{1}{6}(3f_1^2 - 2f_1 - 2f_1^2 + 1) = \frac{1}{6}(f_1^2 - 2f_1 + 1) = \frac{1}{6}m_1^2$$

the constant term in (5.36) reduces to:

(5.43) $$\frac{\mu_2^2}{2\mu^4} - \frac{\mu_3}{3\mu^3} + \frac{2(\mu-1)}{6\mu^2} + \frac{(\mu-1)^2}{6\mu^2} = \frac{\mu_2^2}{2\mu^4} - \frac{\mu_3}{3\mu^3} - \frac{1}{6\mu^2} + \frac{1}{6}.$$

For different distributions for lengths of weather spells, Table 5.1 shows the coefficients of the asymptotic expression of the variance for both the wet-dry process and the rainfall process (model I) of the winter season of Winterswijk

TABLE 5.1. Coefficients of the asymptotic expression of var $\underline{S}_n(k)$ and var $\underline{S}_x(k)$ for the winter season of Winterswijk ($\delta = 0.8$ mm). The asymptotic formula is of the form a_2k+b_2 (see Equation (5.3)). The asymptotic variance of the rainfall process is based on model I with a SGD for rainfall amounts on wet days.

Type of wet-dry	var $\underline{S}_n(k)$		var $\underline{S}_x(k)$	
process (see III, 6.1)	b_2	a_2	b_2	a_2
GD–GD	–0.32	0.46	– 7.3	18.8
GD–LSD	–1.85	0.71	–42.8	24.5
SNBD–SNBD	–1.03	0.62	–23.7	22.4
TNBD–TNBD	–1.09	0.62	–24.8	22.1

TABLE 5.2. Exact and approximated values of var $\underset{\sim}{S}_n(k)$ and var $\underset{\sim}{S}_x(k)$ for the winter season of Winterswijk ($\delta = 0.8$ mm). The calculated values are based on model I with TNBDs for lengths of weather spells and a SGD for rainfall amounts on wet days.

	var $\underset{\sim}{S}_n(k)$ in days^2		var $\underset{\sim}{S}_x(k)$ in mm^2	
k	Exact	Approximated	Exact	Approximated
5	2.18	1.99	90.1	85.8
10	5.12	5.07	197.4	196.3
15	8.17	8.16	307.2	306.9
20	11.24	11.24	417.5	417.4
30	17.41	17.41	638.5	638.5

($\delta = 0.8$ mm). The asymptotic variance of model I of the rainfall process was obtained by substituting (5.36) in (5.2). The mean and the variance of the rainfall amounts were based on ML estimates of the SGD. The central moments in (5.36) were obtained from the expressions in Table 3.3 of Chapter III.

The seasonal value of a particular parameter in the model was obtained by averaging monthly estimates (cf. III, 6.1). The SNBD was treated in the same way as the TNBD.

The coefficients strongly depend on the type of wet-dry process. The intercept always turns out to be negative. From (5.36) it is seen that this can only be so when the lengths of weather spells have large third central moments (skew distributions). The largest negative value of the intercept is found for the GD–LSD process, which is a consequence of the large third central moment of the LSD. On the basis of Figure 6.3 of Chapter III (winter) and Table 5.1 it can be concluded that the intercept is comparatively small.

For different values of k, Table 5.2 compares the exact variances (obtained from (5.1)) and the approximated variances (obtained from (5.36)) for a TNBD–TNBD process of the winter season of Winterswijk ($\delta = 0.8$ mm). This table also gives the variances for model I of the rainfall process. The approximation is reasonable for values of k larger than 10.

APPENDIX

A1. Generating functions

The generating function (gf) $A(s)$ of a sequence, $a_0, a_1, a_2, \ldots$ of real numbers is defined by:

$$(\text{A1.1}) \qquad A(s) = \sum_{n=0}^{\infty} a_n s^n .$$

The sequences considered in this chapter have the property that $0 \leqslant a_n < 1$ for $n \geqslant 1$ and therefore $A(s)$ converges within the unit circle $|s| = 1$.

Let $\underline{x}$ be a non-negative integral valued random variable with probability distribution $\{a_n\}$, then the function $A(s)$ is called the probability generating function (pgf) of $\underline{x}$. This function has the following properties:

a. $A(s) = E(s^{\underline{x}})$.

b. $A(1) = 1$.

c. $A^{(k)}(1) = \sum_{n=k}^{\infty} n(n-1)\ldots(n-k+1)a_n = E[\underline{x}(\underline{x}-1)\ldots(\underline{x}-k+1)]$

which is the kth factorial moment $\mu_{[k]}$ of $\underline{x}$.

Because of this last property $A(s)$ can also be written as:

$$(\text{A1.2}) \qquad A(s) = 1 + \sum_{n=1}^{\infty} [\mu_{[n]}(s-1)^n/n!].$$

For this reason $A(s)$ is sometimes called the factorial moment generating function of $\underline{x}$ (cf. Lindgren (1968), 2.4.2).

Let $\{a_n\}$ and $\{b_n\}$ be sequences with gfs $A(s)$ and $B(s)$ and let $\{c_n\}$ be their convolution, that is:

$$(\text{A1.3}) \qquad c_n = \sum_{k=0}^{n} a_k b_{n-k} \qquad\qquad n=0,1,\ldots$$

then the gf $C(s)$ of $\{c_n\}$ satisfies:

$$(\text{A1.4}) \qquad C(s) = A(s) \cdot B(s)$$

(cf. Feller (1968), XI.2).

If $\underline{x}$ and $\underline{y}$ are independent non-negative integral valued random variables it can be shown that the probability distribution of $\underline{x}+\underline{y}$ is the convolution of the probability distributions of $\underline{x}$ and $\underline{y}$, and from (A1.3) and (A1.4) it follows that the pgf of $\underline{x}+\underline{y}$ is the product of the pgfs of $\underline{x}$ and $\underline{y}$. This follows also immediately from:

$$(\text{A1.5}) \qquad E(s^{\underline{x}+\underline{y}}) = E(s^{\underline{x}})\,E(s^{\underline{y}}).$$

V. ANALYSIS OF RAINFALL DATA FROM FOREIGN STATIONS

1. Introduction

In Chapters III and IV a rainfall model was developed for Dutch rainfall series. In this chapter the adequacy of this model is tested for some foreign rainfall data. The following rainfall series are considered:

a. From India: Bangalore (1901–1930; 1933–1970), Calcutta (1901–1970) and New Delhi (1901–1970).

b. From Indonesia: Jakarta-27 (1942–1973) and Pasar Minggu (1880–1885; 1889–1911; 1913–1920; 1922–1944; 1951–1959).

c. From Surinam: Paramaribo (1899–1968) and Domburg (1910–1968).

d. From Sudan: Khartoum (1902–1940).

e. From Egypt: Alexandria (1901–1940).

For Bangalore the years of 1931 and 1932 were omitted because data are missing for several days; these could not easily be supplemented with rainfall data from nearby stations. For the same reason the years of 1912, 1921 and the period 1945–1950 of the Pasar Minggu series were not taken into consideration; the period 1886–1888 was omitted, because the data were highly suspicious. Figure 1.1 compares annual totals of Pasar Minggu and Jakarta-27 for the periods 1880–1899 and 1910–1939. This figure shows that annual totals of

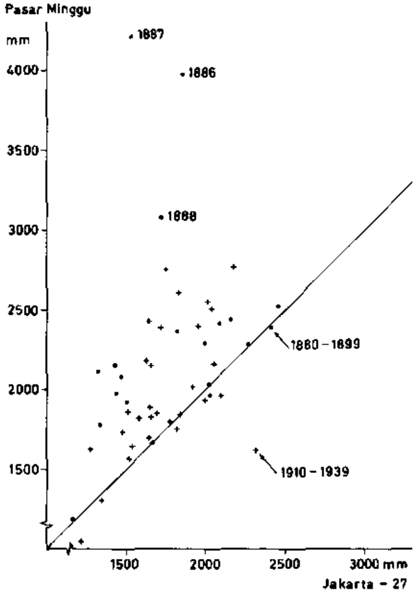


FIG. 1.1. Annual totals of Pasar Minggu and Jakarta-27 for two different periods.

Pasar Minggu are extremely large for the years of 1886, 1887 and 1888. A reasonable explanation for this phenomenon is not known.

Spells or cycles bounded by a gap in the series were discarded in the analysis.

The rainfall series from India, Indonesia and Surinam were analysed in the same manner as Dutch rainfall series (see Section 2). For the series of Khartoum and Alexandria the techniques applied to Dutch rainfall series can be troublesome since these series are characterized by long periods with no rainfall at all. Therefore, these series are discussed in a separate section (Section 3).

2. Analysis of rainfall data from India, Indonesia and Surinam

For each month daily means are given in Figure 2.1 for the series of Bangalore, Calcutta, New Delhi, Jakarta-27, Pasar Minggu and Paramaribo. This figure also shows the means of wet days. As in previous chapters a wet day is defined as a day with a rainfall amount of at least δ millimeters. In Figure 2.1 the value of δ is 0.2 mm for Indian rainfall series, 3.0 mm for Indonesian rainfall series and 0.3 mm for Paramaribo. All rainfall series given in Figure 2.1 are characterized by wet and dry periods. It should be noted, however, that seasonal variations in the mean rainfall amounts on wet days are less obvious, especially for Indonesian (Pasar Minggu) and Surinam data.

On the basis of Figure 2.1 months were grouped in seasons. The different seasons are given in Table 2.1. The driest season is always denoted by Sd and the wettest season by Sw. The seasons Sdw and Swd are transition periods. For Surinam the season Sdw includes the short rainy season (December-January), the short dry season (February-March) and the transition period to the long rainy season.

If a seasonal estimate of a parameter is given from an analysis by wet-dry or dry-wet cycles, the season Sdw relates to the period 1900–1969 for Paramaribo and to the period 1909–1967 for Domburg.

During the wet season the number of wet days can be quite large. For instance, the fraction of wet days during the wettest month is 0.77 for Calcutta A2 and 0.87 for Paramaribo A3. This fraction is lower for the other series,

TABLE 2.1. Classification of seasons. The dry season is denoted by Sd; the wet season by Sw; the period between the dry and wet season by Sdw, and the period between the wet and dry season by Swd.

	Season			
Country	Sdw	Sw	Swd	Sd
India	April-May	June-August	September-October	November-March
Indonesia	October-December	January-February	March-May	June-September
Surinam	December-April	May-June	July-August	September-November

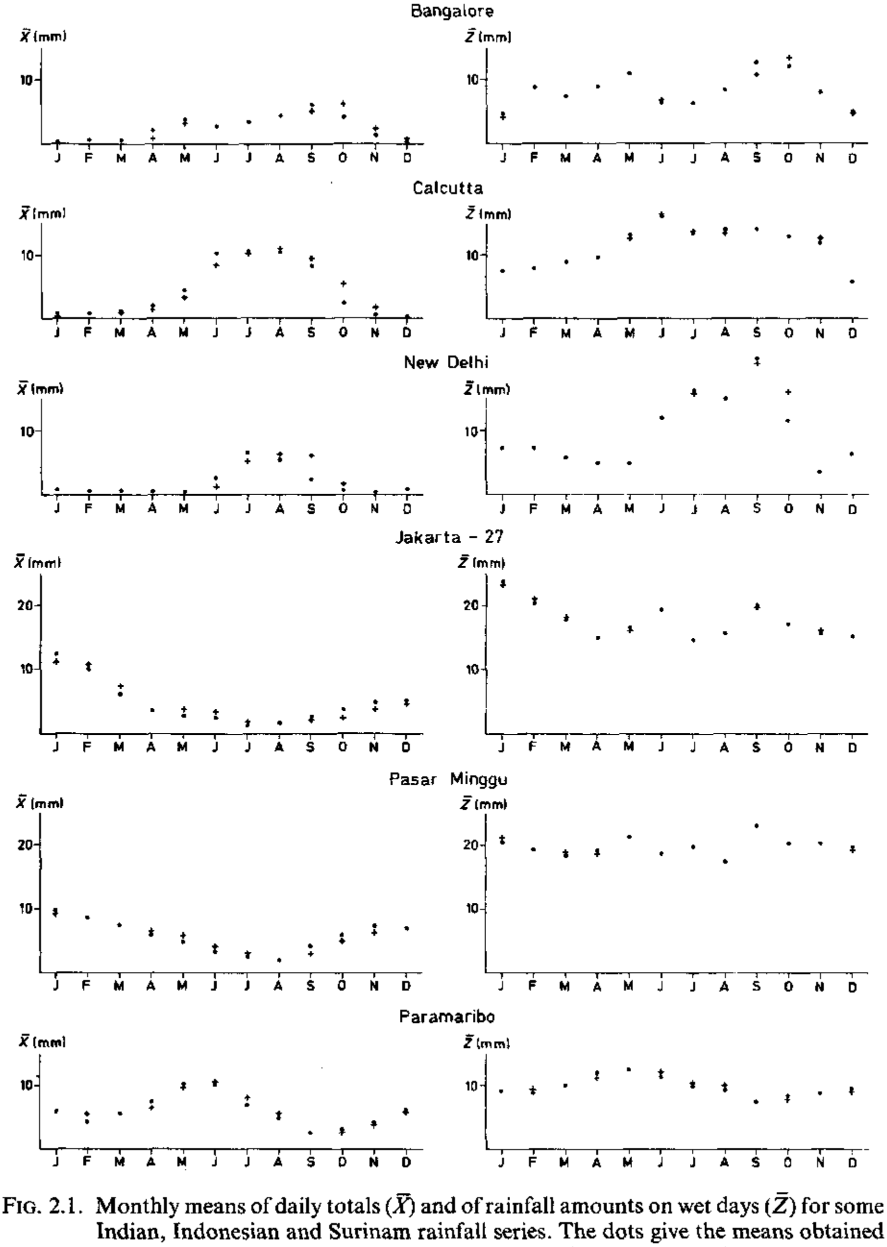


FIG. 2.1. Monthly means of daily totals ($\bar{X}$) and of rainfall amounts on wet days ($\bar{Z}$) for some Indian, Indonesian and Surinam rainfall series. The dots give the means obtained by method A; if there is a difference between the A and B estimates the mean obtained by method B is denoted by a plus sign.

namely: 0.56 for Bangalore A2, 0.40 for New Delhi A2, 0.52 for Pasar Minggu A30 and 0.68 for Paramaribo A22. The large difference between Paramaribo A3 and Paramaribo A22 is an indication for a J-shaped frequency distribution of the rainfall amounts.

There is some difference between the dry periods of the rainfall series given in Figure 2.1. For Jakarta-27, Pasar Minggu and Paramaribo there is still a considerable rainfall amount during the dry period; for Indian series the daily mean is very small for months in the dry period. There are many months with no rainfall at all (see Figure 3.3 of Chapter II), but it is also possible that showers of more than 40 mm occur during the dry period.

For data from India, Indonesia and Surinam homogeneity of the rainfall series and of the wet-dry series is discussed in Section 2.1. Because the analysis of homogeneity was based on the annual totals and the annual number of wet days, some features of the distribution of these quantities are also given in that section. Section 2.2 deals with the distribution of lengths of wet and dry spells, whereas in Section 2.3 the behaviour of rainfall amounts on wet days is discussed. Serial correlation coefficients and variances of k-day totals of the historic series and various rainfall models are compared in Section 2.4. Section 2.5 gives a comparison between features of the historic series of Pasar Minggu and those of some generated sequences.

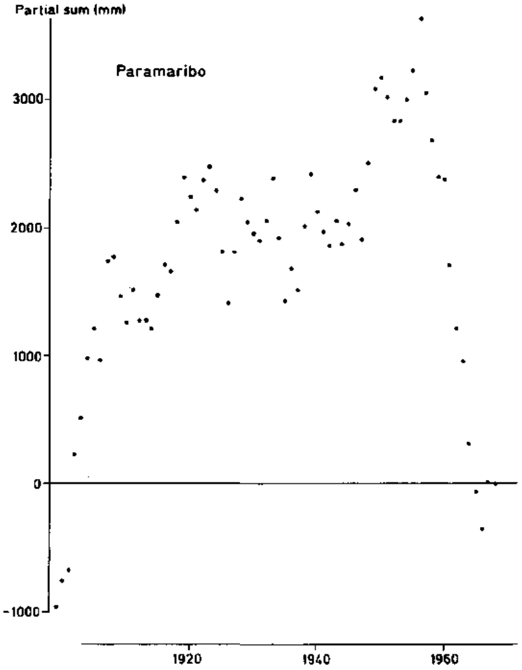
2.1. *Homogeneity*

Homogeneity was tested with the Von Neumann's ratio (see II, 3.1 and II, 6).

For annual totals the Von Neumann's ratio (d) is given in Table 2.2. This table also gives estimates of the mean, the standard deviation and the coefficient of skewness. Only for the series of Paramaribo does the Von Neumann's ratio show lack of homogeneity at the 5 per cent level. This is mainly a consequence of the low annual mean during the period 1955–1968, as is seen from the partial sums of departures from the mean (see II, 5.3), given in Figure 2.2. In nearly all cases there is no evidence for departures from normality. Only for the series of New Delhi is there evidence for positive skewness at the 5 per

TABLE 2.2. Mean (m), standard deviation (s), coefficient of skewness ($\hat{\gamma}$) and Von Neumann's ratio (d) of annual totals.

Station	m(mm)	s(mm)	$\hat{\gamma}$	d
Bangalore	916	184	0.307	1.181
Calcutta	1600	284	0.244	1.097
New Delhi	690	242	0.638	1.017
Jakarta-27	1790	291	–0.535	1.170
Pasar Minggu	2111	412	–0.250	0.853
Paramaribo	2198	361	0.013	0.767
Domburg	2087	357	0.442	0.850



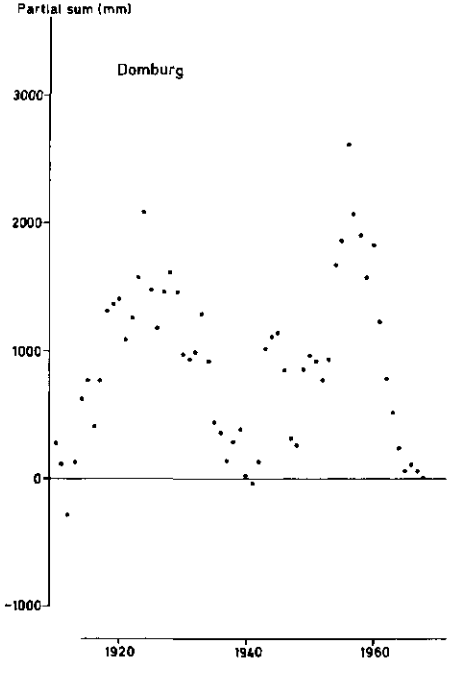


FIG. 2.2. Partial sums of departures from the mean of annual totals of Paramaribo and Domburg.

cent level. This phenomenon is characteristic for rainfall stations for which a few showers form the main contribution to the annual total (see Section 3).

Secular variations in the annual totals of India were assumed by Parthasarathy and Dhar (1974, 1975). Using about 3,000 rainfall stations, these authors found a significant positive difference of about 5 per cent between the means of the periods 1931–1960 and 1901–1930. There is, however, a large regional variation in this difference. For the rainfall series of Bangalore and New Delhi the mean of the period 1931–1960 is larger than the mean of the period 1901–1930. A two-sample Student test, however, gave no evidence for a difference in mean at the 5 per cent level (two-sided).

Jagannathan and Parthasarathy (1973) found a first serial correlation coefficient (scc) of –0.237 for Bangalore (1837–1967). This value supports the existence of serial correlation at the 5 per cent level. The fact that the Von Neumann's ratio is larger than 1 for Bangalore (see Table 2.2) is also an indication for a negative serial correlation.

Table 2.3 gives the Von Neumann's ratio for the annual number of wet days, together with estimates of the mean, the standard deviation and the coefficient of skewness. There is no evidence for non-normality at the 5 per cent level. The Von Neumann's ratio supports non-homogeneity for the series of Pasar Minggu ($\delta = 1$ mm), Paramaribo and Domburg ($\alpha = 0.05$). It should be noticed from Table 2.3 that the number of days with a rainfall amount between 1 and 3 mm is smaller for Pasar Minggu than for Jakarta-27. Besides at the lower threshold (1 mm) the number of wet days is smaller for Pasar Minggu than for Jakarta-27, though its annual mean is larger (see Table 2.2). The number of days with a rainfall amount between 0.3 and 2.2 mm is also much smaller for Domburg than for Paramaribo. So it can be concluded that the low values of the Von Neumann's ratio for Pasar Minggu ($\delta = 1$ mm) and Domburg ($\delta = 0.3$ mm) are due to man-made non-homogeneities. This is not so for the other Surinam wet-dry series. From the partial sums of departures from the mean of the wet-dry series of Domburg and Paramaribo, given in Figure 2.3, it is seen that the mean number of wet days is low during the period 1955–1968. For $\delta = 2.2$ mm the curves of Domburg and Paramaribo are of the same form.

Table 2.3. Mean (m), standard deviation (s), coefficient of skewness ($\hat{\gamma}$) and Von Neumann's ratio (d) of the annual number of wet days.

Station	δ(mm)	m	s	$\hat{\gamma}$	d
Bangalore	0.2	103.0	11.9	–0.196	1.068
Calcutta	0.2	119.2	9.9	0.180	0.913
New Delhi	0.2	52.2	11.2	0.371	0.885
Jakarta-27	1.0	128.4	17.6	–0.001	1.191
	3.0	95.0	12.9	0.003	1.072
Pasar Minggu	1.0	119.5	27.2	–0.294	0.582
	3.0	104.5	20.5	–0.278	0.798
Paramaribo	0.3	217.0	20.0	–0.488	0.758
	2.2	153.0	17.8	–0.316	0.762
Domburg	0.3	184.3	28.0	–0.222	0.570
	2.2	151.7	20.5	0.412	0.626

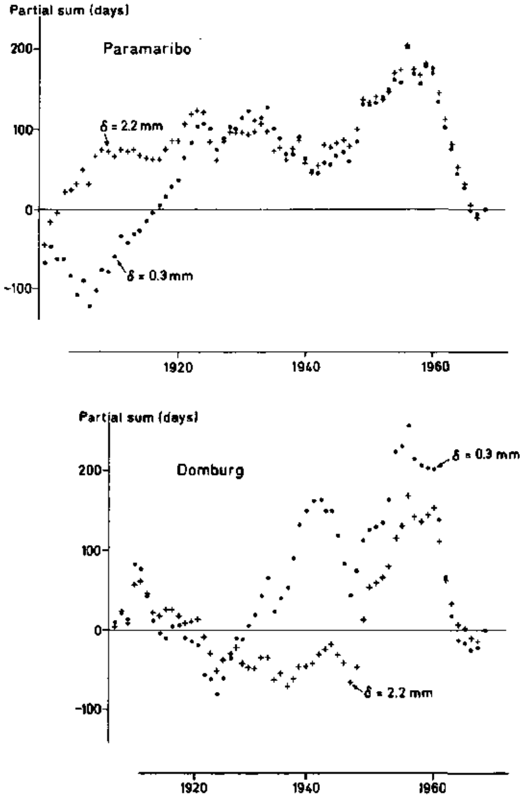


FIG. 2.3. Partial sums of departures from the mean of the annual number of wet days of Paramaribo and Domburg.

Moreover, their behaviour is similar to that of the curves of the annual totals (see Figure 2.2).

The adjusted ranges of the Surinam data are quite large. For instance, from Table 2.3 and Figure 2.3 it follows that the rescaled adjusted ranges of the wet-dry series of Paramaribo are 16.1 and 13.6 for $\delta = 0.3$ and 2.2 mm, respectively. From Figure 6.4 of Chapter II it is seen that these values are significant at the 5 per cent level (the length of the Paramaribo series is 70 years).

Since rainfall amounts smaller than δ mm are set to zero in the model, there is some reduction of the annual mean. For Jakarta-27 a threshold of 3.0 mm leads to a reduction of 46 mm (nearly 3 per cent); for Paramaribo a threshold of 2.2 mm gives a reduction of about 70 mm (3 per cent).

2.2. *The occurrence of wet and dry days*

For Dutch rainfall series it was shown (see III, 3) that the sequence of wet and dry days could be described by a seasonal alternating renewal process. Modifications of the negative binomial distribution were fitted to the lengths of weather spells. In this section it is investigated whether such a process can

TABLE 2.4. Lengths (in days) of the three longest dry spells and the three longest wet spells.

	Dry spells			Wet spells		
Bangalore A2	127	124	119	24	22	19
Calcutta A2	136	126	123	33	32	28
New Delhi A2	153	139	125	15	15	15
Jakarta-27 A10	105	84	67	24	23	22
A30	145	84	82	17	16	14
Paşar Minggu A30	156	115	87	13	13	13
Paramaribo A3	30	30	26	56	43	41
A22	49	42	35	24	20	19
Domburg A22	56	54	50	38	32	25

TABLE 2.5. Estimated correlation coefficients between lengths of wet and dry spells. Values differing more than $2/\sqrt{N}$ (N is the number of paired observations) from zero are denoted by an asterisk.

	June	July	August	September
Bangalore WD2	0.034	–0.043	–0.029	–0.170*
Calcutta WD2	–0.070	–0.059	0.028	–0.095
New Delhi WD2	–0.057	–0.063	–0.011	–0.022
	December	January	February	March
Jakarta-27 WD10	0.090	–0.038	–0.094	–0.053
WD30	–0.063	–0.008	–0.147	–0.017
Pasar Minggu WD30	0.034	–0.020	–0.022	–0.056
	Sdw	Sw	Swd	Sd
Paramaribo WD3	–0.106*	–0.023	–0.051	–0.074*
WD22	–0.074*	–0.055	–0.067*	–0.013
Domburg WD22	–0.086*	0.013	–0.073*	–0.048

also describe the behaviour of wet-dry sequences of rainfall data from India, Indonesia and Surinam.

To get some idea about the wet-dry processes of these series, lengths of the longest dry and wet spells are given in Table 2.4. The longest wet spells are always shorter than the longest dry spells, except for Paramaribo A3. Rainfall series from monsoon climates (India, Indonesia) are characterized by very long dry spells during the dry period.

As in III, 3.1 the adequacy of an alternating renewal process was tested with correlation coefficients between successive wet and dry spells. Estimated correlation coefficients of rainfall series analysed by wet-dry cycles (WD) are given in Table 2.5. For Indian and Indonesian data this table gives the estimated correlation coefficients for the wet months only. During the dry months of these series testing for an alternating renewal process with correlation coefficients between successive wet and dry spells is senseless, since these months are characterized by short wet spells (nearly all these spells have a length of 1 or 2 days) and occasionally very long dry spells. From Table 2.5 it is seen that most estimated correlation coefficients are negative, though only a few of them (denoted by an asterisk) differ more than twice their approximated standard deviation from zero.

These negative estimates are not a consequence of seasonal variations. For instance for Paramaribo WD3 the means of monthly estimated correlation coefficients are -0.111, -0.066, -0.079 and -0.084 for the seasons Sdw, Sw, Swd and Sd, respectively. These values correspond quite well to those given in Table 2.5.

Analysing the series by dry-wet cycles (DW) gives similar results and therefore the assumption of an alternating renewal process seems reasonable.

The goodness of fit of modifications of the negative binomial distribution to the distribution of lengths of weather spells was tested in the same way as in III, 3.2. Table 2.6 indicates significant values of the X^2-test of goodness of fit at the 5 and 10 per cent level for the truncated negative binomial distribution (TNBD), the geometric distribution (GD) and the logarithmic series distribution (LSD).

For dry intervals the following remarks can be made:

a. The TNBD fits well in nearly all cases. For Indian data there is lack of fit during the dry season because the frequency distribution of dry intervals

TABLE 2.6. Results of the X^2-test of goodness of fit for different distributions for lengths of weather spells. Critical levels in the interval (0, 0.05), (0.05, 0.10) and (0.10, 1) are denoted by *, (*) and blank, respectively. A question mark is used when there is no information (number of classes is too small for application of the X^2-test and/or the likelihood equations do not have a solution within the parameter space).

		Dry spells											
Month		1	2	3	4	5	6	7	8	9	10	11	12
Bangalore A2	TNBD	*								*			
	GD	*	*			(*)				*	*	*	*
	LSD	*	(*)	*	*	*	*	*		*			
Calcutta A2	TNBD						*			(*)		(*)	
	GD				*	*	*	*		(*)	*	(*)	
	LSD	*	*	*			*			*	*	*	*
New Delhi A2	TNBD							*			*		
	GD		*	*				*	*	*	*		
	LSD	*	*	*	*	*	*	*			*	*	*
Jakarta-27 A10	TNBD					(*)	(*)				?		
	GD				(*)	*	*	*	*	*	*		
	LSD	*		*	*		(*)	*					*

TABLE 2.6. (continued)

Month		1	2	3	4	5	6	7	8	9	10	11	12
Jakarta-27 A30	TNBD				(*)		*	*					
	GD		*		(*)	*	*	*		*			
	LSD	*		*	*		*	*		*	(*)		(*)
Pasar Minggu A30	TNBD						*						
	GD				*	*	*	*	*		*	*	*
	LSD	(*)			*	*	*		*	*		*	*
Paramaribo A3	TNBD	*		*			?						
	GD	*	*	*		*	*		(*)			*	*
	LSD	*		*					(*)	*	*		
Paramaribo A22	TNBD										*	*	(*)
	GD	*	*	*		(*)		(*)			*	*	*
	LSD		*		*		(*)	(*)	*	*	*	*	*
Domburg A22	TNBD												
	GD	*	*	*	*				*		*	*	*
	LSD		*					*	*	*	*	*	
							Wet spells						
Bangalore A2	TNBD		?	?									*
	GD		?	?		*		(*)	*	*			(*)
	LSD		?	?				*			*		*
Calcutta A2	TNBD	?		?				*				?	?
	GD						*	(*)				*	?
	LSD	*		*				*	*	*	*	*	?
New Delhi A2	TNBD	?	?	?	?	?					?	?	?
	GD		*				(*)		(*)			?	(*)
	LSD	*	*	*				*				?	*
Jakarta-27 A10	TNBD						?	?	?	?		*	
	GD		*						*		*	*	
	LSD			*		(*)	*		*			*	
Jakarta-27 A30	TNBD	*				?	?	?	?	?			
	GD	*	(*)			(*)	?	?	?		*		
	LSD	*		*			?	?	?				
Pasar Minggu A30	TNBD							?	?				
	GD	*	(*)	*								*	*
	LSD							*					
Paramaribo A3	TNBD				*								*
	GD			*	*							*	*
	LSD	*	*		*	*	*	*	*	*			*
Paramaribo A22	TNBD			*					(*)			(*)	
	GD			(*)					*			*	
	LSD	*	*	*	*	*	*	*	*		*	(*)	*
Domburg A22	TNBD					(*)							
	GD			*	*	(*)							*
	LSD	*			*	*	*					*	

TABLE 2.7. Frequency distributions of the length of dry spells.

Class with upper bound (in days)	10	20	30	40	50	60	70	80	90	100	∞
Observed frequencies for:											
Bangalore A2 (January)	18	4	7	8	1	5	6	7	4	0	0
New Delhi A2 (October)	14	13	8	2	0	5	3	4	6	6	3

sometimes has two tops. Two examples are given in Table 2.7. For the month of October of Jakarta-27 A10 the iterative solution of the likelihood equations by means of the Newton-Raphson procedure fails because the optimum value of r is close to zero. Then the TNBD is nearly equivalent to the LSD (see III, 3.2).

b. The GD and the LSD usually give a poor fit.

For wet intervals it can be concluded:

a. The TNBD fits well in nearly all cases. For Indian and Indonesian data there are some months during the dry season for which the likelihood equations of the TNBD have no solution within the parameter space. In all these months there are no wet spells longer than 5 days. Because of these troubles it would be desirable if a one parameter distribution (GD, LSD) could fit the distribution of wet intervals during the whole year.

b. For Indian data the LSD prevails over the GD for Bangalore, whereas for Calcutta the GD is preferred. For New Delhi the LSD fits poorly during the dry season. Therefore the GD might be preferred to the LSD for this station.

c. For Indonesian data the LSD gives a reasonable fit when a threshold of 3.0 mm is used. At this threshold the LSD is preferable to the GD, since the last distribution fits the data poorly for all months of the wet season of Pasar Minggu. For Jakarta-27 A10 the fit of the LSD is poor.

d. For Surinam data the LSD gives a poor fit. The fit of the GD is only slightly worse than the fit of the TNBD.

Instead of a TNBD for lengths of weather spells one can also think in terms of a shifted negative binomial distribution (SNBD), because of its simpler form. It is remarkable, however, that for dry spells of Indonesian and Surinam data the TNBD usually gives a slightly better fit. This is seen from the critical levels of the X^2-test of goodness of fit, given in Table 2.8.

2.3. *The behaviour of rainfall amounts on wet days*

In this section four features of the behaviour of rainfall amounts on wet days are discussed:

a. The correlation between the length of a dry period and the rainfall amount on the day following that period.

b. The correlation between rainfall amounts within a wet spell.

c. The marginal distribution of rainfall amounts.

d. The dependence of the distribution of the rainfall amounts on the day of occurrence.

TABLE 2.8. Critical levels of the X^2-test of goodness of fit for the SNBD and the TNBD, fitted to lengths of dry spells. For June of Paramaribo A3 the number of occupied classes is too small for a X^2-test.

Month	Pasar Minggu A30		Paramaribo A3		Paramaribo A22	
	SNBD	TNBD	SNBD	TNBD	SNBD	TNBD
January	0.549	0.726	0.025	0.042	0.003	0.566
February	0.899	0.977	0.260	0.350	0.049	0.194
March	0.764	0.964	0.000	0.034	0.372	0.633
April	0.156	0.448	0.836	0.770	0.135	0.397
May	0.000	0.184	0.965	0.975	0.156	0.224
June	0.000	0.008			0.053	0.133
July	0.960	0.940	0.108	0.464	0.197	0.258
August	0.072	0.335	0.233	0.204	0.225	0.512
September	0.314	0.332	0.276	0.559	0.303	0.347
October	0.015	0.200	0.329	0.274	0.002	0.002
November	0.142	0.136	0.247	0.686	0.012	0.016
December	0.137	0.105	0.216	0.445	0.000	0.094

TABLE 2.9. Estimated correlation coefficients between the rainfall amount on the first day of a wet spell and the length of its preceding dry spell. Values differing more than $2/\sqrt{N}$ (N is the number of paired observations) from zero are denoted by an asterisk.

	Sdw	Sw	Swd	Sd
Bangalore A2	-0.083	-0.037	-0.042	0.003
Calcutta A2	-0.021	0.057	-0.086	-0.044
New Delhi A2	0.043	-0.014	0.136*	0.010
Jakarta-27 A10	-0.024	0.037	-0.041	-0.030
A30	-0.011	0.034	0.011	-0.065
Pasar Minggu A30	0.033	0.094*	0.066*	-0.022
Paramaribo A3	-0.055*	-0.041	-0.053	0.028
A22	-0.064*	0.002	-0.045	0.034
Domburg A22	0.018	0.081*	0.003	0.080*

Estimated correlation coefficients between the rainfall amount on the first day of a wet spell and the length of its preceding dry spell are given in Table 2.9. The correlation coefficients were obtained in the same way as in III, 4.1. Most correlation coefficients are small; only a few of them (these are denoted by an asterisk in Table 2.9) differ more than twice their approximated standard deviation from zero.

Estimated first serial correlation coefficients of rainfall amounts within a wet spell are given in Table 2.10. For the wet season of India and Indonesia, and the season Sdw of Surinam the data usually support a small positive correlation between successive rainfall amounts, which differs significantly from zero. The existence of serial correlation is less obvious for the seasons Sw, Swd and Sdw of Surinam data.

TABLE 2.10. Estimated first serial correlation coefficients of rainfall amounts within a wet spell. Values differing more than $2/\sqrt{N}$ (N is the number of paired observations) from zero are denoted by an asterisk.

	June	July	August	September
Bangalore A2	0.116*	0.124*	0.128*	0.088*
Calcutta A2	0.249*	0.182*	0.136*	0.226*
New Delhi A2	0.190*	0.065	0.200*	0.060
	December	January	February	March
Jakarta-27 A10	0.166*	0.198*	0.097	-0.017
A30	0.106	0.200*	0.109	-0.001
Pasar Minggu A30	0.156*	0.160*	0.082	0.118*
	Sdw	Sw	Swd	Sd
Paramaribo A3	0.144*	0.015	0.018	0.062*
A22	0.073*	0.002	0.015	0.041
Domburg A22	0.124*	0.066*	0.080*	-0.035

In III, 5.2 the shifted gamma distribution (SGD) was fitted to the rainfall amounts on wet days, because for shifted rainfall amounts (rainfall amounts minus some positive value in order to make the lower bound of the carrier zero) the monthly mean of $\hat{\gamma}/\hat{C}$ (γ is the coefficient of skewness, see II, (2.1), and C is the coefficient of variation, see II, (3.10e)) was close to the theoretical value 2 of the gamma distribution. For the data analysed in this section the monthly means of this ratio are given in Table 2.11. The values given in this table are close to 2 and therefore the use of the SGD looks promising.

The parameters of the gamma distribution were estimated by the modified ML method, given in III, 5.2. and III, A1.

TABLE 2.11. Goodness of fit for the SGD for rainfall amounts. The first column gives the monthly mean of the quotient of the coefficient of skewness ($\hat{\gamma}$) and the coefficient of variation ($\hat{C}$). The other columns give the results of the X^2-test of goodness of fit. Critical levels in the interval (0, 0.05), (0.05, 0.10) and (0.10, 1) are denoted by *, (*) and blank, respectively.

	$\hat{\gamma}/\hat{C}$	1	2	3	4	5	6	7	8	9	10	11	12
Bangalore A2	2.05	*			*		*	*	*	*		(*)	
Calcutta A2	2.09	*	*		(*)			*	*	*	(*)		
New Delhi A2	2.22				*			*	*	(*)			
Jakarta-27 A10	1.81	(*)				*							
Jakarta-27 A30	1.81												
Pasar Minggu A30	1.96			(*)		*				(*)			
Paramaribo A3	2.12	(*)											
Paramaribo A22	2.14												
Domburg A22	2.11		(*)			*			*				*

For the parameter ε in III, (5.1) the value 1.0 mm was chosen for Indian series and 2.0 mm for Indonesian and Surinam series. This gives on the average a relative error of about 0.1 per cent in the integral of III, (5.2).

The fit of the SGD was tested in the same way as in III, 5.2. Frequency distributions of rainfall amounts were constructed to apply the X^2-test of goodness of fit.

The class boundaries were at 2.1(1.0)10.1(2.0)20.1(5.0)50.1(10.0)100.1(25.0)150 mm for Indian series and at 2.0(1.0)14.0(2.0)22.0(5.0)52.0(10.0)102.0(25.0)152.0 mm for Jakarta-27 A10. Notice that the class boundaries are chosen such that the first class contains all rainfall amounts of which the actual value was ignored in the modified ML procedure. For this reason the first two classes of Jakarta-27 A10 were omitted for Jakarta-27 A30 and Pasar Minggu A30. For Paramaribo A22 and Domburg A22 the classes were the same as for Indian data, except that the first class had its upper bound at 4.1 mm. The values of the boundaries of the Indian series were augmented by 0.1 mm for Paramaribo A3. The first class of this series had its upper bound at 2.2 mm.

In Table 2.11 it is indicated whether the X^2-test leads to significant values at the 5 and 10 per cent level. For Indian data the SGD gives a poor fit; for the other stations the fit looks reasonable.

When the SGD does not fit the data well, the empirical distribution is often so anomalous that it might be expected that commonly used probability distributions fail to fit the data.

In III, 4.2 the relation between the intensity and the length of wet spells was investigated. It was shown in that section that the mean intensity depends on the length of the corresponding wet spell. Differences of the mean intensities at various lengths of wet spells were mainly caused by the mean rainfall amount of a particular day depending on the number of adjacent wet days. Rainfall amounts on days with i adjacent ($i = 0, 1, 2$) wet days were called type i amounts.

An F-test, based on the regression model III, (4.1), was used to test equality of mean intensities at different lengths of wet spells. This test was also used for testing equality of mean intensities when the first and last day of wet spells (that is all type 0 and 1 amounts) were omitted.

The last test is less powerful than the first, because there is a reduction in the number of data. For the period June-September of Indian series the fraction of type 2 amounts ranges from 0.26 to 0.47 for Bangalore A2; from 0.53 to 0.69 for Calcutta A2 and from 0.18 to 0.36 for New Delhi A2. During the wettest months (November-March) the fraction of type 2 amounts ranges from 0.25 to 0.40 for Pasar Minggu A30. For Paramaribo A3 this fraction is 0.75 for the season Sw, but only 0.29 for the season Sd; for Paramaribo A22 these values are 0.50 and 0.15, respectively.

The equality of means of type 0, 1 and 2 amounts was also tested by an F-test.

Table 2.12 shows the results of testing at the 5 and 10 per cent level for rainfall series analysed by method A. For New Delhi and the seasons Sdw and Sw of the Surinam data, differences between mean intensities are found at the 5 per cent level, which usually disappear when only type 2 amounts are taken into

TABLE 2.12. Results of tests for dependence of the mean rainfall amount on a wet day and its position in the wet-dry sequence. The tests considered are:

1. An F-test for equality of mean intensities at different lengths of wet spells (cf. III, Table 4.2).
2. An F-test for equality of means of type 0, 1 and 2 amounts (cf. III, Table 4.3).
3. An F-test for equality of mean intensities of type 2 amounts at different lengths of wet spells, omitting the first and last day (cf. III, Table 4.5).

Critical levels in the interval (0, 0.05), (0.05, 0.10) and (0.10, 1) are denoted by *, (*) and blank, respectively.

	June			July			August			September		
	1	2	3	1	2	3	1	2	3	1	2	3
Bangalore A2				*		*	*	(*)	*	*	*	
Calcutta A2	*	*			(*)			*		*	*	
New Delhi A2	*	*	*	*	*		*	*		*	*	
idem, cube root	*	*	*	*	*		*	*		*	*	
	December			January			February			March		
	1	2	3	1	2	3	1	2	3	1	2	3
Jakarta-27 A10	*	(*)		*	*	*		*			*	
A30	(*)	(*)		*	*		*	*				
Pasar Minggu A30	*		*	*		*						
	Sdw			Sw			Swd			Sd		
	1	2	3	1	2	3	1	2	3	1	2	3
Paramaribo A3	*	*	*	*	*		*	(*)	*	(*)		
A22	*	*		*	*							
Domburg A22	*	*			*				*			

account. There are also significant differences in the means of different types of rainfall amounts for these series. For New Delhi the same results are obtained with a cube root transformation on the data. For the other rainfall series and the seasons Swd and Sd of the Surinam data non-homogeneity of mean intensities and differences between different types of rainfall amounts are less obvious.

Sometimes there are significant differences between different types of rainfall amounts, but no significant departures from homogeneity at various lengths of wet spells (Calcutta A2 (July) and Jakarta-27 A10 (February, March)). In these three cases only the mean of type 2 amounts is significantly larger than the mean of type 0 and 1 amounts, which can be tested by a two-sample Student test (see III, 4.2). No differences in mean intensities are found since the fraction of type 2 amounts is only slightly more than 0.30 here.

For Bangalore A2 (July) and Pasar Minggu A30 (January and February) there are no significant differences between means of type 0, 1 and 2 amounts. However there is evidence for dependence between the mean rainfall amount

TABLE 2.13. ML estimates of ν (shape parameter) and ν/λ (shifted mean) of the SGD fitted to rainfall amounts (mm) of different types.

		Type 0		Type 1		Type 2	
		$\hat{\nu}$	$\hat{\nu}/\hat{\lambda}$	$\hat{\nu}$	$\hat{\nu}/\hat{\lambda}$	$\hat{\nu}$	$\hat{\nu}/\hat{\lambda}$
Bangalore A2	Sw	0.375	6.4	0.437	6.5	0.516	7.6
Calcutta A2	Sw	0.447	11.0	0.545	11.5	0.592	15.9
New Delhi A2	Sw	0.541	8.5	0.567	14.4	0.628	19.2
Jakarta-27 A10	Sw	0.620	11.0	0.614	11.6	0.684	22.5
A30	Sw	0.706	12.1	0.724	17.1	0.843	25.8
Pasar Minggu A30	Sw	0.915	17.0	0.900	17.1	0.881	18.6
Paramaribo A3	Sdw	0.288	4.0	0.479	7.4	0.643	12.1
	Sw	0.570	6.3	0.711	11.5	0.777	12.0
Paramaribo A22	Sdw	0.619	7.8	0.712	10.9	0.792	15.5
	Sw	0.692	10.0	0.830	13.0	0.889	13.2

and the length of the corresponding wet spell because of a small number of long spells with high intensity. For instance, for Pasar Minggu A30 the January mean of the rainfall amount on a wet day is 20.8 mm (see Figure 2.1) and its standard deviation is 19.5 mm, but at a length of 12 days two spells are found in this month with a mean of 36.0 mm.

Table 2.13 gives estimated parameters of the SGD fitted to different types of rainfall amounts. As for Dutch series (see III, 5.2), the mean and shape parameter usually increase with the number of adjacent wet days. This increase is most obvious for the series of New Delhi and Paramaribo. For New Delhi A2 and Paramaribo A3 (Sdw) the mean of type 0 amounts is less than 50 per cent of the mean of type 2 amounts. A remarkable point is the large difference between Jakarta-27 A30 and Pasar Minggu A30. The behaviour in time of the mean rainfall amount on a wet day is also different for these series (see Figure 2.1).

Significant differences were found (even at the 1 per cent level) between the means of the first and last day of a wet spell for the season Sdw of Paramaribo. The means of the shifted rainfall amounts of the first and last wet day of spells longer than two days are 8.6 and 6.4 mm, respectively, for Paramaribo A3. For Paramaribo A22 these values are 12.0 and 9.6 mm, respectively.

2.4. *Persistence*

The rainfall series analysed in the previous sections show positive serial correlation. This is indicated by the wet-dry process not being describable by a Bernoulli process (see 2.2) and by rainfall amounts within a wet spell usually being correlated (see 2.3).

Serial correlation coefficients were estimated in the same way as in III, 6.1. Both for the wet-dry process and the entire rainfall process estimated first sccs are given in Table 2.14. For Indian and Indonesian series first sccs are given

TABLE 2.14. Estimated and theoretical first sccs. The theoretical sccs are based on models with TNBDs for wet and dry intervals and SGDs for rainfall amounts on wet days. For Paramaribo ($\delta = 2.2$ mm) the first sccs between brackets are means of monthly values.

			Wet-dry process		Rainfall process				
Station	δ(mm)	Season	Estimated	Model	Estimated	Model I	Model II	Model III	Model IV
New Delhi	0.2	July-August	0.346	0.344	0.202	0.090	0.142	0.160	0.222
Jakarta-27	1.0	Sw	0.217	0.238	0.252	0.049	0.137	0.147	0.249
	3.0	Sw	0.209	0.231	0.239	0.075	0.137	0.190	0.264
Pasar Minggu	3.0	Sw	0.221	0.220	0.155	0.085	0.132	0.103	0.153
Paramaribo	0.3	Sdw	0.329	0.323	0.231	0.058	0.141	0.145	0.236
	2.2	Sdw	0.319	0.314	0.228	0.115	0.141	0.205	0.235
			(0.314)	(0.310)	(0.224)	(0.112)	(0.138)	(0.198)	(0.229)
	0.3	Sw	0.132	0.146	0.059	0.017	0.033	0.040	0.056
	2.2	Sw	0.134	0.131	0.067	0.037	0.040	0.063	0.066
			(0.129)	(0.129)	(0.064)	(0.037)	(0.040)	(0.062)	(0.066)

for the wettest months; for Paramaribo these values are given for different seasons. For Paramaribo ($\delta = 2.2$ mm) the influence of seasonal variation was investigated by comparing seasonal estimates with means of monthly estimates. The differences between these estimates are negligible which supports the estimation on a seasonal base. From Table 2.14 it is seen that the first scc of the entire rainfall process is small for the wet season (Sw) of Paramaribo and also for the seasons Sd and Sdw of this series. A remarkable point is the large difference between the first sccs of the entire rainfall processes of Jakarta-27 and Pasar Minggu. For Jakarta-27 the first scc of the entire rainfall process is larger than the first scc of the wet-dry process.

For different models one can calculate theoretical first sccs. With respect to the behaviour of rainfall amounts theoretical values were calculated for models I, II, III and IV, defined in III, 6.1. Throughout this section it is assumed that the wet-dry process is a seasonal alternating renewal process with TNBDs for lengths of wet and dry spells (TNBD-TNBD process, see III, 6.1). Further, the assumption is made that the rainfall amounts have a SGD with seasonally changing parameters.

Seasonal values of the parameters were obtained by averaging monthly A and B estimates. The only difference from Section III, 6 is, that both parameters of the TNBD were seasonally changing for the wet-dry process. There can be large differences between ML estimates of the parameters of the TNBD for consecutive months as was the case for Dutch series (see III, Figure 3.1). It might be questionable, therefore, whether the mean of monthly estimates is a good seasonal estimate.

Table 2.14 shows a nice correspondence between the theoretical values and the estimated values for the wet-dry process; for the entire rainfall process this is only true for model IV, as was also the case for Dutch series (see III, 6.1). The height of the threshold δ hardly influences the theoretical first sccs.

TABLE 2.15. Estimated and theoretical variances of the number of wet days and of the rainfall amount for a period of considerable length. For Paramaribo the length of the period is 30 days, whereas for New Delhi and Indonesian series the length of the whole season is taken. Theoretical values are based on models with TNBDs for wet and dry intervals and SGDs for rainfall amounts on wet days. For Paramaribo ($\delta = 2.2$ mm) the variances between brackets are means of monthly values. Variances of the rainfall process are in cm^2.

			Wet-dry process		Rainfall process				
Station	δ(mm)	Season	Esti- mated	Model	Esti- mated	Model I	Model II	Model III	Model IV
New Delhi	0.2	July-August	57.8	45.7	361	207	220	244	261
Jakarta-27	1.0	Sw	42.8	30.7	680	273	312	381	426
	3.0	Sw	50.7	30.3	680	298	326	398	430
Pasar Minggu	3.0	Sw	56.1	30.9	320	223	238	231	246
Paramaribo	0.3	Sdw	29.3	17.7	120.8	48.4	54.5	60.9	68.6
	2.2	Sdw	27.5	17.0		59.4	61.4	71.5	73.9
			(28.6)	(16.9)	(123.6)	(59.6)	(61.7)	(71.6)	(74.0)
	0.3	Sw	7.8	6.7	88.5	57.7	59.3	60.9	62.6
	2.2	Sw	9.6	9.6		60.7	61.0	63.8	64.2
			(12.7)	(10.0)	(89.4)	(61.8)	(62.2)	(65.3)	(65.7)

For Paramaribo A22 the seasonally calculated sccs correspond quite well with the mean of monthly calculated values, which supports the averaging of monthly estimates of the parameters.

To get some idea about the tail of the correlogram, estimated and theoretical variances were compared for periods of a considerable length. For New Delhi the period July-August (62 days) was taken, whereas for Indonesian series the period January-February (59 days) was taken; for Paramaribo a 30-day period was considered. Estimated and theoretical values are given in Table 2.15.

For Paramaribo A22 it was investigated whether there are differences between a seasonal value and a value obtained by averaging monthly values (the estimated 30-day variance of e.g. January was obtained by multiplying the monthly estimate by 30/31). In general, the differences between the two values are small, except for the estimate of the wet-dry process for the wet season.

From Table 2.15 it is seen that the model underestimates the variances. However, it should be emphasized that the standard deviation of the estimate is large. For instance, the estimated values of New Delhi and Pasar Minggu are based on 70 and 69 observations, respectively. It follows from III, (6.2a) that the standard deviation of $\underline{s}^2$ should be approximately 0.17 σ^2 when the distribution of the rainfall amount of the wet period is assumed to be Gaussian. Because of departures from normality the standard deviation is usually a bit larger than this value (see III, 6.2). Therefore the large differences between the estimated and theoretical values can still be acceptable.

Figure 2.4 compares cumulative frequencies and the cumulative distribution function (cdf) of model I for the wet season of New Delhi. There is a good correspondence between the two curves even though the variance of model I is much smaller than the estimated value. Something similar holds for Pasar Minggu, which will be shown in the next section.

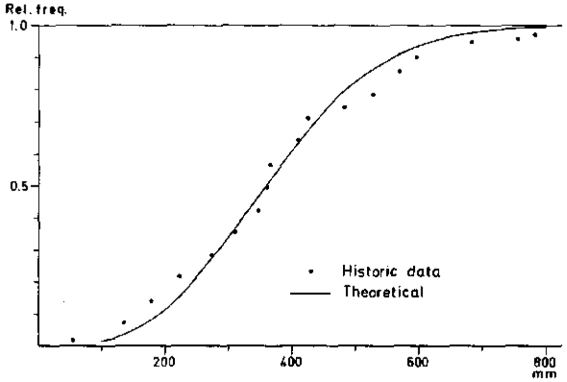


FIG. 2.4. Cumulative frequencies of the rainfall amount in the period July–August (62 days) for the historic series of New Delhi and a calculated cdf. The theoretical curve is based on a TNBD–TNBD process and a SGD for rainfall amounts on wet days with a threshold of 0.2 mm.

In fact, only for the Paramaribo series is there a serious underestimation of the variance during the season Sdw.

For this season the variance-time curves of both the wet-dry process and the entire rainfall process are approximately straight lines on double logarithmic paper. This is seen from Figure 2.5 where estimated variance-time curves are given for Domburg and Paramaribo ($\delta = 2.2$ mm). The same result is obtained at a threshold of 0.3 mm. A straight line on double logarithmic paper means that the correlogram falls off very slowly. For Paramaribo DW3 sccs up to lag 35 were estimated. Both for the wet-dry process and for the entire rainfall process all 35 values were positive. Asymptotically a straight line of the variance-time curve

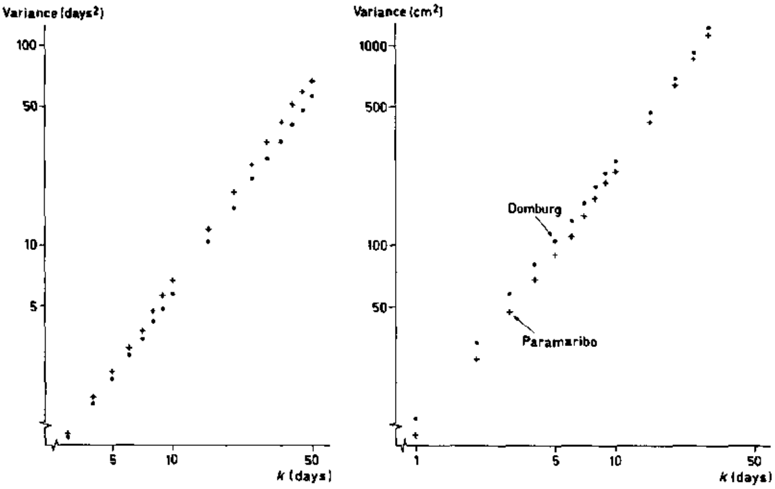


FIG. 2.5. Estimated variances of the number of wet days and the rainfall amount for a k-day period in the Sdw season for Paramaribo and Domburg ($\delta = 2.2$ mm). The estimates are given for different values of k and are based on data with a (k–1) day overlap.

on double logarithmic paper is theoretically possible, e.g. if the wet-dry process is a renewal process with an infinite variance (and higher order moments) of the recurrence times (see IV, 5.1).

2.5. *Comparison of historic and synthetic data for Pasar Minggu*

In the previous sections features of the rainfall process were often investigated for the wet season only when dealing with series from India and Indonesia. Hardly any attention was paid to the dry season and transition periods between wet and dry monsoons. Especially during the transition periods it is difficult to evaluate the fit of a certain model by statistical tests or numerical computations, since there is a strong seasonal variation. However, features of a certain model can often be compared with those of the historic sequence by Monte Carlo simulations.

In this section some synthetic sequences for Pasar Minggu are compared with the historic sequence after a modification (see below). The series of Pasar Minggu was preferred to the Indian series because it is seen from the results of previous sections that a stochastic model underlying this series can have a simple form. There are hardly any differences between type 0, 1 and 2 amounts (see Table 2.13), the SGD provides a good fit to the rainfall amounts on wet days (see Table 2.11) and the TNBD fits lengths of dry spells well (see Table 2.6). Only for wet spells during the dry season (east monsoon) are there sometimes troubles with the solution of the likelihood equations of the TNBD (see 2.2 and Table 2.6). Therefore the LSD was taken for the distribution of wet spells. In the wet season the fit of this distribution was only slightly worse than the fit of the TNBD.

A model with a seasonal LSD-TNBD wet-dry process and rainfall amounts with a SGD can describe the distribution of the total rainfall amount during the wet season (Sw) quite well, as is shown in Figure 2.6. Theoretical cdfs are based on estimated parameters for January or February or on the average of these estimates. There is only a small difference between the three theoretical curves. If a TNBD is taken for the distribution of wet spells instead of a LSD one gets cdfs which are indistinguishable from those given in Figure 2.6.

Two independent series of 70 years were generated. The synthetic series are denoted by S1 and S2, respectively. A third series (S2′) was generated with the same wet-dry process as S2, but with other rainfall amounts. The first year of the synthetic sequence is discarded in the comparison with the historic series; so the historic series and synthetic sequences have the same length.

There are 5 parameters in the generation model, namely 1 for the LSD of the wet intervals, 2 for the TNBD of the dry intervals and 2 for the SGD of the rainfall amounts. Because of seasonal variation the value of a certain parameter was changed from month to month. In contrast with the generation model for Dutch series, described in III, 7.2, the estimates of the parameters were not smoothed. The month to which the first day of a weather spell belongs determines its distribution; during a wet spell the parameters of the SGD are constant and are taken from the month to which the wet spell belongs.

For generating the LSD–TNBD wet-dry process use was made of a geometric approximation (see III, 7.2.1) for lengths of spells longer than 32 days. Gamma variates were obtained

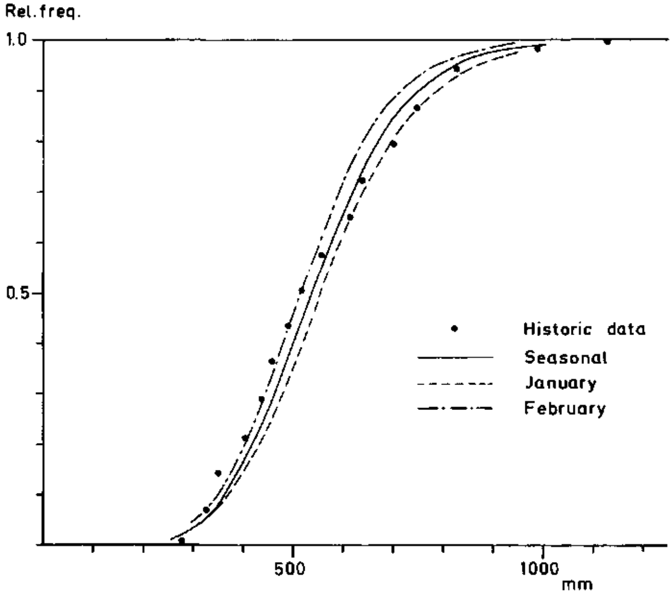


FIG. 2.6. Cumulative frequencies of the rainfall amount in the period January–February (59 days) for the (modified) historic series of Pasar Minggu and calculated cdfs. The theoretical curves are based on a LSD–TNBD process and a SGD for rainfall amounts on wet days with a threshold of 3.0 mm.

by using JÖHNK's (1964) algorithm. For the synthetic series the length of the month of February is taken to be 28 days.

For a comparison of features of the historic sequence with those of the model, special attention is paid to the seasons Swd, Sd and Sdw. In the historic series values of 1 and 2 mm are set to zero and the 29th of February is discarded (modified series). At the end of this section annual totals are compared.

Figure 2.7 gives the cdfs of the rainfall amounts of the dry season (Sd) of the historic series and of the synthetic series S1 and S2. The fit of the model is poor for small rainfall amounts because of the strange shape of the cdf of the historic sequence for rainfall amounts smaller than 120 mm. There are also considerable differences between the cdfs of the two synthetic series. However, because of the small number of data (69 years) these differences are acceptable, as can be shown by the Smirnov test. On the average the cdfs of the synthetic series correspond quite well to the calculated cdf, given in Figure 2.8. This figure also gives the cdf of a model with a LSD for the distribution of dry spells. The fitted LSD has a longer tail than the fitted TNBD and, hence, small rainfall amounts are more probable for this model. The difference between the two theoretical cdfs is only small.

For the wet season (Sw) the cdfs of the synthetic series fit well, as was to be expected from Figure 2.6. For this season there is a large difference between the estimated variances of the

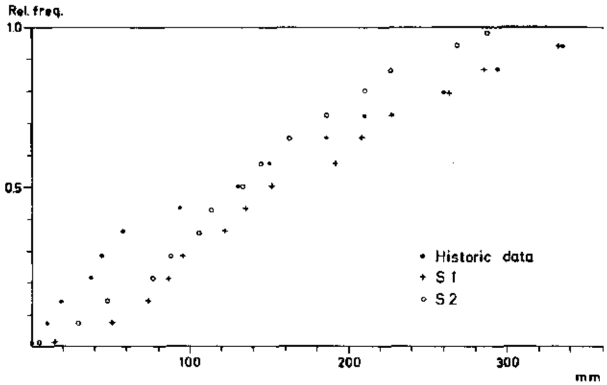


FIG. 2.7. Cumulative frequencies of the rainfall amount in the period July–August (62 days) for the (modified) historic series of Pasar Minggu and two (S1, S2) synthetic series, based on model I.

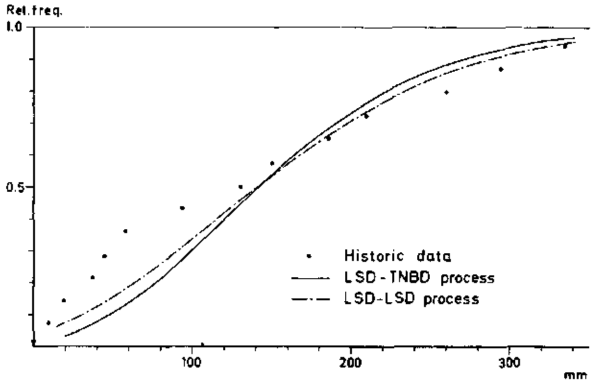


FIG. 2.8. Cumulative frequencies of the rainfall amount in the period July–August (62 days) for the (modified) historic series of Pasar Minggu and calculated values, based on model I with a SGD for rainfall amounts on wet days ($\delta = 2.2$ mm).

series S1 and S2. The estimated variances are 331 and 202 cm^2, respectively. The first value corresponds quite well to the value 320 cm^2 given in Table 2.15. Differences between these estimated variances can be explained by differences in the right tail of the empirical distribution. For the (modified) historic series the three largest rainfall amounts are 947, 982 and 1126 mm; for the series S1 these values are 963, 981 and 1033 mm, and for the series S2 these values are 800, 869 and 878 mm.

A criterion has to be found for a comparison of the historic series with synthetic series during the transition from the dry to the wet season. SCHMIDT and VAN DER VECHT (1952) considered the cumulative rainfall amount after August 31st. The day on which this cumulative rainfall amount exceeded the 350 mm level was defined as the beginning of the wet season. The value of 350 mm was chosen because then the soil is in general sufficiently wet for preparing the seed beds for the rice-crop in the wet season. The so defined beginning date of the

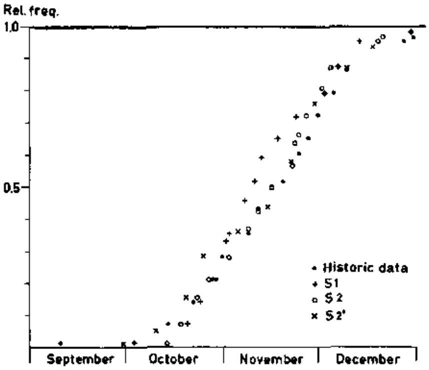


FIG 2.9. Cumulative frequencies of the beginning date of the wet season (west monsoon) of the (modified) historic series of Pasar Minggu and three (S1, S2, S2') synthetic sequences, based on model I. The first day of the wet season is defined as the day for which the cumulative rainfall amount after August 31st exceeds the 350 mm level.

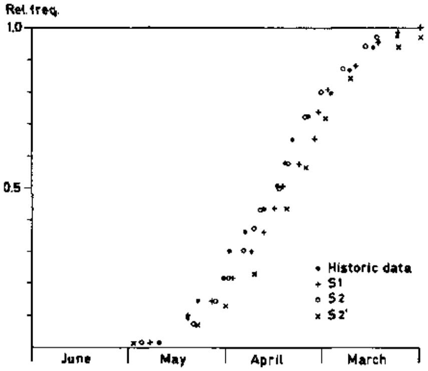


FIG. 2.10. Cumulative frequencies of the beginning date of the dry season (east monsoon) of the (modified) historic series of Pasar Minggu and three (S1, S2, S2') synthetic sequences, based on model I. The first day of the dry season is the day for which the cumulative rainfall amount from July 1st, backwards, exceeds the 350 mm level.

wet season, is used here for a comparison between the historic series and synthetic sequences during the transition from the east to the west monsoon. Figure 2.9 shows the cdfs of this date for both the historic sequence and the three generated sequences. In this figure the cdfs do not always reach the value 1, since it may happen that the 350 mm level is reached after December 31st. There is a reasonable correspondence between the cumulative frequencies of the historic series and those of the synthetic sequences.

An analogous method can be used to compare the historic series with the synthetic series during the transition from the wet to the dry season, by taking into account the cumulative rainfall amount, starting on July 1st, but going backwards in time (cf. SCHMIDT and VAN DER VECHT (1952)). The cdfs of the date on which the 350 mm level is reached are given in Figure 2.10 for both the

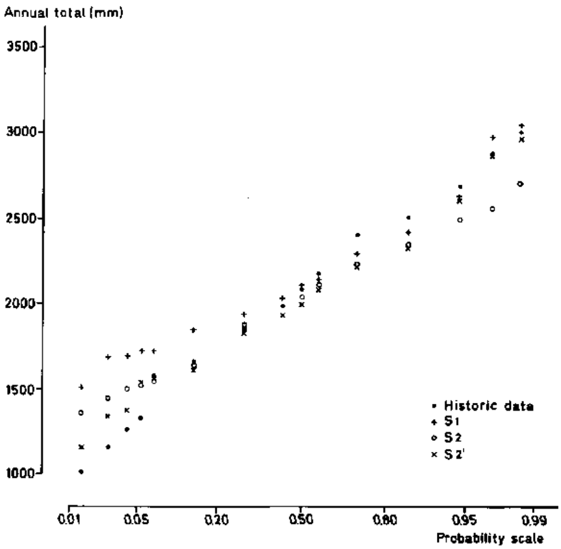


FIG. 2.11. Normal probability plots of annual totals of the (modified) historic series of Pasar Minggu and three (S1, S2, S2') synthetic sequences, based on model I.

historic series and the generated sequences. In this figure the cdf does not reach the value 1 when the 350 mm level is not reached before March 1st in at least one year. Figure 2.10 shows a reasonable correspondence between the cumulative frequencies of the historic series and those of the synthetic sequences.

Normal probability plots of annual totals are given in Figure 2.11 for the historic series and the synthetic sequences. The plots of the synthetic sequences have a smaller slope than the plot of the historic series, which indicates a smaller variance. The estimated standard deviations of the three synthetic sequences are 300, 304 and 365 mm, respectively. For the modified historic series this value is 411 mm, which corresponds quite well to the value 412 mm of the original historic series (see Table 2.2).

3. ANALYSIS OF RAINFALL DATA FROM SUDAN AND EGYPT

In this section daily rainfall data from Khartoum and Alexandria are analysed. In contrast to the rainfall series discussed in the previous section these series are characterized by some period with no rainfall at all. For Khartoum the months of December, January and February are always completely dry and for Alexandria there is no rain in July (except for some drops). In such a situation generating a wet-dry process by sampling alternately lengths of wet and dry spells can lead to serious problems. For instance, it is difficult to fit a probability distribution to the length of weather spells. This point is illustrated

TABLE 3.1. Critical levels of the X^2-test of goodness of fit for the TNBD and the GD fitted to the lengths of dry spells.

Month	Alexandria A1		Alexandria B1	
	TNBD	GD	TNBD	GD
November	0.674	0.740	0.000	0.000
December	0.108	0.133	0.284	0.219
January	0.647	0.666	0.327	0.411
February	0.762	0.003	0.205	0.138
March	0.002	0.000	0.115	0.167

in Table 3.1 where the TNBD and the GD are fitted to lengths of dry spells of Alexandria. For months in the wet period (January, December) these distributions fit well but in transition periods there is the problem that very long dry spells (even longer than 200 days) can occur (Alexandria A1, March and Alexandria B1, November). The frequency distribution of dry intervals has two peaks for these months, as was the case for some Indian series (see Table 2.7). For the drier months one has the problem that there are only a few spells beginning or ending in a particular month. Even months with no wet spells occur.

When generating synthetic sequences for such stations it can be worthwhile to start with generating lengths of the wet and the dry period or equivalently the first and last day of the wet season. For Khartoum the beginning of the wet season of a particular year is defined as the day on which the cumulative rainfall amount after January 1st reaches the 2.0 mm level for the first time. The end of the wet season of a particular year is defined as the first day after which the total rainfall amount is less than 2.0 mm. For Alexandria the first and last day of the wet season can be defined analogously with the exception that the cumulative rainfall amount is not counted from the beginning of a calendar year but from July 1st.

Table 3.2 shows realizations of the Von Neumann's ratio and estimates of the mean, the standard deviation and the coefficient of skewness for such quantities as the annual number of wet days, the annual total, the first and last day of the wet season, etc. A wet day is defined here as a day with a rainfall amount of at least 0.1 mm.

The annual means of both stations lie in the same order, but there is a difference with respect to the number of wet days. For Khartoum there are only a few wet days (on the average less than 20) for which the rainfall amount can be considerable. The annual total consists of a sum of a small number of rainfall amounts with a very skew distribution and its distribution is therefore also markedly skew. For Alexandria, however, there is a considerable number of wet days in a year and because of the central limit theorem the annual total is approximately normally distributed.

TABLE 3.2. Mean (m), standard deviation (s), coefficient of skewness ($\hat{\gamma}$) and Von Neumann's ratio (d) of the annual number of wet days (n_a), the annual total (t_a), the date (1 corresponds to January 1st) of the beginning and the end of the wet season (b_w and e_w, respectively), the length of the wet season (l_w), the average number of wet days during the wet season (n_w/l_w) and the average rainfall amount during the wet season (t_w/l_w) for the stations of Khartoum and Alexandria. The 29th of February is excluded.

	Khartoum				Alexandria			
	m	s	$\hat{\gamma}$	d	m	s	$\hat{\gamma}$	d
n_a	18.8	6.3	0.366	0.475	38.4	8.3	0.469	1.006
t_a (mm)	167.0	83.7	1.135	0.724	181.8	53.1	0.228	1.080
b_w	156.1	33.3	-0.314	0.878	302.5	19.4	-1.176	0.840
e_w	269.6	18.2	0.156	0.810	97.6	27.5	-0.171	0.824
l_w (days)	114.6	38.9	0.236	0.796	161.1	35.1	0.296	0.804
n_w/l_w	0.163	0.063	1.010	0.845	0.241	0.057	0.270	0.927
t_w/l_w (mm/day)	1.66	1.11	1.586	0.906	1.18	0.38	0.132	1.023

For the series of Khartoum the realizations of the Von Neumann's ratio of the annual number of wet days and of the annual totals give evidence for non-homogeneity or serial correlation at the 5 per cent level. However, there is no evidence for non-homogeneity or serial correlation when these quantities are divided by the length of the wet season. It is assumed, therefore, that the low values of the Von Neumann's ratio of the annual number of wet days and of the annual total are caused by persistence in the process underlying the beginning and the end of the wet season.

For Khartoum the beginning and the end of the wet season as well as the duration of the wet season are approximately normally distributed, whereas the distribution of the beginning date of the wet season of Alexandria is markedly skew. This phenomenon can occur, for instance, in a seasonally changing Bernoulli process.

During the wet season the SGD usually provides a reasonable fit to rainfall amounts on wet days. For Khartoum A1 the critical levels of the X^2-test of goodness of fit are 0.168, 0.153, 0.104, 0.287 and 0.401 for May, June, July, August and September, respectively. The critical levels for Alexandria A1 are 0.736, 0.313, 0.572, 0.031, 0.148, 0.022 and 0.055 for October, November, December, January, February, March and April, respectively.

The parameters of the SGD were estimated by the modified ML procedure, which was discussed in III, 5.2 and III, A1. The parameter ε in III, (5.1) was taken to be 0.5 mm for Alexandria and 1.0 mm for Khartoum. The upper bounds of the classes of the frequency distributions of the rainfall amounts were 0.6, 1.1(1.0)10.1(2.5)30.1, 40.1 and 50.1 mm for Alexandria A1. For Khartoum A1 the first class was omitted.

For Khartoum the mean rainfall amount on a wet day is seasonally changing. The means of May, June, July, August and September are 3.6, 5.8, 10.8, 11.9 and 6.8 mm, respectively. Seasonal variation of the mean is less obvious for Alexandria.

When the beginning and the end of the wet season are generated first, the application of TNBDs to lengths of weather spells is complicated. For instance, if one starts generating lengths of wet and dry spells from the first day of the wet season, one has the problem that the last day of the wet season is fixed. This problem does not arise when a Bernoulli wet-dry process is used during the wet season. This process can be applied when there is no evidence for serial correlation. For Khartoum the estimated first sccs of July and August are –0.014 and –0.001, respectively, and therefore the wet-dry process could be approximated well by a Bernoulli process. The estimated first sccs of Alexandria are 0.249 and 0.288 for the months of December and January, respectively, which shows evidence for serial correlation.

The estimated first scc of a particular month is based on all pairs of observations with a time-lag of 1 day in that month. The estimation technique used previously (see Section 2.4 and III, 6.1) can be cumbersome since dry spells can be quite long. The estimated correlation coefficients of Alexandria relate to the period July 1900–June 1940.

For Alexandria Figure 3.1 gives the empirical distribution of the rainfall amount in the period December–January and some calculated cdfs. The calculated cdfs are based on model I with a SGD for the distribution of the rainfall amounts on wet days and a Bernoulli or a GD-GD (first order Markov chain) wet-dry process. The GD-GD process is chosen here, because the GD fits the distribution of wet and dry intervals well for the months of December and January. For dry spells this is seen from Table 3.1; for wet spells the critical levels of the X^2-test of goodness of fit are 0.591 and 0.725 for December and January, respectively (method A).

As in the previous section the parameter values of the model were obtained by averaging monthly A and B estimates. All spells beginning (method A) or ending (method B) in January and December were taken into account.

The cdf fits the empirical frequencies well when a GD-GD process is as-

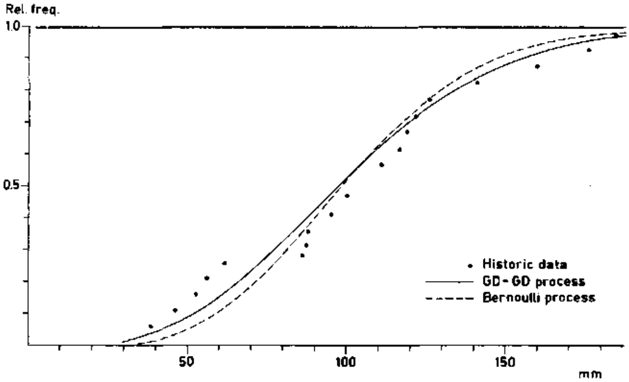


FIG. 3.1. Cumulative frequencies of the rainfall amount in the period December-January (62 days) for the historic series of Alexandria and calculated values, based on model I with a SGD for rainfall amounts on wet days (δ = 0.1 mm).

sumed for the sequence of wet and dry days. The fit becomes poorer when a Bernoulli wet-dry process is used, but the difference between the cdf based on this process and that based on a GD-GD process is only small.

The theoretical first scc of the model with a GD-GD process is 0.118. This value is much smaller than the estimated values given earlier. It should be emphasized that this value is based on model I which underlies the assumption of iid rainfall amounts within a wet spell. This assumption seems invalid here. For instance, the hypothesis of homogeneity of the mean intensity at various lengths of wet spells is rejected at the 5 per cent level by an F-test based on the regression model described in III, 4.2. When a more advanced model is used (e.g. model III) the difference between the estimated and theoretical first scc becomes smaller.

From the results of the analysis given above it can be concluded that synthetic data for Khartoum can be obtained as follows. First, the beginning and the end of the wet season are generated. Both the first and the last day of the wet season can be obtained from a normal distribution but serial correlation must be built in. For a model of serial correlation it seems advisible to extend the series with data after 1940. Second, the rainfall process during the wet season is generated. This process can be approximated by a Bernoulli wet-dry process with a SGD for the distribution of wet days. Both the probability of a day being wet and the mean of the SGD are seasonally changing.

For Alexandria a similar procedure can be followed. However, since there is no evidence for serial correlation in the first and last dates of the wet season it is not necessary to generate these dates first. An alternative procedure can consist of generating a wet-dry sequence by a first order Markov chain. During the dry period this Markov chain might be simplified to a Bernoulli process with a very small probability of a day being wet (cf. LOWRY and GUTHRIE (1968)). This last method gives a reasonable description of the wet-dry process. For the lengths of dry spells this is seen from Table 3.3. In this table cumulative

TABLE 3.3. The distribution of the lengths of dry spells for Alexandria ($\delta = 0.1$ mm). Cumulative relative frequencies are given for the historic series and 3 synthetic sequences.

Length (days)	Historic series	Synthetic series		
		S1	S2	S3
1	0.185	0.158	0.136	0.131
2	0.301	0.258	0.234	0.238
3	0.405	0.364	0.358	0.328
4	0.475	0.452	0.444	0.415
5	0.547	0.515	0.512	0.489
10	0.735	0.719	0.715	0.681
20	0.868	0.866	0.848	0.848
30	0.914	0.902	0.899	0.897
40	0.929	0.920	0.925	0.918
50	0.939	0.929	0.932	0.937
100	0.947	0.947	0.949	0.944
150	0.959	0.960	0.954	0.958
200	0.979	0.987	0.981	0.985

frequencies of these lengths are given for the historic series and for 3 synthetic sequences (denoted by S1, S2 and S3, respectively).

The transition probabilities of the first order Markov chain were estimated for each month separately. For a particular month the estimates of these probabilities were obtained by counting the number of wet-dry, dry-wet, dry-dry and wet-wet transitions in all pairs of successive observations beginning in that month. For July the probability of a day being dry was taken to be 1 and the probability of a day being wet was taken to be 0.

In the synthetic sequences the month of February always has 28 days.

The model only underestimates the number of short spells. This often occurs when dealing with geometric distributions (see III, 3.2).

4. Summary and concluding remarks

In this chapter rainfall sequences of some foreign stations were investigated. With respect to the behaviour of the rainfall process during the dry season one can distinguish:

a. Surinam data. For these series there is a considerable rainfall amount during the dry season. Dry spells are usually not much longer than those of Dutch series.
b. Indonesian data. Dry spells can be quite long for these series. There are, however, only a few months with no rain.
c. Indian data. These series are characterized by long dry spells during the dry season. There are many months without rain.
d. Sudanese and Egyptian data. For these stations there is even a period with no rainfall at all.

Generating wet-dry series by sampling alternately from distributions of wet and dry intervals gives problems for the last two categories, since it is sometimes hard to fit distributions to lengths of dry spells beginning in a certain month. An alternative is to describe the wet-dry series of such a rainfall process by a Markov chain of a certain order. For instance, for Alexandria a first order Markov chain describes the wet-dry series quite well.

With respect to the behaviour of rainfall amounts on wet days it can be concluded:

a. There usually exists a small positive serial correlation between successive rainfall amounts within a wet spell.
b. It is often necessary to discriminate between different types of rainfall amounts.
c. The shifted gamma distribution usually provides a good fit, except for Indian data.

As was the case for Dutch series the model usually underestimates variances of k-day totals for large values of k. But it should be emphasized again that the standard deviations of the estimators are quite large.

For the station of Pasar Minggu some synthetic sequences were generated based on a stochastic model with 5 parameters for each month. This model gives a good fit to the rainfall amount during the wet season and it can also describe the beginning of both the wet and the dry season reasonably well. The fit to the rainfall amount during the dry season is poor due to an anomalous shape of the empirical distribution function.

For the station of Khartoum it might be advisable to generate the beginning and the end of the wet season first, since there is some evidence for persistence in the lengths of successive wet and dry seasons. Within the wet season the rainfall process can be described by a Bernoulli wet-dry process with a shifted gamma distribution for the rainfall amounts on wet days. The parameters of this process show seasonal variation.

VI. REMARKS ON THE DATA

In this chapter it is described how the data, used in the previous chapters, were obtained. It is also indicated what was done with unreliable data and how missing data were supplemented.

With respect to the monthly and annual data, used for the analysis of homogeneity of Dutch stations (see II), the following remarks can be made:

a. Belgian series. Annual totals of Ghent Observatory were obtained from the University. The data of the Belgian national network were supplied by the Belgian Royal Meteorological Institute. For the rainfall station of Moerbeke Waas rainfall data from a nearby station (Oorderen) were used to get an uninterrupted sequence.

b. German series. Annual totals up to and including 1965 were supplied by the 'Deutsche Wetterdienst'. The other annual totals and monthly totals were obtained from yearbooks (nearly all these books are available in the library of the KNMI).

c. Dutch series. Monthly totals up to and including 1970 were obtained from notebooks in the archives of the KNMI. The other monthly and annual totals were obtained from yearbooks of the KNMI. However, the series of Zwanenburg-Hoofddorp (1735–1944) and Leyden (1736–1786) were adopted from LABRIJN (1945). The other rainfall totals of these series were taken from yearbooks of the KNMI. The rainfall series of Den Helder (see II, Figure 5.3) was taken from LABRIJN (1948).

Annual totals were always rounded to the nearest millimeter, also monthly totals of German stations and of Hoofddorp. The monthly totals of the other Dutch stations are in tenths of millimeters.

Daily rainfall sequences of Dutch stations were obtained in the following way. The data for the period 1953–1971 were copied from magnetic tapes of the KNMI. The daily data of Winterswijk and Hoofddorp for the period before 1953 came from copies of the punch cards on which the frequency tables of the KNMI (1956) are based. These frequency tables also give information about missing data for this period. Values for the 29th of February and the 31st of May, July, August, October and December, which were not given on the original punch cards, were inserted at a later stage. The other daily data were obtained from yearbooks of the KNMI. Gaps in the series of Heino (October 1954), Lochem (June-July 1968) and Hengelo (April-May 1910, June-November 1912, March-May 1919) were filled up with rainfall data from nearby stations (Vilsteren, Markelo and Enschede, respectively). A survey of supplements and corrections of incomplete and unreliable data of Hengelo (December 1907–1940) is given in Appendix A1. For the period after 1940 corrections were carried out by the KNMI and are marked on their magnetic tapes (after 1952) and their punch lists. From 1968, onwards, supplements and corrections are also given in their yearbooks.

For daily data of Indian series use was made of duplicates of punch cards of the Indian Meteorological Office. Missing observations (denoted by a blank) were assumed to be zero. For the period 1879–1900 monthly totals of Bangalore (see II, Figure 3.3) were obtained from a publication of the SMITHSONIAN INSTITUTION (1927). The other monthly totals were obtained by adding daily totals. Rainfall observations in inches were converted to millimeters.

Daily data of Jakarta-27 and Pasar Minggu were based on photocopies of the original observations. Data of Jakarta-27 were substituted for missing observations of Pasar Minggu (November 1958, August 1959). During December 1945 there were no daily observations for Jakarta-27. Only the total rainfall amount of this month (108 mm) was given. Because there were no rainfall observations of neighbouring stations available, the data of this month were taken identical to those of December (1960), except for the first day on which the rainfall amount was assumed to be 15 mm (is 16 mm for December 1960) in order to get a correct monthly total.

Daily observations of Domburg and Paramaribo were taken from books available in the library of the KNMI. Missing observations of Domburg were completed by data from Brownsweg (November 1916, November 1929) or Moengo (October 1955). Supplements of incomplete data of Domburg are given in Appendix A2.

The observations of Khartoum and Alexandria were obtained from yearbooks. Most of these books are available in the library of the KNMI. The following remarks can be made with respect to incomplete data of Khartoum:

a. The rainfall amount on the 17th of July 1902 was assumed to be zero.
b. The rainfall amounts on the 21st and the 22nd of July 1932 were assumed to be both 5.1 mm.
c. The rainfall amounts on the 1st and the 2nd of August 1932 were assumed to be 0 and 19.3 mm, respectively.

APPENDICES

A1. Supplements and corrections for Hengelo (December 1907–1940)

This appendix contains values of the series of Hengelo which differ from those given in the yearbooks of the KNMI. The reason for another value can be that the wrong date was recorded for a rainfall observation. It is also possible that the rainfall amount was only observed for a successive number of days. Then the rainfall amount was split up proportional to the daily values of a nearby station (Almelo, Borne, Denekamp, Enschede or Enter). The supplemented and corrected values are given in the next table.

Year	Month(s)	Days	Rainfall amounts (mm)
1908	5	6, 7	0, 6.2
1909	5	18–26	10.4, 0, 0, 0, 0, 0, 0, 0, 2.6
	12	8–11	2.6, 0.4, 0, 0
1910	2	19, 20	5.2, 0
	11	21–23	1.2, 0.2, 0
	12	4, 5	0, 5.1
		19–22	0.3, 4.0, 0.4, 0
1911	1	2, 3	7.3, 2.2
	6,7	30, 1	3.9, 2.1
1912	2	20–23	0, 3.5, 2.7, 9.3
	3	8–20	2.8, 0, 0, 0, 0, 0, 2.4, 0, 2.7, 0, 2.2, 2.4, 2.7
	4	4–6	0, 6.0, 7.3
	5	3–9	4.3, 0, 0, 0, 7.9, 14.9, 2.5
		12, 13	0, 16.7
1915	3	12, 13	0.4, 0.8
1916	2	23–28	0.4, 6.4, 3.7, 1.4, 0, 1.4
	12	22, 23	1.4, 12.6
1919	3	30, 31	3.0, 1.5
	4	13, 14	11.0, 0
		27, 28	15.3, 3.8
1920	4	11, 12	3.4, 11.3
	8	7, 8	5.4, 0
	12	25–29	7.4, 0, 0, 2.9, 3.3
1922	7	27, 28	2.6, 0
	8	1–8	0.2, 0, 1.3, 4.5, 13.6, 0, 0.2, 2.9
	12	25, 26	0, 4.5
		28–31	6.7, 2.1, 6.7, 0
1923	7	24–31	3.0, 0, 9.8, 3.3, 0.2, 6.3, 0.1, 0.6
	12	24–27	0.7, 6.3, 2.0, 0
1924	1	2–5	0.1, 0.3, 4.4, 0
		8, 9	0, 1.2
	10	8, 9	4.8, 4.4
		27, 28	4.9, 6.8
1925	11	28–30	5.4, 1.3, 0
	12	25–28	6.4, 0.1, 4.6, 3.9

Year	Month(s)	Days	Rainfall amounts (mm)
1926	1	18–20	3.8, 0, 0
	8	11–15	4.4, 2.3, 11.7, 5.7, 0
1927	1	20–24	0.1, 0, 5.3, 0, 0.1
	7	3, 4	4.5, 2.0
	8	5–10	0, 0, 7.6, 0, 4.1, 0
		14, 15	7.8, 8.3
	12	25–27	3.0, 14.8, 3.0
1928	2	14, 15	9.0, 2.4
	4	3, 4	0, 7.5
	8	1–5	0.9, 29.0, 0, 0, 6.9
	12	23–31	1.3, 9.6, 4.0, 3.8, 6.7, 0.1, 1.5, 1.9, 9.9
1929	1	24–26	0.1, 0.6, 1.1
	3	3–7	0, 1.3, 0, 0.8, 1.4
	8	1–4	13.4, 2.3, 0, 0
1930	1	3, 4	3.2, 2.5
		12, 13	11.1, 4.0
	10	12, 13	3.0, 0
	12	25–28	0.7, 2.6, 6.2, 5.3
1931	1	1, 2	7.2, 0
		14–16	3.3, 0, 2.7
	2	3–5	4.3, 0, 0.2
	5	27, 28	0, 17.0
	6	7, 8	4.6, 3.0
	7	8, 9	4.1, 3.2
		17, 18	9.5, 12.5
	11	6–8	7.0, 0, 0.5
	12	4–6	5.0, 4.4, 0
1932	1	10–14	0, 0, 0.1, 0.2, 4.3
	2	19, 20	0, 0.7
	6	4, 5	0, 0.3
	7	26, 27	2.1, 6.5
	9	26, 27	1.0, 2.5
	12	25, 26	3.0, 1.6
1933	2	16–19	1.7, 2.2, 0.2, 0.2
	6	17–20	0, 17.8, 1.1, 7.6
	10	8, 9	0, 8.5
		15–18	0, 0.2, 1.6, 5.0
	12	28–30	1.5, 0, 0
1934	8	5, 6	34.8, 7.2
	12	9, 10	0, 5.5
1935	1	16, 17	1.6, 0.4
	3	23, 24	1.2, 12.0
	4	7, 8	4.6, 4.4
	5	20–22	1.2, 4.0, 3.0
	6	10–12	0, 29.0, 0.5
	10	21–26	4.0, 0, 0.3, 0, 0, 2.2
1936	1	19, 20	2.2, 0
	2	16, 17	9.4, 3.5
	10	19–21	0, 11.3, 2.0
1937	3	1, 2	0, 1.5

Year	Month(s)	Days	Rainfall amounts (mm)
1938	11	20, 21	2.6, 0.9
	12	8–10	0.1, 3.5, 1.9
1939	7	7–9	0, 2.8, 1.2
	11	24–26	0.5, 4.3, 12.6
	12	28–31	0.4, 0, 0.2, 9.6
1940	4	17, 18	2.2, 0
	12	18–20	2.6, 7.7, 3.7
		27, 28	7.3, 0

The large number of Christmas days in this table is remarkable.

A2. Supplements for Domburg (December 1909–1968)

For Domburg there are some days for which the observed rainfall amounts relate to a number of successive days. The values obtained by splitting up these accumulated rainfall amounts are given in the next table.

Year	Month(s)	Days	Rainfall amounts (mm)
1912	6	1–30	39.5, 8.5, 2.3, 5.3, 1.9, 0.6, 5.7, 24.4, 19.9, 10.2, 0, 0, 0.8, 0, 13.1, 9.1, 72.3, 0, 0, 0, 0, 0.8, 6.2, 0, 0, 22.8, 0.8, 16.2, 5.1, 0
1912	10	1–31	0, 0, 3.2, 18.8, 0, 0, 0, 0, 0, 7.8, 11.5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.9, 2.3
1956	12	24, 25	6.7, 6.6
		30, 31	29.9, 5.0
1957	1	1, 2	8.0, 6.9
		6, 7	12.0, 2.5
	2	17, 18	16.5, 2.5
1958	4	8, 9	7.7, 4.0
	5	3, 4	24.0, 31.5
	6	28, 29	2.1, 2.0
	7	22, 23	0, 9.1
	9	4, 5	14.1, 14.0
		7, 8	24.0, 4.7
	12	10, 11	5.0, 21.5
1959	1	29, 30	7.6, 10.0
	4	19, 20	9.0, 11.0
	5	10, 11	4.0, 4.4
	6	21, 22	3.2, 27.0
	7	29, 30	8.4, 8.6
	10	28, 29	2.5, 28.0
1960	1	8, 9	0.5, 6.0
1964	3	20, 21	4.0, 4.5
	6	14, 15	36.8, 5.0
1965	1	10, 11	7.2, 20.3
1966	6	1, 2	36.7, 20.0
1967	7	1, 2	0.8, 19.4

SUMMARY AND CONCLUDING REMARKS

Rainfall series of different climatic regions were analysed with the aim of generating daily rainfall sequences. A survey of the data is given in I, 1.

When analysing daily rainfall sequences one must be aware of the following points:

a. Seasonality. Because of seasonal variation of features of the rainfall process the analysis is done for each month or season separately (see III, 2).

b. Non-homogeneity. A rainfall series is called non-homogeneous when it is non-stationary even after elimination of seasonal variation.

c. A large fraction of days with no rain.

d. Dependence between rainfall amounts on successive days (serial correlation).

It is the combination of the last two points which makes the generation of daily rainfall sequences difficult. When dealing with rainfall observations over periods longer than one day this difficulty is mostly obviated because one gets less zeroes and evidence for serial correlation usually disappears. For instance, there is no evidence for serial correlation in monthly data of Dutch stations (see II, 3.1). Besides, theoretical distributions can easily be fitted to the marginal distribution (e.g. the 'loi des fuites', see II, 3.2). The generation of these data is therefore not complicated. For annual totals the Gaussian distribution often fits reasonably well (see II, 2 for Dutch series, and V, 2.1 for foreign series). Departures from normality are found for rainfall stations with a few wet days in a year (New Delhi, Khartoum).

Homogeneity of Dutch rainfall series is discussed in Chapter II. It is assumed that non-homogeneities are man-made, e.g. due to a change in rain gauge installation or a change of observer and therefore non-homogeneities usually consist of jumps.

A problem when dealing with Dutch rainfall series is the lowering of the rain gauges (from 1.50 m to 0.40 m) during the period 1946–1954 (see II, 4). Due to a smaller wind effect it is expected that such a reduction in height results in larger rainfall measurements. To find a jump in the mean, annual totals of Dutch stations were compared with contemporary totals of foreign stations where no change of height took place. For such a comparison two points are important:

a. The distance between the various rainfall stations. In order to obtain a powerful test for a jump, one should choose the stations close together. Therefore Dutch rainfall stations near the Belgian or German border were taken.

b. There are other non-homogeneities, for instance, due to changes of site. The consequence of such non-homogeneities is that the estimates of a jump, caused by a reduction of height, may be biased. Moreover, these non-homogeneities give rise to a smaller correlation between the rainfall series and the tests for a jump become therefore less powerful. The influence of local changes can be reduced by taking averages of different stations in a certain area.

With regression models and plots of partial sums, a jump in the mean of about 2 per cent is found for stations remote from the coast; for coastal stations the height of the jump can be much larger (even more than 10 per cent), but there is a large variation due to differences in the degree of protection against the wind. The results correspond quite well with those of earlier research by BRAAK (1945).

By comparing monthly data of Dutch and German stations in the northern coastal area (see II, 4.2) it is found that the largest jumps occur in the winter season.

Another point of investigation is the homogeneity of the Zwanenburg-Hoofddorp (1735–1972) series (see II, 5). Since here there is no nearby rainfall station, with no changes in the way of measuring during the period of observation, the analysis of homogeneity was merely based on the series under consideration. The tests which were considered are less powerful than the ones based on a comparison between changed and unchanged stations. Yet, there is obvious evidence for differences in the means of Zwanenburg (1735–1860) and Hoofddorp (1861–1972). There is no evidence for departures from homogeneity in the Hoofddorp series. Since there is also a poor correlation between the Zwanenburg data and other old rainfall series, these data can be considered useless for present-day hydrological research.

Because of the large number of zeroes in daily rainfall sequences, it is suggested to generate first the occurrence of wet and dry days and subsequently the rainfall amounts on wet days. Since small rainfall amounts are often registered as zero it is advisable to call a day wet if its rainfall amount exceeds some specified value. For the Netherlands a threshold of 0.8 mm is advisable (see II, 6); for smaller thresholds there are only a few rainfall stations for which the series of wet and dry days (shortly denoted as wet-dry series) is homogeneous.

In Chapter III a model is developed for Dutch rainfall series, using daily data from Winterswijk (1908–1973), Hoofddorp (1867–1971) and Hengelo (1908–1973). Theoretical considerations about the model are given in Chapter IV.

With respect to the wet-dry sequences of these series it can be concluded:

a. There is no evidence for correlation between the lengths of successive wet and dry spells (see III, 3.1).
b. Modifications of the negative binomial distribution (the shifted negative

binomial distribution, see III, (3.2) and the truncated negative binomial distribution, see III, (3.3)) fit the lengths of weather spells well.

Seasonal dependence of the parameters of the truncated negative binomial distribution was extensively studied. For a particular type of spell it was shown that it is reasonable to keep one of the parameters, r, constant throughout the year. Further, for dry spells the other parameter, p, can be smoothed according to a moving average scheme (see III, (3.17)); for wet spells seasonal variation of the parameter p can be described by a Fourier series with one harmonic component (see III, (3.12)).

With respect to the behaviour of rainfall amounts on wet days the following remarks can be made.

a. There is no evidence for correlation between the rainfall amount on the first day of a wet spell and the length of the preceding dry spell (see III, 4.1).
b. The first and the last day of a wet spell have smaller means than the other wet days; the smallest mean is found for solitary wet days (see III, 4.2).
c. There is some evidence for serial correlation of successive rainfall amounts within a wet spell (see III, 5.1). It is assumed that this serial correlation can be described by a first order moving average process (see III, 6.1).

The last two points are most evident during the winter season.

A shifted gamma distribution fits the marginal distribution of the rainfall amounts on wet days reasonably well (see III, 5.2). There is no evidence for seasonal variation of the shape parameter; the mean, however, shows an obvious seasonal variation.

Though synthetic sequences resemble the historic series with respect to features contained in the model (such as the marginal distribution of daily rainfall amounts and the lengths of wet and dry spells), this is not necessarily true for other features. As examples the correlogram and features of k-day sums ($k = 2, 3, \ldots$) were considered. This was done for both the wet-dry process and the entire rainfall process.

Some features of the rainfall model can be obtained by numerical methods. These features are:

a. The cumulative distribution function (cdf) of the number of wet days in a k-day period. Under the assumption of iid rainfall amounts within a wet spell it is not difficult to derive an expression for the cdf of k-day rainfall totals (see IV, 3).
b. The correlogram for both the wet-dry process and the entire rainfall process (see IV, 4).
c. The variance-time curve of the wet-dry process and of the entire rainfall process (see IV, 5). For large values of k ($k > 10$) the variance of the number of wet days in a k-day period can be approximated well by an asymptotic formula (Equation IV, (5.36)) involving only the first three moments of the lengths of wet and dry spells. This approximation can also be done for the variance of k-day rainfall amounts when the rainfall totals within a wet spell are iid.

For the derivation of the formulas, underlying these numerical calculations, the following assumptions are made.

a. The process is stationary.

b. The wet-dry process is an alternating renewal process. A definition of this process is given in IV, 2.2.

These assumptions turn out to be reasonable when the rainfall process is examined for a particular month or season.

For the correlogram it can be concluded (see III, 6.1):

a. There is a good correspondence between the estimated first serial correlation coefficient and the theoretical value for both the wet-dry process and the entire rainfall process. This quantity is usually underestimated when simplifying assumptions are made about the behaviour of rainfall amounts within a wet spell.

b. For larger lags the model usually underestimates the serial correlation coefficients of the rainfall process, especially during the winter season. For the wet-dry process the model usually provides a better fit at the higher lags.

Closely related to the last point is the fact that the model underestimates the variances of 30-day rainfall amounts (see III, 6.2). During winter and autumn sometimes long wet spells occur with very high intensity (see III, 4.2) which inflate the estimated variances of k-day totals for large values of k.

The following remarks can be made on the cdf of k-day sums.

a. For the number of wet days in a k-day period there is a good correspondence between theoretical and empirical cdfs (see III, 7.1).

b. For the entire rainfall process theoretical cdfs fit well for small values of k; poor fit may occur for larger values of k (e.g. $k = 30$). This poor fit usually consists of an underestimation of the probabilities of large values (see III, 7.2.2).

Though the cdf was only investigated under the assumption of independent rainfall amounts within a wet spell, it may be expected that the shape of the cdf is hardly influenced when serial correlation between these rainfall amounts is assumed, since the increase in the variance of k-day totals is only very small for a model with serial correlation (see III, 6.2).

For the rainfall process it was investigated how different features of the model affect the shape of the cdf of 30-day totals. The main results are:

a. The shape of the cdf is hardly influenced by the distribution of the lengths of weather spells (see III, 7.1).

b. The shape of the cdf is to some extent not sensitive to the marginal distribution of the rainfall amounts on wet days (see III, 7.2).

c. The shape of the cdf is hardly altered when rainfall amounts within a wet spell are assumed to be iid.

For Winterswijk (1908–1973) nearly the same results were obtained when the threshold defining a wet day is taken to be 0.3 mm instead of 0.8 mm.

Though there are many corrections and supplements in the series of Hengelo

(1908–1973) the results for this station correspond quite well to those of the adjacent station of Winterswijk.

In Chapter V daily rainfall sequences of stations with a more pronounced seasonal variation than Dutch stations are discussed.

The problems encountered for Dutch stations usually arise here too:

a. In order to get a homogenous wet-dry series one is often forced to call only those days wet for which the rainfall amount exceeds a rather large threshold (see V, 2.1).

b. Rainfall amounts within a wet spell are often non-identically distributed. Moreover, there usually exists a small serial correlation between rainfall amounts within a wet spell (see V, 2.3).

c. The rainfall model underestimates the variances of k-day totals for large values of k (see V, 2.4).

Besides, for the series analysed in Chapter V there are some problems associated with dry seasons with no or hardly any rainfall:

a. It is often not possible to fit the shifted negative binomial distribution or the truncated negative binomial distribution to lengths of wet spells during the dry season. Since there are no long wet spells during this season, the likelihood equations of these distributions often do not have a solution within the parameter space. In such cases it is possible to fit a one-parameter distribution (geometric, logarithmic series) to the lengths of wet spells (see V, 2.2).

b. Dry spells can be quite long. Modifications of the negative binomial distribution sometimes cannot fit the lengths of these spells (see V, 2.2 and V, 3). In such cases it might be advisable to use transition probabilities for the generation of the wet-dry series instead of generating lengths of wet and dry spells. For instance, it was shown, by simulation, that a first order Markov chain describes the right tail of the distribution of the lengths of dry spells well for the station of Alexandria (see V, 3).

The generation of synthetic data for Pasar Minggu (Indonesia) was investigated in more detail (see V, 2.5). Special attention was paid to the beginning of both the wet and the dry monsoon. The model can describe the transitions between these seasons quite well.

A special problem arises for the rainfall series of Khartoum (1902–1940). For this station there is some evidence for serial correlation in the annual totals and in the annual number of wet days (see V, 3). This serial correlation can be explained by persistence in the lengths of successive wet and dry seasons. It is proposed therefore to generate the beginning and the end of the wet season first. Within a wet season the rainfall process can be approximated by a Bernoulli process for the occurrence of wet and dry days and a shifted gamma distribution for the rainfall amounts on wet days. The probability of a day being wet and the mean of the rainfall amount on a wet day show seasonal variation.

The main shortcoming of the daily rainfall model is that it underestimates the variance of k-day totals for large values of k which may result in it poorly fitting the distribution of these totals. It is, however, by no means certain whether this shortcoming is important in practical situations. When dealing with hydrological systems with a long memory one may expect serious problems but studies on such systems can often be based on a time-scale longer than one day. Therefore it is necessary to test the model on some real problems to obtain a better insight into its shortcomings.

One may ask whether improvements of the model are possible. For Dutch series the description of the wet-dry process by a seasonal changing alternating renewal process seems reasonable, since the model fits well the probability distributions of the annual number of wet days (see III, 7.2.2) and of the number of wet days in a 30-day period (see III, 7.1). Therefore one must think of a better model for the behaviour of rainfall amounts on wet days. It is impractical to incorporate serial correlation of higher order between rainfall amounts within a wet spell as the effect on the variance-time curve of the process is negligible, because wet spells usually are of short duration. The model could be improved by:

a. a random slowly changing mean of the rainfall amounts on wet days. This certainly will increase the variance of k-day totals for large values of k. The main problem of this method is the estimation of the parameters. Another problem can be the choice of the type of distribution for the rainfall amounts on wet days.

b. generating a total rainfall amount for a particular period (e.g. a month) and splitting this rainfall amount into the rainfall amounts of the various wet days of that period. Because of the method of generation this model may give a reasonable fit to monthly and annual totals. A disadvantage of this method is that the model contains a large number of parameters.

But before thinking of such improvements one must realize that there are large local differences for the variances of 30-day totals (see III, 6.2). It is therefore necessary to analyse a large number of daily rainfall sequences of the Netherlands and its neighbouring countries.

For some foreign stations analysed in Chapter V one also has the trouble that for large values of k an alternating renewal process leads to a serious underestimation of the variance of the number of wet days in a k-day period. Research still has to be done to get a better model for such series.

ACKNOWLEDGMENTS

This research was done under supervision of Prof Dr Ir W. H. van der Molen and Prof Dr P. van der Laan.

Throughout this study Dr Ir M. A. J. van Montfort made many helpful suggestions.

The procedure for obtaining maximum likelihood estimates of the 'loi des fuites' (Appendix A3 of Chapter II) was suggested by Drs A. Otten. He also made some valuable comments on Chapter IV.

SAMENVATTING EN SLOTOPMERKINGEN

Dit onderzoek behandelt een analyse van neerslagreeksen van stations met een verschillend klimaat met tot doel het genereren van dagelijkse neerslagsommen. Een overzicht van de gebruikte gegevens wordt gegeven in I, 1.

Bij de analyse van dagelijkse neerslagreeksen dient men rekening te houden met de volgende punten:

a. Seizoenmatige veranderingen. Vanwege de jaarlijkse gang van eigenschappen van het neerslagproces wordt de analyse voor iedere maand of seizoen afzonderlijk uitgevoerd (zie III, 2).
b. Inhomogeniteit. Een neerslagreeks wordt inhomogeen genoemd indien deze door andere oorzaken dan een seizoenmatige verandering niet stationair is.
c. Een grote fractie neerslagvrije dagen.
d. Afhankelijkheid tussen neerslaghoeveelheden van opeenvolgende dagen (autocorrelatie).

De combinatie van de laatste twee punten maakt het genereren van dagelijkse neerslagreeksen moeilijk. Bij neerslaggegevens voor perioden van een toenemend aantal dagen treft men deze moeilijkheid doorgaans in geringere mate aan daar men minder nullen krijgt en evidentie voor autocorrelatie gewoonlijk verdwijnt. Voor maandsommen van Nederlandse stations is er bijvoorbeeld geen evidentie voor autocorrelatie (zie II, 3.1). Het is bovendien niet moeilijk de marginale verdeling door een theoretische verdeling te benaderen (b.v. de 'loi des fuites', zie II, 3.2). Het genereren van maandsommen is daardoor betrekkelijk eenvoudig. De normale verdeling geeft vaak een redelijke aanpassing voor jaarsommen (zie II, 2 voor Nederlandse reeksen, en V, 2.1 voor buitenlandse reeksen). Afwijkingen van normaliteit worden gevonden bij neerslagstations waarvoor het aantal natte dagen in een jaar gering is (New Delhi, Khartoum).

In Hoofdstuk II wordt de homogeniteit van Nederlandse neerslagreeksen onderzocht. Verondersteld wordt dat inhomogeniteiten door de mens veroorzaakt zijn, b.v. door een verandering van de installatie van de regenmeter of een verandering van waarnemer; daardoor bestaan inhomogeniteiten gewoonlijk uit sprongen.

Een probleem bij Nederlandse neerslagreeksen is de verlaging van de regenmeters (van 1.50 m naar 0.40 m) in de periode 1946–1954 (zie II, 4). Men mag verwachten dat deze verlaging hogere neerslagmetingen tot gevolg heeft door een geringer windeffect. Teneinde een sprong in het gemiddelde te vinden worden jaarsommen van Nederlandse stations vergeleken met gelijktijdige waarnemingen van buitenlandse stations waar geen verandering van de opstellingshoogte heeft plaatsgevonden. Bij zulk een vergelijking zijn twee punten van belang:

a. De afstand tussen de verschillende neerslagstations. Voor het verkrijgen van een onderscheidende toets dient men de stations dicht bij elkaar te kiezen. Nederlandse stations dichtbij de Belgische of Duitse grens zijn daardoor gekozen.

b. Er zijn andere inhomogeniteiten, bijvoorbeeld door plaatsveranderingen.

Het gevolg van dergelijke inhomogeniteiten is dat schattingenvan een sprong, veroorzaakt door een verlaging van de opstellingshoogte, onzuiver kunnen zijn. Bovendien geven deze inhomogeniteiten aanleiding tot een geringere samenhang tussen de neerslagreeksen en toetsen op een sprong worden daardoor minder onderscheidend. De invloed van lokale veranderingen kan men reduceren door gemiddelden te nemen van verschillende stations in een bepaald gebied.

Door middel van regressiemodellen en tekeningen van partiële sommen wordt een sprong in het gemiddelde gevonden van ongeveer 2 procent voor landstations; voor kuststations is de hoogte van de sprong veel groter (zelfs meer dan 10 procent), maar door verschillen in de mate van beschutting is er een grote variatie tussen deze stations. De resultaten komen redelijk overeen met die van een eerder onderzoek van BRAAK (1945).

Bij een vergelijking van maandsommen van Nederlandse en Duitse stations in het noordelijk kustgebied (zie II, 4.2) worden de grootste sprongen in het winterseizoen aangetroffen.

Een ander punt van onderzoek is de homogeniteit van de Zwanenburg-Hoofddorp (1735–1972) reeks (zie II, 5). Daar men in dit geval niet over een nabij gelegen station beschikt, waarop de neerslag steeds op dezelfde manier gemeten is gedurende de waarnemingsperiode, wordt de homogeniteitsanalyse uitsluitend op de beschouwde reeks gebaseerd. De beschouwde toetsen zijn minder onderscheidend dan toetsen gebaseerd op een vergelijking van stations met en zonder veranderingen. Toch is er evidentie voor verschillen in de gemiddelden van Zwanenburg (1735–1860) en Hoofddorp (1861–1972). Er is geen evidentie voor inhomogeniteiten in de Hoofddorp reeks. Doordat er ook een slechte samenhang bestaat tussen de gegevens van Zwanenburg en andere oude neerslagreeksen, mag men deze gegevens ongeschikt achten voor het huidige hydrologisch onderzoek.

Daar het aantal nullen in dagelijkse neerslagreeksen groot is wordt eerst het optreden van droge en natte dagen gegenereerd en daarna de neerslaghoeveelheden op natte dagen. Doordat kleine neerslaghoeveelheden vaak als nul geregistreerd worden is het raadzaam pas van een natte dag te spreken als de neerslaghoeveelheid op die dag een zekere waarde overschrijdt. Voor Nederland is een drempelwaarde van 0.8 mm redelijk (zie II, 6); voor kleinere waarden van de drempel zijn er slechts weinig neerslagstations waarvoor de reeks van droge en natte dagen homogeen is.

In Hoofdstuk III wordt een model ontwikkeld voor Nederlandse neerslagreeksen, waarbij gebruik wordt gemaakt van dagsommen van Winterswijk (1908–1973), Hoofddorp (1867–1971) en Hengelo (1908–1973). Theoretische beschouwingen over het model worden gegeven in Hoofdstuk IV.

Met betrekking tot de opeenvolging van droge en natte dagen van deze reeksen kan het volgende worden opgemerkt:

a. Er is geen evidentie voor correlatie tussen de lengten van opeenvolgende natte en droge perioden (zie III, 3.1).

b. De verdeling van de lengten van droge en natte perioden kan goed beschreven worden door modificaties van de negatief binomiale verdeling (de verschoven negatief binomiale verdeling, zie III, (3.2) en de afgeknotte negatief binomiale verdeling, zie III, (3.3)).

Uitgebreide aandacht wordt geschonken aan de seizoenafhankelijkheid van de parameters van de afgeknotte negatief binomiale verdeling. Voor zowel droge als natte perioden is aangetoond dat het redelijk is om één der parameters, r, constant te houden gedurende het jaar. Voor droge perioden kan de andere parameter, p, gladgestreken worden volgens een voortschrijdend gemiddelde (zie III, (3.17)); voor natte perioden kan de seizoenafhankelijkheid van de parameter p door een Fourier reeks beschreven worden met één sinusoïde (zie III, (3.12)).

De volgende opmerkingen kunnen worden gemaakt over het gedrag van neerslaghoeveelheden op natte dagen:

a. Er is geen evidentie voor correlatie tussen de neerslaghoeveelheid op de eerste dag van een natte periode en de lengte van de voorafgaande droge periode (zie III, 4.1).

b. De eerste en de laatste dag van een natte periode hebben een kleiner gemiddelde dan de andere natte dagen; het kleinste gemiddelde wordt gevonden bij natte dagen die aan beide zijden door droge dagen begrensd worden (zie III, 4.2).

c. Er is enige evidentie voor autocorrelatie tussen opeenvolgende neerslaghoeveelheden binnen een natte periode (zie III, 5.1). Verondersteld wordt dat deze autocorrelatie beschreven kan worden door een eerste orde voortschrijdend gemiddelde proces (zie III, 6.1).

De laatste twee punten zijn het duidelijkst in het winterseizoen.

Een verschoven gammaverdeling geeft een redelijke aanpassing voor de neerslaghoeveelheden op natte dagen (zie III, 5.2). Er is geen evidentie voor seizoenafhankelijkheid van de vormparameter; het gemiddelde vertoont echter een duidelijke seizoenmatige verandering.

Eigenschappen van synthetische reeksen zullen gelijken op die van de historische reeks indien zij in het model ingebouwd zijn (zoals de marginale verdeling van dagelijkse neerslaghoeveelheden en de lengten van natte en droge perioden); dit is niet noodzakelijk waar voor andere eigenschappen. Als voorbeelden worden het correlogram en eigenschappen van k-daagse sommen ($k = 2, 3, \ldots$) beschouwd. Dit wordt gedaan voor zowel de reeks van droge en natte dagen als voor het gehele neerslagproces.

Sommige eigenschappen van het neerslagmodel kunnen afgeleid worden met behulp van numerieke methoden. Deze eigenschappen zijn:

a. De cumulatieve verdelingsfunctie (cdf) van het aantal natte dagen in een k-daagse periode. Onder de veronderstelling van onafhankelijke neerslaghoeveelheden met dezelfde kansverdeling binnen een natte periode kan men een uitdrukking voor de cdf van k-daagse neerslagsommen afleiden (zie IV, 3).

b. Het correlogram van het proces van droge en natte dagen en het gehele neerslagproces (zie IV, 4).

c. De variantie-tijd curve van het proces van droge en natte dagen en van het gehele neerslagproces (zie IV, 5). Voor grote waarden van k ($k > 10$) kan de variantie van het aantal natte dagen in een k-daagse periode goed benaderd worden door een asymptotische formule (vergelijking IV, (5.36)) waarin slechts de eerste drie momenten van de lengten van natte en droge perioden voorkomen. Dit is ook het geval voor de variantie van k-daagse neerslagsommen indien de neerslaghoeveelheden binnen een natte periode onafhankelijk en isomoor zijn.

Voor de afleiding van de formules, die aan deze numerieke berekeningen ten grondslag liggen, worden de volgende veronderstellingen gemaakt:

a. Het proces is stationair.

b. Het proces van droge en natte dagen is een alternerend vervangingsproces. Een definitie van dit proces wordt gegeven in IV, 2.2.

Deze veronderstellingen blijken redelijk te zijn indien het neerslagproces bestudeerd wordt voor een bepaalde maand of een bepaald seizoen.

Voor het correlogram kunnen de volgende opmerkingen worden gemaakt (zie III, 6.1):

a. Er is een goede overeenstemming tussen de geschatte eerste orde autocorrelatiecoëfficiënt en de theoretische waarde voor zowel het proces van droge en natte dagen als het gehele neerslagproces. Deze correlatiecoëfficiënt wordt gewoonlijk onderschat indien vereenvoudigende aannamen worden gemaakt over het gedrag van neerslaghoeveelheden binnen een natte periode.

b. Het model onderschat gewoonlijk de hogere orde autocorrelatiecoëfficiënten van het neerslagproces, vooral in het winterseizoen. Voor het proces van droge en natte dagen geeft het model gewoonlijk een betere aanpassing voor hogere orde autocorrelatiecoëfficiënten.

Nauw verwant aan het laatste punt is het feit dat het model de varianties van 30-daagse neerslagsommen onderschat (zie III, 6.2). In de winter en de herfst komen soms lange natte perioden voor met een zeer hoge intensiteit (zie III, 4.2), die de geschatte varianties van k-daagse sommen verhogen voor grote waarden van k.

De volgende opmerkingen kunnen worden gemaakt over de cdf van *k*-daagse sommen:

a. Voor het aantal natte dagen in een *k*-daagse periode is er een goede overeenkomst tussen theoretische en empirische cdf's (zie III, 7.1).

b. Voor het gehele neerslagproces geven theoretische cdf's een goede aanpassing voor kleine waarden van *k*; de aanpassing kan slecht zijn voor grote waarden van *k* (b.v. k = 30). De slechte aanpassing bestaat gewoonlijk uit een onderschatting van de kansen op grote waarden (zie III, 7.2.2).

Hoewel de cdf slechts onderzocht wordt onder de veronderstelling van onafhankelijke neerslaghoeveelheden binnen een natte periode mag verwacht worden dat de vorm van de cdf nauwelijks verandert indien autocorrelatie tussen deze neerslaghoeveelheden verondersteld wordt, daar de toename in de variantie van *k*-daagse sommen slechts zeer klein is voor een model met autocorrelatie (zie III, 6.2).

Voor het neerslagproces is onderzocht hoe verschillende eigenschappen van het model de vorm van de cdf van 30-daagse sommen beïnvloeden. De belangrijkste resultaten zijn:

a. De vorm van de cdf wordt nauwelijks beïnvloed door de verdeling van de lengten van droge en natte perioden (zie III, 7.1).

b. De vorm van de cdf is enigszins ongevoelig voor de marginale verdeling van de neerslaghoeveelheden op natte dagen (zie III, 7.2).

c. De vorm van de cdf verandert nauwelijks indien verondersteld wordt dat neerslaghoeveelheden binnen een natte periode onafhankelijk en isomoor zijn.

Voor Winterswijk verkrijgt men vrijwel dezelfde resultaten als men voor de drempel, die een natte dag definieert, een waarde kiest van 0.3 mm in plaats van 0.8 mm.

Hoewel men vele correcties en aanvullingen in de reeks van Hengelo (1908–1973) aantreft komen de resultaten voor dit station redelijk overeen met die van het nabijgelegen station Winterswijk.

In Hoofdstuk V worden dagelijkse neerslagreeksen van stations onderzocht met een meer uitgesproken seizoenvariatie dan Nederlandse stations.

De problemen, die bij Nederlandse stations voorgekomen zijn, doen zich hier ook voor:

a. Om een homogene reeks van natte en droge dagen te krijgen is men vaak gedwongen slechts die dagen nat te noemen, waarop de neerslaghoeveelheid een tamelijk hoge drempelwaarde overschrijdt (zie V, 2.1).

b. Neerslaghoeveelheden binnen een natte periode hebben vaak niet dezelfde kansverdeling. Bovendien bestaat er doorgaans een geringe autocorrelatie tussen neerslaghoeveelheden binnen een natte periode (zie V, 2.3).

c. Het neerslagmodel onderschat de varianties van *k*-daagse sommen voor grote waarden van *k* (zie V, 2.4).

Voor de reeksen, die geanalyseerd zijn in Hoofdstuk V, geven droge seizoenen met geen of nauwelijks enige regen aanleiding tot enkele problemen:

a. Voor lengten van natte perioden in het droge seizoen is het vaak niet mogelijk om de verschoven negatief binomiale verdeling of de afgeknotte negatief binomiale verdeling aan te passen. Daar er geen lange natte perioden in dit seizoen zijn hebben de aannemelijkheidsvergelijkingen van deze verdelingen vaak geen oplossing binnen de parameterruimte. In zulke gevallen is het mogelijk om voor lengten van natte perioden een één-parameter verdeling (geometrische, logaritmische verdeling) aan te passen (zie V, 2.2).

b. Droge perioden kunnen zeer lang zijn. Modificaties van de negatief binomiale verdeling kunnen soms niet aangepast worden aan de lengten van deze perioden (zie V, 2.2 en V, 3). In zulke gevallen kan het raadzaam zijn om overgangswaarschijnlijkheden te gebruiken voor het genereren van een reeks van droge en natte dagen in plaats van het genereren van lengten van natte en droge perioden. Zo is door simulatie aangetoond dat een eerste orde Markov keten een goede aanpassing geeft voor de rechter staart van de verdeling van droge perioden voor het station Alexandrië (zie V, 3).

Het genereren van synthetische gegevens voor Pasar Minggu (Indonesië) wordt uitvoeriger onderzocht (zie V, 2.5). Speciale aandacht wordt besteed aan het begin van de natte en van de droge moesson. Het model kan de overgang tussen deze seizoenen redelijk beschrijven.

Een speciaal probleem doet zich voor bij de neerslagreeks van Khartoum (1902–1940). Voor dit station is er enige evidentie voor autocorrelatie in de jaarsommen en in het jaarlijks aantal natte dagen (zie V, 3). Deze autocorrelatie kan verklaard worden door persistentie in de lengten van opeenvolgende natte en droge seizoenen. Daarom worden eerst het begin en het einde van het natte seizoen gegenereerd. Binnen een nat seizoen kan het neerslagproces benaderd worden door een Bernoulli proces voor de opeenvolging van droge en natte dagen en een verschoven gamma verdeling voor de neerslaghoeveelheden op natte dagen. De kans op een natte dag en de gemiddelde neerslaghoeveelheid op een natte dag zijn seizoenafhankelijk.

De belangrijkste tekortkoming van het dagelijks neerslag model is de onderschatting van de variantie van *k*-daagse sommen voor grote waarden van *k*, die kan leiden tot een slechte aanpassing voor deze sommen. Het is echter geenszins zeker of deze tekortkoming belangrijk is in praktische situaties. Voor trage hydrologische systemen kan men ernstige problemen verwachten, maar studies over dergelijke systemen kunnen vaak gebaseerd worden op tijdseenheden van meer dan een dag. Het is daarom noodzakelijk om het model op enkele reële problemen te toetsen om een beter inzicht te krijgen in de tekortkomingen.

Men kan zich afvragen of het mogelijk is om het model te verbeteren. Voor Nederlandse reeksen kan het proces van droge en natte dagen goed beschreven worden door een seizoenafhankelijk alternerend vervangingsproces, daar dit model tot een goede aanpassing leidt voor de kansverdelingen van het jaarlijks aantal regendagen (zie III, 7.2.2) en van het aantal natte dagen in een 30-daagse periode (zie III, 7.1). Men moet daarom in de eerste plaats denken aan een beter model voor het gedrag van neerslaghoeveelheden op natte dagen. Het is onpraktisch om autocorrelatie van hogere orde in te bouwen tussen de neerslaghoeveelheden binnen een natte periode, omdat het effect hiervan op de variantie-tijd curve gering is, daar natte perioden doorgaans van korte duur zijn. Voor verbetering van het model kan men denken aan:

a. een stochastische, langzaam veranderende verwachting van de neerslaghoeveelheden op natte dagen. Hierdoor zal voor grote waarden van *k* de variantie van *k*-daagse sommen zeker toenemen. Het belangrijkste probleem bij deze methode is gelegen in hèt schatten van de parameters. Een ander probleem kan de keuze van het type verdeling voor de neerslaghoeveelheden op natte dagen zijn.

b. het genereren van een totale neerslaghoeveelheid voor een bepaalde periode (b.v. een maand) en het opsplitsen van deze neerslaghoeveelheid in de neerslaghoeveelheden voor de verschillende natte dagen van die periode. Door de manier van genereren kan dit model een redelijke aanpassing geven voor maand- en jaarsommen. Een nadeel van deze methode is dat het model een groot aantal parameters bevat.

Maar voordat men aan dergelijke verbeteringen begint dient men zich te realiseren dat er grote lokale verschillen voor de varianties van 30-daagse sommen bestaan (zie III, 6.2). Het is daarom noodzakelijk om een groot aantal neerslagreeksen uit Nederland en omringende landen te analyseren.

Bij enkele buitenlandse stations, die in Hoofdstuk V geanalyseerd zijn, heeft men ook nog de moeilijkheid dat een alternerend vervangingsproces de variantie van het aantal natte dagen in een *k*-daagse periode onderschat voor grote waarden van *k*. Veel onderzoek moet nog gedaan worden om een beter model voor zulke reeksen te verkrijgen.

REFERENCES

Abramowitz, Milton and Stegun, Irene A. (1970). Handbook of Mathematical Functions. Dover Publications, Inc., New York, seventh printing.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. Journal of the Royal Statistical Society, Series B, **8**, 27–41.

Barton, D. E., David, F. N. and Merrington, M. (1963). Tables for the solution of the exponential equation $\exp(b) - \frac{b}{1-p} = 1$. Biometrika, **50**, 169–172

Bernier, J. (1967). Sur la théorie de renouvellement et son application en hydrologie. Electricité de France. Direction des études et recherches.

Bernier, J. and Fandeux, D. (1970). Théorie du renouvellement. Application à l'étude statistique des précipitations mensuelles. Revue de Statistique Appliquée, **XVIII**, 75–87.

Boer, H. J. de (1956). De cumulatieve frequentieverdelingen van k-daagse neerslagsommen van Hoofddorp. WR56–004 (III–186) KNMI, De Bilt.

Boer, H. J. de (1957). De cumulatieve frequentieverdelingen van k-daagse neerslagsommen van Winterswijk. WR57–007 (III–202) KNMI, De Bilt.

Boer, H. J. de (1958). On the Cumulative Frequency Distributions of k-Day Period Amounts of Precipitations for any Station in the Netherlands, while $k > 30$. Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B, Band 9, 2. Heft, 244–253.

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis forecasting and control. Holden–Day, San Francisco.

Braak, C. (1945). Invloed van den wind op regenwaarnemingen (Influence of the wind on rainfall measurements). Mededelingen en Verhandelingen, KNMI, Nr 48, 's-Gravenhage.

Brass, W. (1958). Simplified methods of fitting the truncated negative binomial distribution. Biometrika, **45**, 59–68.

Brazier, M. C. E. (1927). Sur la mesure correcte de la pluie. La Météorologie, Tome **III** (Tome **LXX**, ancienne série), 385–395.

Caskey, James E., Jr. (1963). A Markov chain model for the Probability of Precipitation Occurrence in Intervals of Various Lenth. Monthly Weather Review, **91**, 298–301.

Choi, S. C. and Wette, R. (1969). Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. Technometrics, **11**, 683–690.

Cole, J. A. and Sherriff, J. D. F. (1972). Some single- and multi-site models of rainfall within discrete time increments. Journal of Hydrology, **17**, 97–113.

Cooke, D. S. (1953). The duration of wet and dry spells at Moncton, New Brunswick. Quarterly Journal of the Royal Meteorological Society, **79**, 536–538.

Cox, D. R. (1962). Renewal theory. Methuen & Co., London.

Cox, D. R. and Lewis, P. A. W. (1966). The Statistical Analysis of Series of Events. Methuen & Co., London.

Cox, D. R. and Miller, H. D. (1965). The theory of stochastic processes. Methuen & Co., London.

D'Agostino, Ralph B. (1970). Transformation to normality of the null distribution of g_1. Biometrika, **57**, 679–681.

Das, S. C. (1955). The fitting of truncated type III curves to daily rainfall data. Australian Journal of Physics, **8**, 298–304.

Dumont, A. G. and Boyce, D. S. (1974). The Probabilistic Simulation of Weather Variables. Journal of agricultural Engineering Research, **19**, 131–145.

Elliott, E. O. (1965). A Model of the Switched Telephone Network for Data Communications. The Bell System Technical Journal, **44**, 89–109.

Feller, W. (1949). Fluctuation theory of recurrent events. Transactions of the American Mathematical Society, **67**, 98–119.

FELLER, W. (1968). An introduction to probability theory and its applications, Vol I, Third edition. John Wiley & Sons, Inc., New York.

FISHER, R. A. (1953). Note on the efficient fitting of the negative binomial. Biometrics, **9**, 197–200.

FISHER, SIR RONALD A. and CORNISH, E. A. (1960). The percentile points of distributions having known cumulants. Technometrics, **2**, 209–225.

GABRIEL, K. R. (1959). The Distribution of the Number of Successes in a Sequence of Dependent Trials. Biometrika, **46**, 454–460.

GABRIEL, K. R. and NEUMANN, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society, **88**, 90–95.

GRACE, R. A. and EAGLESON, P. S. (1966). The synthesis of short-time incremental rainfall sequences. Massachusetts Institute of Technology. Dept. of Civil Engineering, Ralph M. Parsons Laboratory, Report No. 91.

GREEN, J. R. (1964). A Model for Rainfall Occurrence. Journal of the Royal Statistical Society, Series B, **26**, 345–353.

GREENWOOD, ARTHUR J. (1974). A Fast Generator for Gamma-Distributed Random Variables. COMPSTAT 1974, Proceedings in Computational Statistics, Physica Verlag Wien, pp. 19–27.

GREENWOOD, ARTHUR J. and DURAND, DAVID (1960). Aids for fitting the gamma distribution by maximum likelihood. Technometrics, **2**, 55–65.

HALDANE, J. B. S. (1941). The fitting of binomial distributions. Annals of Eugenics, **11**, 179–181.

HARTER, H. L. (1961). Expected Values of Normal Order Statistics. Biometrika, **48**, 151–165.

HERMANS, J. (1969). De chi-kwadraat toets voor aanpassing van continue verdelingen. Statistica Neerlandica, **23**, 277–285.

HIEMSTRA, LOURENS A.V. and CREESE, ROBERT C. (1970). Synthetic generation of seasonal precipitation. Journal of Hydrology, **11**, 30–46.

HOGG, R. V. (1961). On the resolution of Statistical Hypotheses. Journal of the American Statistical Association, **56**, 978–989.

JAGANNATHAN, P. and PARTHASARATHY, B. (1973). Trends and Periodicities of Rainfall over India. Monthly Weather Review, **101**, 371–375.

JENKINS, GWILYM M. and WATTS, DONALD G. (1969). Spectral Analysis and its applications. Holden-Day, San Francisco.

JÖHNK, M. D. (1964). Erzeugung von betaverteilten und gammaverteilten Zufallszahlen. Metrika, **8**, 5–15.

JOHNSON, NORMAN L. and KOTZ, SAMUEL (1969). Distributions in Statistics – Discrete Distributions. Houghton Mifflin Company, Boston.

JOHNSON, NORMAN L. and KOTZ, SAMUEL (1970). Distributions in Statistics – Continuous univariate distributions – 1. Houghton Mifflin Company, Boston.

KAVVAS, M. L. and DELLEUR, J. W. (1975). Methodology for the selection and application of probability models for the simulation of daily rainfall and runoff. Paper presented at the International Symposium and Workshops on the Application of Mathematical Models in Hydrology and Water Resources Systems – Bratislava.

KENDALL, MAURICE G. and STUART, ALAN D. (1969). The Advanced Theory of Statistics. Three-volume edition: Volume 1 Distribution Theory, Third edition. Charles Griffin & Company Limited, London.

KENDALL, MAURICE G. and STUART, ALAN D. (1973). The Advanced Theory of Statistics. Three-volume edition: Volume 2 Inference and Relationship, Third edition. Charles Griffin & Company Limited, London.

KNMI (1956). Frequenties van *k*-daagse neerslagsommen op Nederlandse stations (Frequency distribution of total amounts of rainfall in *k*-day periods at stations in the Netherlands). KNMI, Publikatie 140, De Bilt. (140–1 Winterswijk 1880–1953; 140–2 Hoofddorp 1867–1953).

LABRIJN, A. (1945). Het klimaat van Nederland gedurende de laatste twee en een halve eeuw (The climate of the Netherlands during the last two and a half centuries). Mededelingen en Verhandelingen, KNMI, Serie A, Nr 49, 's-Gravenhage.

LABRIJN, A. (1948). Het klimaat van Nederland. Temperatuur/Neerslag en wind. (The climate of the Netherlands. Temperature/Precipitation and Wind). Mededelingen en Verhandelingen, KNMI, Serie A, Nr 53, 's-Gravenhage.

LAWRENCE, E. N. (1954). Application of mathematical series to the frequency of weather spells. Meteorological Magazine, London, **83**, 195–200.

LECLERC, GUY and SCHAAKE, JOHN C., JR. (1973). Methodology for assessing the potential impact of urban development on urban runoff and the relative efficiency of runoff control alternatives. Massachusetts Institute of Technology. Dept. of Civil Engineering, Ralph M. Parsons Laboratory, Report No. 167.

LILLIEFORS, HUBERT W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, **62**, 399–402.

LINDGREN, B. W. (1968). Statistical Theory, Second Edition. The Macmillan Company, New York.

LOUTER, A. S. and KOERTS, J. (1970). On the Kuiper test for normality with mean and variance unknown. Statistica Neerlandica, **24**, 83–87.

LOWRY, WILLIAM P. and GUTHRIE, DONALD (1968). Markov chains of order greater than one. Monthly Weather Review, **96**, 798–801.

MONTFORT, M. A. J. VAN (1966). Statistische beschouwingen over neerslag en afvoer. H. Veenman & Zonen N.V., Wageningen.

MONTFORT, M. A. J. VAN (1968). Enige statistische opmerkingen over neerslag en afvoer. (Some statistical remarks on precipitation and runoff). Commissie voor hydrologisch onderzoek T.N.O., Verslagen en Mededelingen, No. 14, Regenwaarnemingscijfers (II), 's-Gravenhage, pp. 25–40 (Committee for hydrological research T.N.O., Proceedings and Informations, No 14, Precipitation Data (II), The Hague, pp. 25–40).

PARTHASARATHY, B. and DHAR, O. N. (1974). Secular variations of regional rainfall over India. Quarterly Journal of the Royal Meteorological Society, **100**, 245–257.

PARTHASARATHY, B. and DHAR, O. N. (1975). Trend analysis of annual Indian rainfall. Hydrological Sciences Bulletin, **XX**, 257–260.

PARZEN, EMANUEL (1962). Stochastic Processes. Holden-Day, San Francisco.

PAYNE, W. H., RABRING, J. R. and BOGYO, T. P. (1969). Coding the Lehmer Pseudo Random Number Generator. Communications of the ACM, **12**, 85–86.

PEARSON, E. S. and HARTLEY, H. O. (1962). Biometrika Tables for Statisticians, Volume 1. Cambridge University Press.

PEARSON, E. S. and HARTLEY, H. O. (1972). Biometrika Tables for Statisticians, Volume 2. Cambridge University Press.

QUÉLENNEC, R. E. (1973). Contribution à l'étude probabiliste des phénomènes pluvieux. Application aux bassins de la Charante et de la Seudre. LaHouille Blanche, No. 1–1973, 21–33.

RAMABHADRAN, V. K. (1954). A statistical study of the persistency of rain days during the monsoon season at Poona. Indian Journal of Meteorology and Geophysics, **5**, 48–55.

RAO, C. R. (1973). Linear Statistical Inference and Its Applications, Second edition. John Wiley & Sons, Inc., New York.

RICHARDS, F. S. G. (1961). A Method of Maximum-likelihood Estimation. Journal of the Royal Statistical Society, Series B, **23**, 469–475.

SAMPFORD, M. R. (1955). The truncated negative binomial distribution. Biometrika, **42**, 58–69.

SCHMIDT, F. H. and VAN DER VECHT, J. (1952). East monsoon fluctuations in Java and Madura during the period 1880–1940. Verhandelingen, No. 43, Djawatan Meteorologi dan Geofisik.

SHAPIRO, S. S. and FRANCIA, R. S. (1972). An Approximate Analysis of Variance Test for Normality. Journal of the American Statistical Association, **67**, 215–216.

SHAPIRO, S. S. and WILK, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, **52**, 591–611.

SHENTON, L. R. and BOWMAN, K. O. (1970). Remarks on Thom's estimators of the gamma distribution. Monthly Weather Review, **98**, 154–160.

SMITH, W. L. (1954). Asymptotic renewal theorems. Proceedings Royal Society Edinburgh, Series A, **64**, 9–48.

SMITH, W. L. (1959). On the cumulants of renewal processes. Biometrika, **46**, 1–29.

SMITH, R. E. and SCHREIBER, H. A. (1973). Point Processes of Seasonal Thunderstorm Rainfall, 1, Distribution of rainfall events. Water Resources Research, **9**, 871–884.

SMITH, R. E. and SCHREIBER, H. A. (1974). Point Processes of Seasonal Thunderstorm Rainfall, 2, Rainfall Depth Probabilities. Water Resources Research, **10**, 418–423.

SMITHSONIAN INSTITUTION (1927). World Weather Records. Smithsonian Miscellaneous Collections Volume 79, Washington DC.

SNEYERS, R. (1957). Sur la détermination de l'homogénéité des séries climatologiques. Institut Royal Météorologique de Belgique, Contributions No. 34.

STOL, PH. TH. (1970). Het vergelijken van empirische frequentieverdelingen met een toepassing op reeksen neerslaggegevens uit de Gelderse Achterhoek. Instituut voor Cultuurtechniek en Waterhuishouding, Wageningen, Mededelingen 129.

THOM, H. C. S. (1958). A note on the gamma distribution. Monthly Weather Review, **86**, 117–122.

TODOROVIC, P. and WOOLHISER, D. A. (1971). Stochastic Model of Daily Rainfall. Proceedings of the USDA-IASPS Symposium on Statistical Hydrology, Tucson, Arizona, pp. 232–246.

TODOROVIC, P. and YEVJEVICH, V. (1969). Stochastic process of precipitation. Colorado State University, Hydrology Paper No. 35, Fort Collins, Colorado.

VEN TE CHOW (1964). Handbook of Applied Hydrology, A compendium of Waterresources Technology. Mc Graw-Hill Book Company, New York.

WALLIS, J. R. and MATALAS, N. C. (1971). Correlogram analysis revisited. Water Resources Research, 7, 1448–1459.

WALLIS, JAMES R. and O'CONNELL, P. ENDA (1973). Firm reservoir yield-How reliable are historic hydrological records? Hydrological Sciences Bulletin, **XVIII**, 347–365.

WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. Journal of the Royal Statistical Society, Series B, **20**, 44–72.

WEISS, L. L. (1964). Sequences of wet and dry days described by a Markov chain probability model. Monthly Weather Review, **92**, 169–176.

WILLIAMS, C. B. (1952). Sequences of wet and dry days considered in relation to the logarithmic series. Quarterly Journal of the Royal Meteorological Society, **78**, 91–96.

WISE, M. E. (1946). The use of the negative binomial distribution in an industrial sampling problem. Journal of the Royal Statistical Society, Series B, **8**, 202–211.

WOOLHISER, D. A., ROVEY, E. and TODOROVIC, P. (1972). Temporal and Spatial Variation of Parameters for the Distribution of *N*-Day Precipitation. Proceedings of the Second International Symposium in Hydrology, Fort Collins, Colorado, pp. 605–614.

YEVJEVICH, VUJICA (1972). Stochastic Processes in Hydrology, Water Resources Publications, Fort Collins, Colorado, U.S.A.

YEVJEVICH, V. and JENG, R. J. (1969). Properties of Non-Homogeneous Hydrologic Series. Colorado State University, Hydrology Paper No. 32, Fort Collins, Colorado.