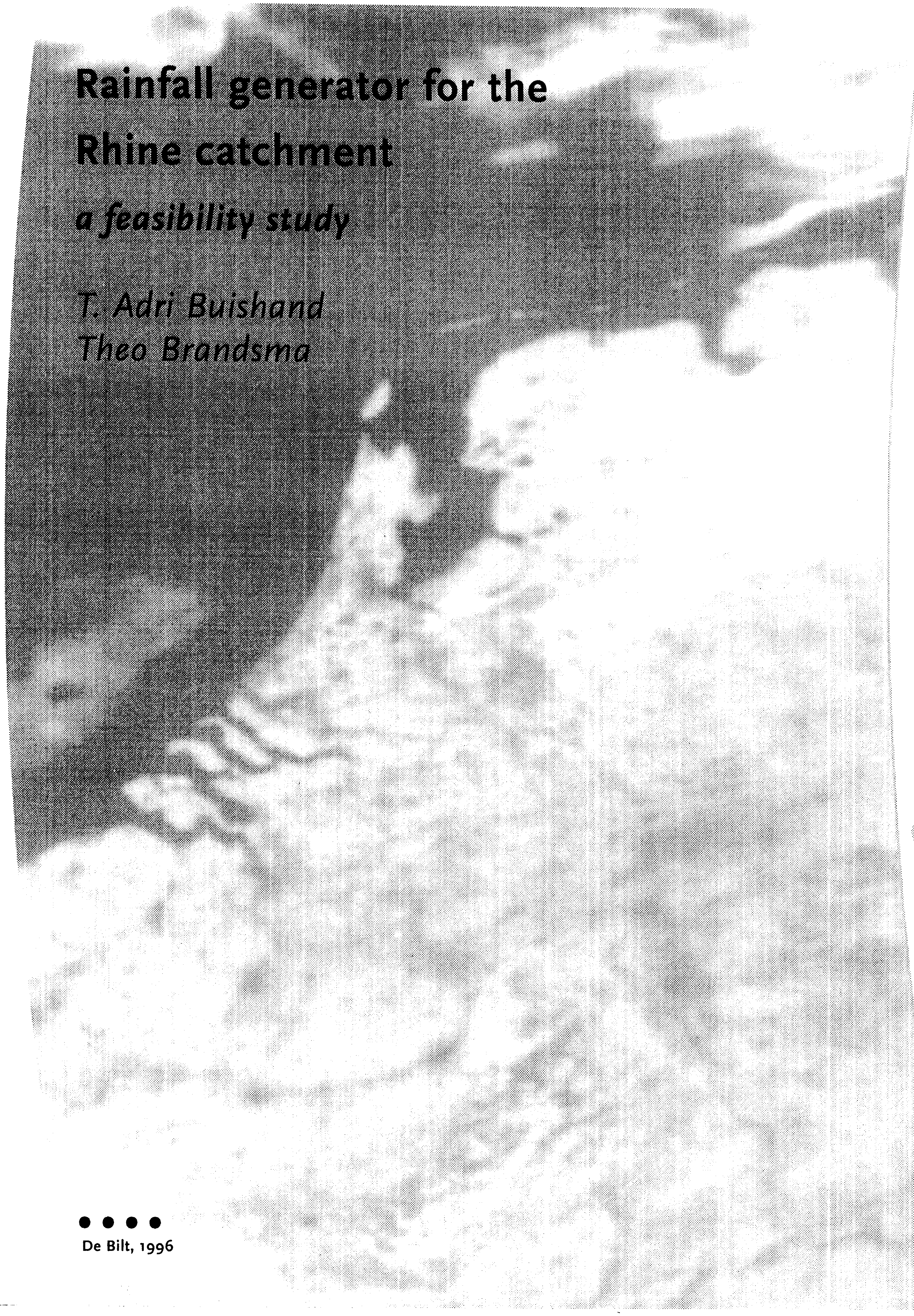
Technical report TR-183

KNMI

Rainfall generator for the Rhine catchment

a feasibility study

T. Adri Buishand
Theo Brandsma



De Bilt, 1996

Koninklijk Nederlands Meteorologisch Instituut

Technisch rapport TR-183

De Bilt, 1996

Postbus 201
3730 AE De Bilt
Wilhelminalaan 10
Telephone +31.30.220 69 11, telefax +31.30.221 04 07

UDC: 551.579.1
(282.243.1)
ISSN: 0169-1708
ISBN: 90-369-2096-5

Rainfall Generator for the Rhine Catchment

a feasibility study

T. Adri Buishand
Theo Brandsma

Technical Report TR-183

Work performed under contract RI – 1673 for Ministry of Transport, Public Works and Water Management, Institute for Inland Water Management and Waste Water Treatment RIZA,
P.O. Box 17, 8200 AA Lelystad (The Netherlands)
Telephone: +31.320.298411, telefax +31.320.249218

Preface

The planning of river dikes in the Netherlands is mainly based on the design water levels that are computed by RIZA and decided on by the Minister of Transport and Public Works. The design discharge is an important parameter in the determination of the design water levels. For the upper parts of the rivers in the Netherlands a design discharge with a recurrence time of 1250 years is applied, which is determined on the basis of statistical extrapolation of observed discharge maxima. To reduce uncertainties about the design discharge, RIZA aims at the development of an alternative method to determine this discharge. The method will first be developed for the River Rhine and is based on a hydrological/hydraulic model in combination with a statistical rainfall generator for the Rhine basin.

To study the possibilities for statistical rainfall generation on the scale of the Rhine basin in relation to extreme discharges, KNMI has carried out a preliminary study on this subject under contract to RIZA. The results of the study are presented in this report.

Bart Parmet
(Project leader RIZA)

Table of contents

Summary

Methods for generating multivariate daily rainfall sequences for the Rhine catchment are reviewed. The request for such a review arose from the need to study the likelihood of extreme river discharges in the Netherlands, using a hydrological/hydraulic model. The rainfall generator must provide simultaneous records of daily average amounts over about 30 subcatchments, each with an area of about 5000 km^2.

The autoregressive or AR model for univariate time series of normally distributed random variables can easily be extended to the multivariate case. Especially the multivariate first-order AR model has been used to generate multisite daily rainfall data. This model generates dry days as negative values. A transformation of the daily amounts (often a power transformation) is necessary to achieve normality.

Parameter estimation for the AR model is not straightforward because dry days must be treated as censored values. Replacing negative values by zeroes leads to an attenuation of autocorrelation and cross-correlation coefficients. Another problem is that the number of model parameters is quite large in the application to 30 subcatchments. A large reduction in the number of parameters is possible by introducing so-called contemporaneous models. Estimates of the lag-1 cross-correlation coefficients are then no longer required.

Promising results have been reported in the literature about the reproduction of a number of properties that are not included in the fitting procedure, like the extreme-value distribution of 1-day amounts, the decay of the autocorrelation coefficients with increasing lag, and the run-length distributions of wet and dry days, in particular in studies where the model parameters were linked to the atmospheric circulation. This is, however, not enough for the intended application to the Rhine catchment, because high river discharges strongly depend on the magnitude and spatial distribution of multi-day amounts during the winter half year, which have not been considered in the model evaluations.

The state of the atmospheric circulation is an important indicator of both the occurrence of rain and the rainfall amounts. Furthermore, the atmospheric circulation explains a part of the spatial and temporal correlation of rainfall. The link to the atmospheric circulation is usually achieved by relating the values of the model parameters to a classification of circulation (or weather) types. There are several methods to make such a classification. Two major categories can be distinguished: manual classification and automated classification. The British Lamb weather types and the German Grosswetterlagen are important examples of the first category while the objective Lamb classification scheme and the P-27 classification scheme, developed at KNMI, are important examples of the latter. Automated schemes are reproducible and much easier to use than a

manual classification. There are also more possibilities for generating time series of weather types when an automated classification is used.

A study of the behaviour of the atmospheric circulation during the winters of 1880/81 to 1992/93 revealed that especially in the last 20 years there have been some changes in the atmospheric circulation. However, the statistical significance of these changes is not strong.

Different approaches can be followed to simulate characteristics of the atmospheric circulation. A first approach is to simulate weather types with Markov chain models. A second aproach is to simulate circulation indices or principal component amplitudes using a multivariate parametric time series model. The latter approach only applies to the automated schemes.

The link between rainfall and the atmospheric circulation offers the possibility to generate multivariate daily values by nonparametric resampling techniques. This can be done by sampling from the days in the historical records with the same weather type as the day of interest or by choosing the day with the most similar circulation pattern (selection of analogues). These nonparametric alternatives to the multivariate AR model ensure that the spatial relationships between the 1-day amounts are preserved, but they have difficulties to reproduce the temporal dependence of daily rainfall. Modifications are thus necessary to achieve a satisfactory agreement between observed properties of multi-day amounts. Combination with a parametric model for the rain state or the daily average amounts of the whole catchment is a possible option to improve the reproduction of temporal dependence.

The multivariate AR model is a very suitable model for the joint simulation of daily values of different meteorological variables. It may thus provide additional daily temperatures to account for snow accumulation and snow melt in the Rhine catchment. Even when the precipitation sequences have been obtained by a resampling technique, it is possible to generate the concurrent temperatures with a multivariate parametric AR model. The normalising transformation for daily temperature is quite different from that of daily precipitation.

Statistical time series models may be used to generate synthetic data in the situation of an altered climate. This can be done by changing model parameters or by linking the time series model to the large-scale daily output of a General Circulation Model (statistical downscaling). For the case of a global warming, the effect of higher temperatures on precipitation in the Rhine catchment needs further study. Downscaling daily temperatures differs from the generation of daily temperatures for the present-day climate, because the former must be consistent with the values produced by the General Circulation Model.

A future project concerning daily rainfall generation in the Rhine catchment is delineated. Further evaluation of the multivariate AR model and improvement of non-parametric resampling techniques is given first priority. The German part of the catchment ($\approx$ 100 000 km^2) should be considered first. It is extremely important to verify the reproduction of properties of multi-day amounts in this phase. Statistical simulation of circulation characteristics and temperatures should be considered thereafter. By the end of that phase the need to incorporate extra long-term variation should be considered. Low priority is given to a new classification scheme of the atmospheric circulation. Model development for climate change studies is seen as a project on its own.

1. Introduction

The Rhine is the most important river in the Netherlands. The river flows through several countries (Figure 1). Large parts of its catchment are situated in Switzerland, Germany, France and the Netherlands. The maintenance and reconstruction of the dikes along the Rhine in the Netherlands is a point of continuous concern. An important piece of information for the design heights of these dikes forms the discharge record at Lobith (near Arnhem), where the river enters the country. Several probability distributions have been fitted to the discharge maxima of that record. Because of the enormous consequences of flooding, a safe design requires an extrapolation beyond the length of the observed record. Different distributions then lead to quite different design discharges. The fact that the parameters of these distributions have to be estimated from a finite record introduces another uncertainty.

In the most recent re-evaluation of the design discharge at Lobith the question arose whether the uncertainties of extrapolation could be reduced by taking the physical behaviour of the catchment into account (Delft Hydraulics and EAC-RAND,1993). For this purpose, a hydrological/hydraulic model for the whole catchment has to be developed. It further needs a statistical model to generate the spatial and temporal distribution of precipitation over the catchment to study unprecedented extreme situations. For instance, an unfavourable succession or spatial distribution of heavy rainfall may lead to more extreme discharges at Lobith than those experienced in the past century.

The aim of the present study is to review statistical techniques for generating multisite daily rainfall in the Rhine area, and to give an indication of the additional value of the use of these techniques for the determination of the design discharge at Lobith in the Netherlands. Multisite daily rainfall means here that simultaneous daily average amounts over about 30 subcatchments, each with an area of about 5000 km^2, are required. Some attention is given to the generation of temperature which is relevant for the distinction between rainfall and snow and for modelling snow melt. Accumulation and melt of snow are important aspects for discharge modelling of the River Rhine, especially for the southern parts of the basin. A brief discussion is also presented on time series modelling in the situation of an altered climate.

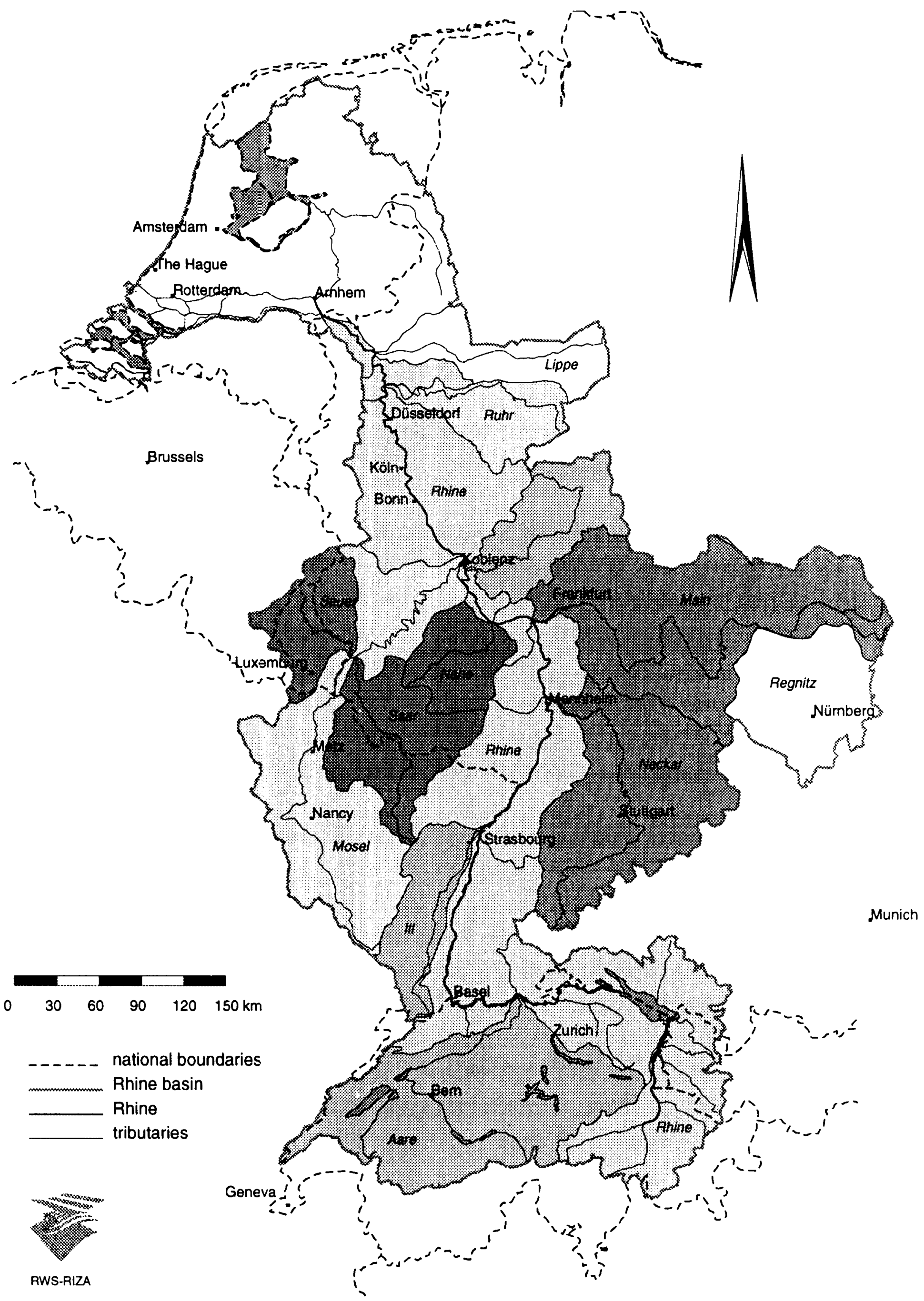


Figure 1: Drainage basin of the River Rhine.

The idea of statistical simulation was already used in hydrology in 1927 by Charles E. Studler who generated artificial sequences of annual runoff values by randomly drawing cards. This technique does not require mathematical expressions for probability distributions and is thus an example of a nonparametric approach. Nonparametric sampling of daily rainfall has recently been considered in the literature. Klemeš (1993) obtained the extreme upper tail of the annual maximum distribution of the daily precipitation amounts at Coquitham Lake in British Columbia by sampling all possible combinations of precipitable water, the synoptic storm efficiency and the orographic storm efficiency. Another recent development of nonparametric sampling is the generation of precipitation amounts at several sites in a region (Hughes *et al.*, 1993; Zorita *et al.*, 1995). In that method, resampling is restricted to those days in the record having more or less similar characteristics of the atmospheric circulation than the day of interest. Temporal dependence (autocorrelation) may cause serious problems in the nonparametric approach. Parametric time series models have been introduced to reproduce autocorrelation. In contrast to the nonparametric resampling techniques, the extension to the multisite case is not straightforward. In recent years, the linkage of parametric precipitation models to the atmospheric circulation has received much attention in the scientific literature (Bárdossy and Plate, 1992; Woolhiser *et al.*, 1993).

Many publications are restricted to precipitation modelling only. Most hydrological applications require, however, more meteorological input than precipitation. The joint simulation of precipitation and temperature is important for discharge modelling of the River Rhine both under present-day and altered climate conditions.

The report is organised as follows. Section 2 deals with parametric modelling of daily precipitation. Much attention is given to a multisite model based on the multivariate normal distribution. The classification and statistical modelling of the atmospheric circulation is addressed in Section 3. The long-term variation of the circulation and precipitation are also considered in that section. Section 4 is devoted to nonparametric resampling techniques. The inclusion of daily temperatures and the application to climate change impact assessment of extreme river discharges are discussed in Sections 5 and 6, respectively. The conclusions and recommendations are presented in Section 7.

2. Parametric modelling

Time series of daily rainfall are characterised by a large proportion of zero values. A popular method to model such data is to describe the occurrence of wet and dry days by a separate process (quite often a two-state Markov chain) and to represent the distribution of the precipitation amounts on the wet days by a gamma distribution or another skewed distribution, see Woolhiser (1992) for a review. This approach has, however, only been confined to data at a single site. The extension to multisite data over a large area meets serious difficulties. It takes for instance a large effort to reproduce days on which rainfall is restricted to a part of the area.

In the past ten years there have been many publications on rainfall modelling by clustered point processes, like the Neyman-Scott and Bartlett-Lewis rectangular pulses models (Rodríguez-Iturbe *et al.*, 1987, 1988; Cowpertwait, 1991; Onof and Wheater, 1994). These models give a stochastic description of the occurrence, duration and intensity of rain cells at a site. Cowpertwait (1994) investigated a multiple cell model, in which the 'heavy' short-duration cells were interpreted as convective precipitation and the 'light' long-duration cells as widespread rainfall. Although the development of clustered point process models is an active research area with continuous progress, their application is today mainly restricted to single-site data. The extension to spatial-temporal clustered point processes is still in an early stage (Cox and Isham, 1994; Chandler *et al.*, 1995; Cowpertwait, 1995). A spatial-temporal clustered point process model is basically continuous in space and time. Simplifying assumptions concerning the form, intensity and movement of cells are needed to estimate the model parameters from aggregated data. It is questionable whether such models can be fitted to spatial averages of daily amounts over areas as large as 5000 km^2.

The normal distribution plays a central role in the literature on multivariate time series models. It is in fact the only distribution which can easily be put in a multivariate framework. The multivariate normal distribution has therefore also been considered to generate multisite daily rainfall (Richardson, 1977; Bárdossy and Plate, 1992). Fitting time series models for normal random variables to positively skewed daily rainfall data with a large proportion of zero values needs, however, special care. To get a good insight into this problem we restrict our attention first to the generation of single-site data. The multisite extension is discussed thereafter. Limitations of the model are identified at the end of this section.

2.1 Generating single-site data by transforming a normal autoregressive process

This section is focused on the use of the first-order autoregressive or AR(1) process:

$$z_t = \varphi\, z_{t-1} + a_t, \qquad |\varphi| < 1 \tag{1}$$

The a_t's form a sequence of independent random variables with the same distribution (white noise). The variables z_{t-1} and a_t in the right-hand side of (1) are therefore independent. The parameter φ determines the correlation between the z_t's:

$$\text{corr}(z_t, z_{t-k}) = \varphi^k, \qquad k \geq 0 \tag{2}$$

In this report we restrict ourselves to positive autocorrelation, i.e. $\varphi > 0$.

When the a_t's come from a normal distribution, then the z_t's will also be normally distributed. For non-normal distributions the situation is more complicated (Gaver and Lewis, 1980). There is in fact only a limited class of distributions for which a stationary sequence with an AR(1) autocorrelation structure can be obtained using Equation (1). In any case, it is not possible to obtain a sequence where positive values and zeroes follow each other as in a daily rainfall sequence. This can easily be seen as follows. In a sequence of daily rainfall data there are cases that $z_{p-1} > 0$ and $z_p = 0$ and thus $a_p < 0$ if $\varphi > 0$. Hence it is necessary that a_t can take negative values. Let now $z_{s-1} = 0$. Because z_{s-1} and a_s are independent, it is possible that $a_s < 0$, which would result in a negative value of z_s.

To make use of a normal AR(1) process for generating daily rainfall data a transformation is applied such that the distribution of the transformed amounts looks like a normal distribution and generated negative values are replaced by zeroes. Quite often it has been assumed that a simple power transformation (square root, cube root) would be sufficient. The result of such a transformation is presented in Figure 2. The distribution

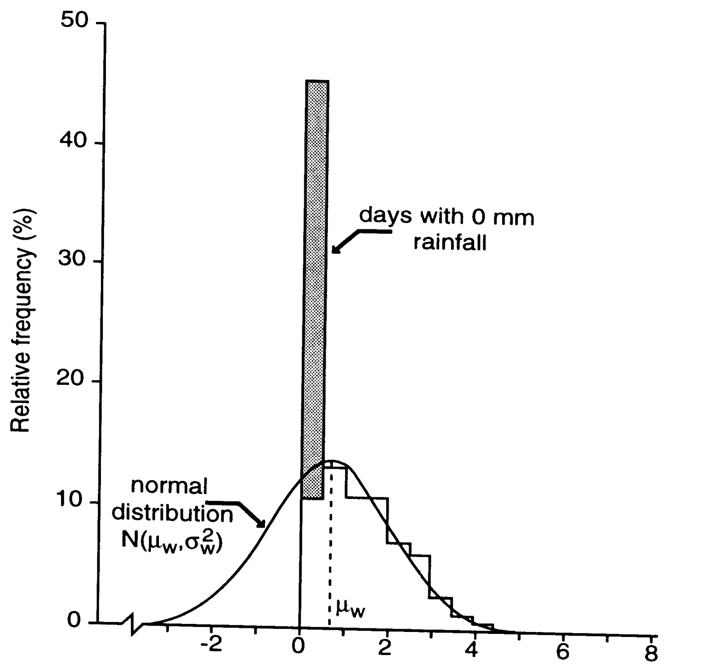


Figure 2: Histogram of the square root of the January daily rainfall amounts at a site in the Netherlands and probability density of a fitted normal distribution.

of the square root of the rainfall amounts on the wet days is quite well approximated by a normal distribution. The probability that the fitted normal distribution can take negative values is almost equal to the relative frequency of dry days.

The following five steps describe the generation of an artificial sequence of daily rainfall data.

Step 1 First a starting value z_0 is generated from a normal distribution with mean 0 and variance 1. It is also necessary to generate a sequence $a_1, a_2, \ldots$ of independent normal variables with mean 0 and variance $1-\varphi^2$.

Step 2 Equation (1) is applied for $t = 1,2,\ldots$ to get a sequence $z_0, z_1, \ldots$ of correlated variables with lag-one autocorrelation coefficient φ. Further $E(z_t) = 0$ and var $(z_t) = 1$.

Step 3 In order to achieve that the generated daily data have the correct mean and variance the following transformation is necessary:

$$w_t = \mu_w + \sigma_w z_t, \qquad t = 0,1,\ldots \tag{3}$$

Because this transformation is linear, the lag-one autocorrelation coefficient of the w_t's also equals φ. Further $E(w_t) = \mu_w$ and var $(w_t) = \sigma_w^2$.

Step 4 Negative values in the sequence $w_0, w_1, \ldots$ are replaced by zeroes:

$$y_t = \max(0, w_t), \qquad t = 0,1,\ldots \tag{4}$$

This non-linear transformation affects the autocorrelation structure. The lag-one autocorrelation coefficient $\rho_{yy}(1)$ of the y_t's is generally smaller than φ.

Step 5 Artificial daily rainfall data x_t are finally obtained by transforming the y_t's as:

$$x_t = y_t^{1/\lambda}, \qquad t = 0,1,\ldots \tag{5}$$

(λ = 1/2 corresponds to the square root transformation of the daily rainfall amounts, λ = 1/3 to the cube root). This transformation also affects the autocorrelation structure. The autocorrelation coefficients of the x_t's differ from those of the AR(1) process given by Equation (2).

The basic model parameters are φ, μ_w and σ_w. The problem is that transformation of the observed daily rainfall sequence does not supply the w_t's but the y_t's. Special techniques are therefore required to estimate the unknown model parameters.

2.1.1 Estimation of μ_w and σ_w

The w_i's in the generation algorithm have a normal distribution with mean μ_w and standard deviation σ_w.. As is seen in Figure 2, this distribution can take negative values. These negative values correspond to dry days. The problem is now to estimate the parameters μ_w and σ_w from the number m of dry days and the values $w_1,\ldots,w_n$ of the transformed rainfall amounts on the wet days.

In the statistical literature the term censored data is used for a sample that is curtailed at a known threshold w_0. Here w_0 corresponds to the smallest observed rainfall amount. The maximum likelihood method lends itself very well for parameter estimation from censored data. The maximum likelihood estimates are obtained by maximising the following log-likelihood:

$$L = m\ln[\Phi\{(w_0 - \mu_w)/\sigma_w\}] - \frac{1}{2}n\ln(2\pi) - n\ln(\sigma_w) - \frac{1}{2}\sum_{i=1}^{n}[(w_i - \mu_w)/\sigma_w]^2 \qquad (6)$$

where Φ is the standard normal distribution function. The first term in the right-hand side represents the contribution of negative values. $\Phi\ \{(w_0 - \mu_w)/\sigma_w\}$ is the probability of a dry day. The maximisation of L with respect to μ_w and σ_w is discussed in David (1981) and Hutchinson *et al.* (1993). The latter point out that the goodness of the fit is sensitive to the value of w_0. It is often advantageous to consider wet days with little rainfall as dry days.

Because μ_w and σ_w vary over the year the daily data are usually pooled by month (or season). The estimation procedure is performed for each pooled sample separately, assuming that these parameters are constant for the duration of a month (or season).

The log-likelihood in Equation (6) assumes that the daily values are independent. The maximisation of L still provides good estimates in case of autocorrelation. It is, however, no longer allowed to derive the standard errors of these estimates from the second-order derivatives of L.

In the literature on daily rainfall generation the term censoring is often confused with truncation. The latter refers to the conditional distribution of the values exceeding w_0. Truncation does not consider the number of observations below the threshold. The form of the log-likelihood function in case of truncation is almost identical to that for censored data (David, 1981). Ignoring the frequency of dry days leads to an increase in the standard errors of parameter estimates.

As an alternative to maximum likelihood estimation the parameters μ_w and σ_w can be chosen such that the frequency of dry days and the long-term monthly mean are exactly matched (Hutchinson *et al.*, 1993). This alternative might particularly be useful for climate change impact studies, because it directly gives the new model parameters in case of prescribed changes in the frequency of dry days and the long-term mean. Bárdossy and Plate (1992) used an additional expression for the second moment of the untransformed daily amounts to obtain parameter estimates. It is unknown how their method compares with maximum likelihood estimation. Another possibility is to base the estimates of μ_w and σ_w on a least squares fit to a probability plot of the transformed precipitation amounts (Marshall,1977). Probability plots are also helpful for the choice of the transformation and the threshold w_0 defining a wet day.

2.1.2 Estimation of φ

For samples from a normal distribution censoring leads to an attenuation of correlation coefficients. This effect has been studied analytically by Bhuiya and Yevjevich (1968) and Muthén (1990). Figure 3 shows that the attenuation is negligible if only a small fraction of the data is censored. It can, however, not be ignored when 50% of the days is dry (a realistic value for daily point precipitation in the Netherlands). The relation between $\rho_{yy}(1)$ and φ can be used to derive an estimate of φ from the estimated lag-one autocorrelation coefficient $\rho_{yy}(1)$ of the transformed daily rainfall amounts y_t..

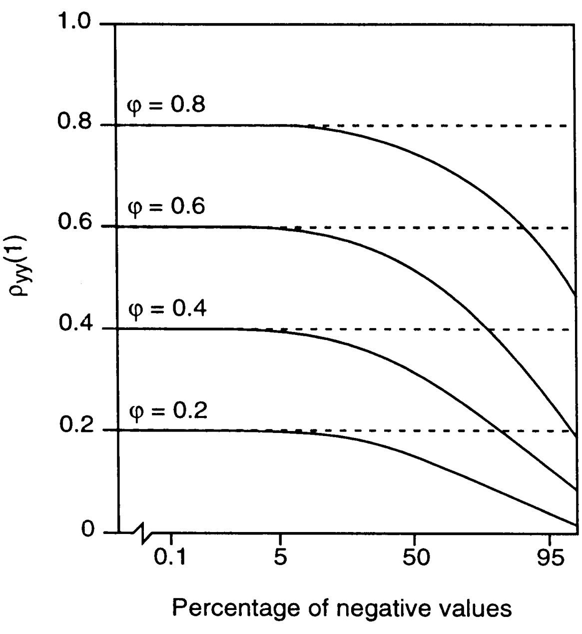


Figure 3: Change of the lag-one autocorrelation coefficient $\rho_{yy}(1)$ with percentage of censored data for normal AR(1) processes with φ = 0.2, 0.4, 0.6 and 0.8.

Bárdossy and Plate (1992) follow, however, a quite different approach using quantile exceedances. The pth quantile q_p in the distribution of daily rainfall amounts is the value that is exceeded with probability $1 - p$:

$$\Pr(x_t > q_p) = 1 - p \tag{7}$$

For any p greater than the proportion of dry days Bárdossy and Plate (1992) introduce the indicator series:

$$\begin{aligned} I_t &= 0 \quad \text{if} \quad x_t < q_p \\ I_t &= 1 \quad \text{if} \quad x_t \geq q_p \end{aligned} \tag{8}$$

where $t = 0,1,...$ Because for the wet days there is a monotonic relation between w_t and x_t, the variable I_t is also an indicator of an exceedance of the pth quantile of the w_t's.

The estimate of φ is based on the probability of two successive exceedances of the pth quantile:

$$p_{11} = \Pr(I_{t-1}, I_t = 1) \tag{9}$$

This probability increases with increasing φ. For instance, for $p = 1/2$ the value of p_{11} equals 1/4 for a purely random series ($\varphi = 0$) and p_{11} tends to 1/2 as $\varphi \rightarrow 1$. The estimate of φ from Equation (9) is known in the statistical literature as the tetrachoric correlation. This refers to the fact that the pair (I_{t-1}, I_t) can only have four possible outcomes, namely (0,0), (0,1), (1,0) and (1,1).

For $p = 1/2$ the value of φ follows directly from:

$$\varphi = \sin[2\pi(p_{11} - 1/4)] \tag{10}$$

Numerical methods are required for other values of p. This is necessary if the proportion of dry days is greater than 1/2.

2.2 Multisite extension

The daily precipitation amounts $x_{1t}, \ldots, x_{mt}$ at m different sites or over m different subcatchments are generally correlated. We therefore need a multivariate extension of the AR(1) process given in Equation (1). This extension is rather straightforward for the multivariate normal distribution. Matalas (1967) was the first who discussed the use of the multivariate normal AR(1) process for generating multisite synthetic sequences in hydrology. The process is defined by:

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{B}\boldsymbol{\varepsilon}_t \tag{11}$$

Here $\mathbf{z}_t$ and $\mathbf{z}_{t-1}$ are vectors of m correlated normally distributed random variables, each with mean 0 and variance 1. $\mathbf{A}$ and $\mathbf{B}$ are $m \times m$ matrices which control the autocorrelation and (lagged) cross-correlation coefficients between the elements of the $\mathbf{z}_t$'s. The random component $\boldsymbol{\varepsilon}_t$ is a vector of m independent normally distributed random variables with mean 0 and variance 1. The elements of $\boldsymbol{\varepsilon}_t$ are also independent of the elements of $\mathbf{z}_{t-1}$.

For the vectors $\mathbf{z}_t$ we can define the following $m \times m$ cross-correlation matrices:

$$\mathbf{M}_0 = \mathrm{E}(\mathbf{z}_t\mathbf{z}_t^T); \qquad \mathbf{M}_1 = \mathrm{E}(\mathbf{z}_t\mathbf{z}_{t-1}^T) \tag{12}$$

where T stands for the transpose of a vector or matrix. The elements on the main diagonal of $\mathbf{M}_0$ are all equal to 1, because the elements of $\mathbf{z}_t$ have unit variance. The off-diagonal elements of $\mathbf{M}_0$ are the lag-zero cross-correlation coefficients. The matrix $\mathbf{M}_1$ contains the lag-one cross-correlation coefficients. The diagonal elements of $\mathbf{M}_1$ are the lag-one autocorrelation coefficients.

The matrices $\mathbf{M}_0$ and $\mathbf{M}_1$ are related by:

$$\mathbf{M}_1 = \mathbf{A}\mathbf{M}_0 \tag{13}$$

from which it follows:

$$\mathbf{A} = \mathbf{M}_1\mathbf{M}_0^{-1} \tag{14}$$

An estimate of $\mathbf{A}$ can be obtained from the lag-zero and lag-one cross-correlation coefficients. These correlation coefficients are also affected by censoring. The results in Muthén (1990) can be used to determine the magnitude of the attenuation or the correlation estimates should be based on the joint exceedances of a quantile.

The vector $\mathbf{a}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$ corresponds with the noise term a_t in Equation (1). The matrix $\mathbf{B}$ converts the independent elements of $\boldsymbol{\varepsilon}_t$ into a vector of m correlated normal variables. The covariance matrix $\mathbf{C}$ of $\mathbf{a}_t$ is given by:

$$\mathbf{C} = \mathrm{E}(\mathbf{a}_t\mathbf{a}_t^T) = \mathbf{B}\mathbf{B}^T = \mathbf{M}_0 - \mathbf{M}_1\mathbf{M}_0^{-1}\mathbf{M}_1^T \tag{15}$$

This covariance matrix follows thus from $\mathbf{M}_0$ and $\mathbf{M}_1$. The matrix **B** can then be obtained from the Choleski decomposition (Young, 1968) or the spectral decomposition (Bárdossy and Plate, 1992) of **C**. Further details can be found in textbooks on statistical hydrology, like Salas *et al.* (1980) and Bras and Rodríguez-Iturbe (1985). A similar decomposition of $\mathbf{M}_0$ is necessary to get a starting value $\mathbf{z}_0$ in the generation process.

A problem with the model given in Equation (11) is that the number of parameters rapidly grows with the number of subcatchments. There are $m(m-1)/2$ different lag-zero cross-correlation coefficients and m^2 lag-one cross-correlation coefficients. For $m = 30$ we thus have $(3m^2-m)/2 = 1335$ different correlation coefficients. Most of these correlation coefficients show seasonal variation. The smallest correlation coefficients are found in the summer season due to the relatively large contribution of short lasting local showers on daily rainfall in that season. Dividing the year into s different seasons to account for seasonal variation implies in fact that the total number of correlation coefficients becomes as large as $s(3m^2-m)/2$.

It is thus desirable to think about reduction of unknown parameters. One possibility is to take the matrix **A** to be a diagonal matrix (Haltiner and Salas, 1988). Each element of $\mathbf{z}_t$ then behaves as an AR(1) process and the correlation between the sites only enters via the correlation of the elements in the noise term ε_t. This multivariate model is known as the contemporaneous AR(1) or CAR(1) model. The diagonal elements of **A** are the lag-one autocorrelation coefficients of the elements of $\mathbf{z}_t$. The lag-one cross-correlation coefficient $\rho_{uv}(1)$ between the uth and the vth element of $\mathbf{z}_t$ is given by:

$$\rho_{uv}(1) = \rho_{uv}(0)\rho_{uu}(1) \tag{16}$$

where $\rho_{uv}(0)$ is their lag-zero cross-correlation coefficient and $\rho_{uu}(1)$ is the lag-one autocorrelation coefficient of the uth element. An important property of the CAR(1) process is that the lagged cross-correlation coefficients can be factored in a spatial and a temporal component. A detailed discussion of contemporaneous models is presented in Hipel and Mc Leod (1994).

With respect to the above-mentioned example, the number of correlations coefficients decreases from 1335 to 465 which is a reduction of 65%. Methods to achieve a further reduction of unknown parameters have not been considered yet in the hydrological literature.

2.3 Model limitations

Stochastic simulation of daily rainfall can only be successful when the generated data reproduce the essential characteristics of observed records. We may expect that the transformed normal AR(1) process preserves the mean, the standard deviation and the lag-one autocorrelation coefficient of the daily precipitation amounts even though the parameters μ_w, σ_w and φ refer to the transformed amounts. To be meaningful for the estimation of design discharges at Lobith it is, however, required that the distribution of extreme multi-day events can be reproduced, in particular for the winter half year. In addition to the univariate extremes, the spatial dependence of large values is also of interest.

In this section we first consider some aspects of the distribution of N-day amounts. Then the reproduction of the spatial correlation coefficients of the N-day amounts by the transformed multisite normal AR(1) process is dealt with. Finally the distinction between correlation and dependence, which is important for the distribution of multivariate extremes, is discussed.

2.3.1 Some remarks on the distribution of N-day amounts

In the transformed normal AR(1) process the parameters μ_w, σ_w and λ determine the distribution of the 1-day amounts. The transformation parameter λ strongly affects the exceedance probabilities of large values. Although this parameter can be estimated by the method of maximum likelihood (Hutchinson *et al.*, 1993), the choice of the transformation must include a careful examination of the occurrence of extreme values. The censoring threshold w_0 may also influence the quality of the fit in the upper tail. Bárdossy (1994) observed that the 1-day annual maximum distribution at a site in the Ruhr catchment (a tributary of the River Rhine, see Figure 1) and the 1-day maximum average rainfall over the Lenne subcatchment in the winter half year were well reproduced.

For $N > 1$ the distribution of the N-day amounts also depends on the autocorrelation structure. The following relation exists between the variance V_N of the N-day amounts and the autocorrelation coefficients:

$$V_N = \sigma_x^2 \left[N + 2 \sum_{k=1}^{N-1} (N-k) \rho_{xx}(k) \right] \tag{17}$$

with $\sigma_x^2 = V_1$ the variance of the 1-day amounts and $\rho_{xx}(k)$ their lag-k autocorrelation coefficient $(k = 1, \ldots, N-1)$.

From Equation (17) it is clear that positive autocorrelation leads to relatively large values of V_N. Time series of natural phenomena often exhibit long-term persistence, which says that $\rho_{xx}(k)$ decreases slowly with increasing k, so that distant terms still have a marked influence on quantities like V_N. It is often important that a time series model can reproduce this characteristic of the data. Lack of long-term persistence implies that V_N is too low. The model then generates too few extreme N-day amounts. Time series models that use a separate process for the generation of wet and dry days often underestimate the variances of the monthly amounts (Buishand, 1977a; Woolhiser, 1992; Gregory *et al.*, 1993). Lack of long-term persistence generally also implies that the frequency of long runs of wet or dry days is underestimated.

The autocorrelation function of the underlying AR(1) process given by Equation (1) falls off quite rapidly. However, the transformations in Steps 4 and 5 of the generation algorithm in Section 2.1 distort the autocorrelation structure and may lead to a quite different decay of the autocorrelation coefficients with increasing lag. The effect of a logarithmic transformation on the autocorrelation structure has been studied by Mejía and Rodríguez-Iturbe (1974). Here we present a small simulation study to show that some long-term persistence is introduced by our generation algorithm. Table 1 compares the variances of the N-day amounts generated from a transformed normal AR(1) process with those derived from the autocorrelation function of an untransformed AR(1) process. Both the square root $(\lambda = 1/2)$ and cube root $(\lambda = 1/3)$ transformation are considered in Step 5 of the generation algorithm. The parameters σ_x and $\rho_{xx}(1)$ of the untransformed AR(1) process are the same as those for the simulated transformed AR(1) process. The variances are therefore identical for $N = 1$ and $N = 2$ in Table 1. For the 10-day and 30-day amounts the relative differences are 5 and 10 %, respectively, both for the case $\lambda = 1/2$ and $\lambda = 1/3$. The larger variances for the transformed process may lead to a better reproduction of the distribution of extreme N-day amounts. It is further reported in Hutchinson (1995) that the transformation of the normal AR(1) process performs much better than the first-order two-state Markov chain in matching the observed run-lengths of dry days.

The reproduction of the variances of the N-day amounts can further be improved by: (a) distinguishing between circulation types which results in extra persistence; (b) introducing an extra parameter in the autocorrelation function of the underlying time series model. The latter can be achieved by taking an ARMA(1,1) model. The incorporation of the atmospheric circulation in daily precipitation models is discussed in Section 3.

N	Transformed, $\lambda = 1/2$	Untransformed	Ratio	Transformed, $\lambda = 1/3$	Untransformed	Ratio
1	1.3	1.3	1.00	6.7	6.7	1.00
2	3.2	3.2	1.00	16.0	16.0	1.00
5	9.6	9.6	1.00	46.0	45.6	1.01
10	21.3	20.3	1.05	100.0	95.3	1.05
30	70.2	63.4	1.11	326.8	294.0	1.11

Table 1: Variances of N-day amounts of a simulated transformed normal AR(1) process compared with those of an untransformed AR(1) process. The parameters for the transformed AR(1) process are $\varphi = 0.4$, $\mu_w = 0$ and $\sigma_w = 1$. Number of simulated years $J = 100$.

2.3.2 Spatial correlation coefficients of N-day amounts

In Section 2.2 we introduced the CAR(1) process to reduce the number of unknown parameters in the multisite AR(1) process. It was also noted in that section that the lagged cross-correlation coefficients between the elements of the CAR(1) process can be factored in a spatial and a temporal component. A consequence of this space-time separability is that the correlation coefficients between the N-day totals of these elements do not change with N (Buishand, 1977b). Precipitation data show, however, an increasing cross-correlation with increasing period of aggregation (Buishand, 1977b; Cox and Isham, 1994). The relatively low correlation coefficients between the 1-day amounts are due to small-scale events (local showers). Their influence on the cross-correlation coefficients of the N-day amounts decreases with increasing N. In the Netherlands the strongest change in the cross-correlation coefficients is found in the summer season, where a relatively large proportion of the total rainfall is due to convective precipitation. But also in the winter season there is still a marked increase of the correlation coefficients with the period of aggregation.

Table 2 presents the correlation coefficients between the N-day amounts at Twente and Gemert (distance 115 km) for the winter season (December-February). The increase in the correlation coefficients with increasing N is well reproduced by the 18-year simulation run of the transformed bivariate normal CAR(1) process, even though these correlation coefficients are constant in the original CAR(1) process. The transformations in the generation algorithm in Section 2.1 thus also produce a rather realistic change of the cross-correlation coefficients with the period of aggregation. It should be noted, however, that quite long records are needed to obtain an accurate estimate of the change of the cross-correlation coefficients with N. The results for the 100-year run in Table 2 differ markedly from those for the 18-year run.

N	Twente-Gemert (J = 18)	Transformed CAR(1) process, $\lambda = 1/2$ J = 18	J = 100
1	0.771	0.774	0.792
2	0.808	0.791	0.802
3	0.821	0.798	0.805
4	0.830	0.802	0.808
5	0.836	0.805	0.810
10	0.855	0.840	0.820
30	0.870	0.857	0.829

Table 2: Correlation coefficients between N-day amounts for a transformed bivariate normal CAR(1) process compared with those for Twente and Gemert in the Netherlands during the winter seasons (December-February) 1952/53,..., 1969/70 (after Buishand, 1977b). The parameters in the underlying bivariate AR(1) process are $\rho_{uv}(0) = 0.85$, $\rho_{uu}(1) = \rho_{vv}(1) = 0.40$, $\mu_w = 0$ and $\sigma_w = 1$. J is the record length in years.

For the simulated transformed data in Table 2 the parameter μ_w was zero. This implies that the probability of a dry day is 0.5, which is a realistic value for daily point precipitation in the Netherlands. The parameter σ_w appears as a scaling factor in Equations (3), (4) and (5). It does not influence the value of the cross-correlation coefficients. The effect of the transformations on these correlation coefficients is completely determined by the proportion of dry days and the value of the parameter λ.

2.3.3 Association of large values

In the foregoing sections much attention has been given to the reproduction of correlation coefficients. This is, however, not sufficient to preserve the temporal and spatial dependence of large values. Dependence has a much wider meaning than correlation which refers to linear dependence only.

The degree of dependence between the precipitation amounts u_t and v_t at two different sites is determined by the form of their joint probability distribution. Here we will look at the joint probability $G(p) = \Pr(u_t \le u_p,\ v_t \le v_p)$, with u_p and v_p the pth quantiles in the marginal distributions of u_t and v_t. If u_t and v_t are independent, then

$$G(p) = \Pr(u_t \le u_p)\ \Pr(v_t \le v_p) = p^2 \tag{18}$$

whereas in case of complete dependence

$$G(p) = \Pr(u_t \le u_p) = \Pr(v_t \le v_p) = p \tag{19}$$

More general, we may write $G(p)$ as

$$G(p) = p^{h(p)} \tag{20}$$

Here $h(p)$ determines the degree of association at the pth quantile. There is little association if $h(p)$ is close to 2. If this holds for large p, then the probability is small that both u_t and v_t are extreme. The extremes tend to occur simultaneously when there is strong association in the upper tail, i.e. when $h(p)$ is close to 1 for large p.

Preservation of dependence requires that functions like $h(p)$ are reproduced rather than correlation coefficients. For 1-day maxima in the winter half year in the Netherlands Buishand (1984) found an increase of spatial association with event magnitude at distances greater than 50 km. This is different from that expected from the multivariate normal distribution, which is characterised by a relatively weak association in both tails. For bivariate normal variables with correlation coefficient < 1 the function $h(p)$ tends to 2 as $p \to 0$ or $p \to 1$. This remains true after a monotonic transformation like that in

Step 5 of our generation algorithm. Multisite models relying on joint normality may therefore underestimate the frequency of concurrent extremes, like high winter precipitation or droughts over large geographic areas (Leytham, 1987).

Estimates of the degree of association at large quantiles (p close to 1) suffer from high sampling variability. In Buishand (1984) a reduction in sampling variability was achieved by pooling the exceedances from all pairs of stations with comparable inter-station distances.

2.4 Conclusions

A tractable time series model for multisite daily rainfall data can be obtained from a multivariate normal autoregressive process. Transformation of this process is necessary to reproduce the skewed distribution of daily rainfall data. The effect of replacing negative values by zeroes on the correlation coefficients needs special care. Various methods of parameter estimation have been proposed. However, for most estimators it is not known how they compare with others. Contemporaneous models can be used to reduce the number of unknown parameters.

The transformations to obtain a realistic distribution of the 1-day amounts introduce some long-term persistence and an increase in the cross-correlation coefficients with the period of aggregation. These properties are also observed in real rainfall data. The transformations can, however, not assure that the variances and correlation coefficients of the multi-day amounts are preserved. Furthermore, the underlying multivariate normal distribution may not be capable to reproduce the dependence of extreme values. These aspects have not been considered yet in the validation of parametric time series models. The large variances of estimated second moments of multi-day amounts and features of extremes may impede such a validation. Regional estimation of extremes and pooling or averaging information from pairs of subcatchments may alleviate this problem.

The literature indicates that the 1-day annual maximum distribution can reasonably be reproduced and there are also possibilities to achieve a satisfactory agreement between the variances of the observed and simulated N-day amounts. The reproduction of the multi-day annual maximum distribution and the joint occurrences of large N-day amounts at different subcatchments during winter remain questionable. This needs to be investigated further.

3. Atmospheric circulation

Precipitation is partly controlled by the atmospheric circulation. Firstly, the occurrence of rain and the rainfall amounts depend on the pressure distribution. Secondly, the persistence of the large-scale circulation has a strong effect on the autocorrelation coefficients of the daily rainfall amounts. For models of daily point rainfall, Woolhiser *et al.* (1993) and Katz and Parlange (1993) found that the variances of the monthly amounts were better reproduced when the model parameters were perturbed by a circulation index. Finally, linkage of a multisite rainfall model to the large-scale atmospheric circulation might improve the reproduction of the spatial patterns of daily rainfall. Especially in the past few years the conditional modelling of daily rainfall on atmospheric circulation characteristics has received much interest for downscaling large-scale output of General Circulation Models (GCMs) to produce scenarios for climate change impact studies.

The present section deals with the use of atmospheric circulation schemes in relation to daily rainfall. Firstly, we present a general introduction to the topic of atmospheric circulation classification. The emphasis is on those classification schemes referring to the area of the River Rhine catchment. Secondly, we discuss the behaviour of the atmospheric circulation during this century and its possible influence on precipitation. Finally, we consider the possibilities to generate the succession of circulation types.

3.1 Classification schemes

The large-scale atmospheric circulation is an important driving force of the surface climate. Not surprisingly, several attempts have been made to relate the atmospheric circulation to the local weather elements. In general, the first stage is the classification of the atmospheric circulation and the second stage is the assessment of the relationship between the categories of the classification and the region's weather elements. In fact we are relating physical processes at varying scales in order to generate weather elements at the smaller scales.

Figure 4 presents the two major procedures that have been used to develop classification schemes for the atmospheric circulation: manual classification and automated classification.

The results of manual classification are unique because they are based on the investigator's subjective interpretations: no two investigators will produce the same results. On the other hand, automated classification generates reproducible categories and reduces the time and labour spent on the classification process. Although the automated classifications often pretend to be objective, subjective decisions still have to be made. Those decisions involve e.g. the size of the grid and the number of grid points, the type of pressure variable (surface pressure or 500 hPa heights), class boundaries and the number of classes. Both classification procedures are discussed below in more detail with a number of examples.

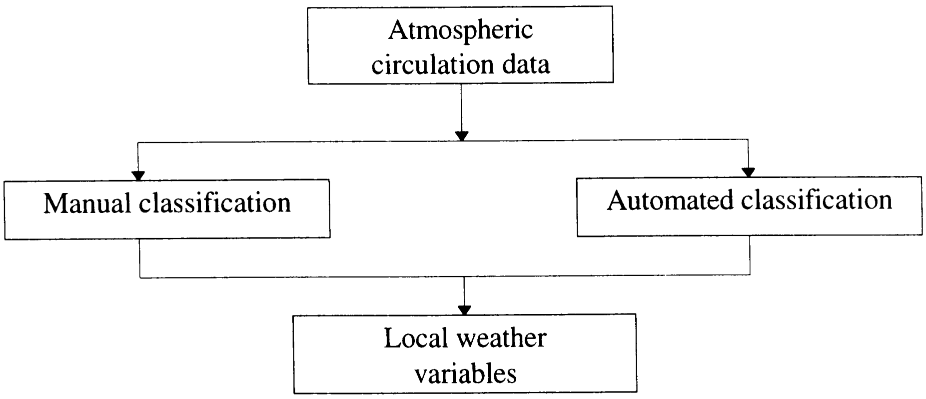


Figure 4: Two major procedures used for relating the atmospheric circulation to local weather variables.

3.1.1 Manual classification

In all manual schemes based on weather maps, investigators use a holistic combination of all weather elements, in addition to the pressure patterns, to determine the ultimate classification of the maps. As a result, the investigator can control the classification completely. However, manual classifications suffer two major drawbacks: (1) they are labour intensive by definition; and (2) they cannot be duplicated.

There are several highly regarded manually derived classification schemes available in the literature like the Lamb weather types, the Muller classification, the Grosswetterlagen and the cyclone model. These classifications have been applied to a wide range of environmental problems. The main difference among the various classifications is their regional focus. Here we restrict ourselves to the Lamb weather types and the Grosswetterlagen only because they are both developed for Europe.

Lamb weather types

The British climatologist H.H. Lamb developed a synoptic-scale, daily weather-map classification for the British Isles. The original catalogue extends from 1861 to 1971. Because Lamb has been the only analyst to classify the data, this long synoptic record is internally consistent. Lamb (1972) recognised seven basic weather types: Anticyclonic (A), Cyclonic (C), Westerly (W), Northwesterly (NW), Northerly (N), Easterly (E), and Southerly (S). Furthermore, Lamb also identified unclassifiable days and hybrid types. Each day is classified by the general circulation patterns throughout the whole day and not at a particular time. Altogether, the classification has 27 possible types, including unclassifiable days.

Several investigators have related the Lamb weather types to precipitation. Jones and Kelly (1982) found that the first two principal components (PCs) of the annual frequencies of the weather types for the period 1861–1980 were significantly correlated at the 5% level with the England and Wales annual rainfall series. The total amount of variance accounted for by PC1 and PC2 was 63%. Briffa *et al.* (1990) expanded the work of Jones and Kelly to include seasonal frequencies. Wigley and Jones (1987) studied, among others, the correlation coefficients between W, A and C types and rainfall for each season for six regions in England and Wales. For A and C types there is a strong correlation in all seasons and for all regions.

Wilby (1994) presented a stochastic precipitation model for regional climate change impact assessment. The model uses the Lamb weather types to simulate daily point precipitation at two sites in Central and Southern England. Wilby *et al.* (1994) applied the same approach to simulate daily precipitation and they used the simulated precipitation data as input to a rainfall-runoff model to simulate daily flows. In a more recent publication (Wilby, 1995), the incorporation of a frontal submodel within the daily rainfall generator is considered.

Grosswetterlagen

Like the Lamb weather types, the Grosswetterlagen of the Deutscher Wetterdienst (German weather service) are a well known example of manual classifications. An important difference between the Grosswetterlagen and the Lamb weather types is that the former apply to Europe which is a much larger area than the British Isles.

Over several decades, Franz Baur developed the Grosswetterlagen concept using surface pressure patterns and knowledge of the observed weather elements. Hess and Brezowsky defined 29 weather types. A list of these circulation patterns and the corresponding meteorological descriptions can be found in Hess and Brezowsky (1969). The latter authors modified the system to include the upper-air data that became available after World War II. Gerstengarbe and Werner (1993) presented a catalogue of Grosswetterlagen extending from 1881 through 1992. Since 1949, the Grosswetterlagen have been determined using both surface and upper-air charts. Older Grosswetterlagen were prepared from surface weather maps only.

During a Grosswetterlage, the main features of the weather are almost constant over Central Europe. Each Grosswetterlage usually lasts three days or more, after which there is a transition to another circulation type. Each day is assigned to one of the 29 Grosswetterlagen or an unclassifiable category. For convenience, the Grosswetterlagen can be broken down into three general categories: zonal, meridional, and mixed.

The Grosswetterlagen have been used in several studies. To understand the regional effects of climate change, Bárdossy and Caspary (1990) studied the temporal and spatial associations of Grosswetterlagen with temperature and precipitation for the period 1881–1989. Since the year 1973, they found an increase in the frequency of zonal types in the months of December and January and a corresponding decrease in meridional circulation types. The climatological consequences of this are more frequent mild and humid winters in central Europe, with precipitation mostly falling as rain.

Bárdossy and Plate (1992) linked a multisite daily rainfall model to the Grosswetterlagen. The model was based on the multivariate AR(1) process given in Equation (11). In their coupled version both the parameters μ_w and σ_w of the underlying normal distribution and the matrices **A** and **B** depend on the circulation type. The model was fitted to daily data from 44 rainfall stations in the Ruhr catchment over a 14-year period. This period is rather short to obtain good estimates of the various parameters of the multivariate AR(1) process for each circulation type. To reduce the number of unknown parameters, circulation types with similar rainfall statistics were put together to form new groups. In this way, the 30 circulation types were reduced to 11 circulation groups. For the spatial correlation coefficients, the number of groups could even be reduced to four (Bárdossy, 1993). Seasonality was incorporated by dividing the year into a winter and summer half year. The approximation of the annual cycle in the mean and standard deviation is therefore rather rough.

Bárdossy and Plate (1992) observed that for a number of thresholds $\leq$ 10 mm the model reproduced the bivariate distribution of 1-day amounts quite well. This was also the case for the average proportion of stations with rainfall under the condition that the daily maximum over the catchment exceeded a given limit. They further reported a good agreement between the simulated and observed run-length distributions of wet and dry days and that the decay of the autocorrelation coefficients with increasing lag agreed well with that of the observed data. More details about this can be found in Bárdossy

(1993). In a later publication Bárdossy (1994) showed that the annual maximum distribution of dry run-lengths was well preserved. Satisfactory results were also obtained for the 1-day annual maxima (Section 2.3.1). The distribution and spatial dependence of multi-day extremes were not considered.

Schubert (1994) developed a weather generator based on the Grosswetterlagen for point meteorological data. Van der Wateren-de Hoog and Middelkoop (1992) tried to relate the Grosswetterlagen to high water stages of the River Rhine at Lobith in the Netherlands. Their results were, however, somewhat disappointing: a clear relationship between the Grosswetterlagen and the occurrence and magnitude of high river stages could not be found.

3.1.2 Automated classification

Automated schemes are quick and easy to use and, in contrast to manual schemes, investigators can replicate the results. The automated schemes are frequently referred to as objective schemes. It was already indicated before that they still involve a number of subjective assumptions which may influence the relationships with the local weather considerably. So whenever we use the term objective in this report, we mean objective only in the sense that the classification is automated and that the results can easily be replicated.

Most automated classification schemes have a strong connection with cluster analysis in the statistical literature. The latter is concerned with the division of observations into groups or clusters in situations that there is little or no prior information about the structure of these groups. Yarnal (1993) presented an overview of the methods used in synoptic climatology during the past decades. Popular methods are the correlation-based map-pattern classification, the eigenvector-based classification and indexing. In the correlation-based map-pattern classification the pattern correlations between maps of mean sea level pressure (MSLP) are determined and the maps are classified and reclassified according to selected keydays. Before the analysis the maps of MSLP are standardised to account for seasonality. The correlation coefficient as a measure of pattern similarity is then uniquely related to the Euclidean distance in cluster analysis. In the eigenvector-based classification the number of variables is reduced first using techniques like principal component analysis (PCA). The class boundaries are then based on the reduced set of variables. Finally, indexing is a way to characterise variations in the atmospheric circulation with a simple time series of single numbers. These numbers may represent e.g. the pressure difference between two stations. Some examples of these methods are discussed below.

Objective Lamb classification scheme

The objective Lamb classification scheme is an example of the indexing method whereby several indices are defined. Initially it was developed by Jenkinson and Collison (1977) for the British Isles. For each day the following air-flow indicators are derived from MSLP data on a regular grid:

W_t = strength of the westerly flow (using the pressure gradient in the north-south direction);

S_t = strength of the southerly flow (using the pressure gradient in the west-east direction);

ZW_t = westerly shear vorticity (a second order difference of the pressure field);

ZS_t = southerly shear vorticity (another second order difference of the pressure field).

From these indices three other indices are obtained: (1) the direction of the flow as $\tan^{-1}(W_t / S_t)$, where 180° is added when W_t is positive; (2) the strength of the flow $F_t = (S_t^2 + W_t^2)^{1/2}$; and (3) the total shear vorticity $Z_t = ZW_t + ZS_t$. The latter is a measure of the rotation of the atmosphere. Positive vorticity corresponds to a low pressure area (cyclonic) and negative vorticity corresponds to a high pressure area (anti-cyclonic). The values of the three air-flow indicators determine the weather type of the day considered. The full classification distinguishes 26 weather types plus one type denoted as undefined. It was shown by Jones *et al.* (1993) that the resulting scheme quite well reproduced the subjective weather typing scheme of Lamb (1972) for the British Isles.

The MSLP data are available on a daily basis for most of the Northern Hemisphere on a 5° latitude by 10° longitude grid extending back to December 1880.

Jones *et al.* (1993) used the objective scheme to classify daily circulation types over the British Isles, along the lines of the subjective method devised by Lamb, for the period 1881– 1989. The aim of the objective scheme is not to replace the Lamb classification but to provide a scheme that produces an acceptable surrogate for the original, that might be readily applied to the daily pressure fields generated by GCMs. They found that the frequencies of the objectively developed types are highly correlated with the traditional Lamb types. They also found rather large correlation coefficients between England and Wales rainfall and the numbers of anticyclonic and cyclonic days as obtained from the objective Lamb scheme.

An important advantage of the objective Lamb scheme is that it can be applied to other parts of Europe as well. For instance, Brandsma and Buishand (1995) calculated the objective Lamb weather types for the Netherlands and compared the performance of this scheme with that of the Grosswetterlagen and the objective P-27 scheme (discussed below) with respect to the prediction of daily temperature and occurrence of rain days. For rainfall occurrence it appeared that the three methods are almost equivalent, but for temperature the objective Lamb weather types are inferior. The classification schemes explained only a quarter of the climatological variance of wet and dry days. Buishand and Brandsma (1995) compared the objective Lamb classification scheme with the Grosswetterlagen concerning their ability to predict monthly temperature and rainfall for De Bilt and the Netherlands. For the prediction of monthly precipitation amounts it appeared that the Lamb scheme performs equally well as the Grosswetterlagen. Both schemes perform better for area-average rainfall over the Netherlands than for rainfall at De Bilt.

Hulme *et al.* (1993) used the objective Lamb weather types to validate two GCM control simulations. The GCMs examined were the UK Meteorological Office high-resolution atmospheric model (UKHI) and a coupled ocean/atmosphere model of the Max-Planck-Institut für Meteorologie, Hamburg (MPI). It appeared that both GCMs are too cyclonic in winter. The seasonality of both anticyclonic and cyclonic types was much too strong in the MPI model and the summer precipitation (as derived from the weather types in this GCM simulation) was much lower than the observed summer precipitation. The MPI model simulated the annual cycle of temperature well, while UKHI successfully reproduced the annual cycle of precipitation.

P-27 classification scheme

The P-27 classification scheme is an objective scheme developed at KNMI (Kruizinga, 1978, 1979). It is an example of an eigenvector-based classification scheme known as principal component analysis (PCA). The P-27 scheme uses daily 500 hPa heights on a regular grid of 36 grid points enclosed by the lines 40°N, 65°N, 20°W and 30°E (nearly whole Europe). The actual 500 hPa height h_{ti} for day t at grid point i is reduced first by subtracting the daily average height $\bar{h}_t$ over the grid. This operation removes a substantial part of the annual cycle. The reduced 500 hPa heights p_{ti} are given by:

$$p_{ti} = h_{ti} - \bar{h}_t, \qquad i = 1,\ldots,36; \;\; t = 1,\ldots,n \tag{21}$$

Subsequently, the vector $\mathbf{p}_t = (p_{t,1},\ldots,p_{t,36})^T$ of reduced 500 hPa heights is approximated as:

$$\mathbf{p}_t \approx s_{1t}\mathbf{a}_1 + s_{2t}\mathbf{a}_2 + s_{3t}\mathbf{a}_3, \qquad t = 1,\ldots,n \tag{22}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are the first three principal component vectors and s_{1t}, s_{2t} and s_{3t} their amplitudes or scores. These amplitudes are calculated each day from $\mathbf{p}_t$. The flow pattern of a particular day is thus described by three amplitudes: s_{1t}, s_{2t} and s_{3t} instead of 36 grid point values. It turns out that s_{1t} characterises the east-west component of the flow, s_{2t} the north-south component and s_{3t} the cyclonicity.

The first three principal components are used to develop the classification scheme in the following way. The range of each amplitude is divided into 3 equiprobable intervals. Then each pattern is, on the basis of its amplitudes, uniquely assigned to one of the 3x3x3=27 possible interval combinations.

As discussed above, Brandsma and Buishand (1995) compared the P-27 scheme with the Lamb objective scheme and the Grosswetterlagen with respect to their ability to predict daily temperature and occurrence of rain. A drawback of the P-27 scheme is that the 500 hPa heights are available only since 1949.

Other classifications

Bogárdi *et al.* (1993) modified the model of Bárdossy and Plate (1992) to simulate daily precipitation conditional on the daily circulation patterns in Nebraska. A more general transformation than the power transformation was introduced to achieve normality. Instead of the subjective Grosswetterlagen an objective classification scheme was devised based on a PCA of the 500 hPa heights followed by k-means clustering. The latter provides an optimal subdivision of the amplitudes of the principal component vectors into k different groups. Nine different circulation types were discriminated for both the summer and winter half year. Unlike the objective Lamb and P-27 schemes the classification rules for the winter half year differed from those for the summer half year. The coupling of PCA with k-means clustering has been compared with three other methods by Wilson *et al.* (1992). With respect to the prediction of daily precipitation, all methods performed about equally well.

Bárdossy (1994) developed objective schemes to classify the Grosswetterlagen from gridded 700 hPa heights. Both neural networks and fuzzy rules based clustering were considered. For daily rainfall estimation the best results were obtained with the fuzzy rules. Nevertheless, the performance of this objective scheme was worse than that of the subjective Grosswetterlagen. Further details about the classification of Grosswetterlagen by fuzzy rules can also be found in Bárdossy *et al.* (1995).

Hughes *et al.* (1993) and Zorita *et al.*(1995) used the Classification And Regression Trees (CART) procedure to discriminate circulation types. In contrast to the previous classification schemes the CART procedure makes use of precipitation data to obtain the optimal classification. In Hughes *et al.*(1993) and Zorita *et al.* (1995) it classifies the observed daily MSLP fields in weather states such that the separation between the patterns of wet at dry days at a set of index stations is maximal. Because the computational demand of the CART procedure increases rapidly with the number of points in the MSLP grid, the number of variables was reduced first using PCA. To account for lags between the large-scale pressure distribution and local precipitation both the amplitudes of the present and the previous day were considered in the CART procedure. Different classifications were obtained for winter, spring, summer and autumn, respectively. In both publications a nonparametric resampling technique was used to generate multisite daily rainfall data conditional on the circulation types. This work is discussed further in Section 4.

3.1.3 Discussion

In the 1990s there has been a rapid growth of papers dealing with the linkage of stochastic precipitation models to the atmospheric circulation. Part of this growth is due to the interest in downscaling the large-scale output from GCMs. The state of the atmospheric circulation is an important indicator of both the occurrence of rain and the rainfall amounts. Furthermore, the atmospheric circulation explains a part of the spatial and temporal correlation of the rainfall.

There are several methods to classify the atmospheric circulation, falling roughly into two major categories: manual classification and automated classification. Some classifications, like the Grosswetterlagen and the Lamb classification, have a large number of classes, resulting in too many parameters for stochastic rainfall models. This problem can be solved by grouping the classes with similar rainfall characteristics. For monthly rainfall, the atmospheric circulation explains a reasonable part of the variance but for daily rainfall the performance is much less. Some improvement is possible, however.

Most classification schemes are not designed to produce an optimal prediction of rainfall. Shortcomings can easily be identified. For instance, the P-27 scheme does not consider the average height of the 500 hPa surface and the objective Lamb scheme does not consider average air pressure at sea level over the grid. The dependence of the occurrence of rain on these quantities has been studied by Hughes and Guttorp (1994). These authors also show that the grid averages are weakly or moderately correlated with the gradients and second-order derivatives of the pressure field. On the other hand, it questionable whether a better classification scheme will lead to a better reproduction of properties of multi-day extremes. The development of new classifications schemes should be given low priority.

3.2 Long-term variation of the atmospheric circulation

Extreme years may not be randomly distributed in time, because of long-term variation in the atmospheric circulation. Such variation can often be interpreted as a systematic change in the mean of certain characteristics of the circulation or as a year-to-year correlation. Quite different techniques have therefore been followed to analyse and model long-term variation.

The presence of long-term variation complicates the simulation of daily precipitation amounts. It requires an extension to increase the interannual variation. Another consequence of long-term variation is that conclusions about the occurrence of extremes will be more uncertain than in the case of pure randomness between years.

This section provides an assessment of the behaviour of the atmospheric circulation in the past hundred years. Firstly we review the literature on changes of the atmospheric circulation in Europe. Secondly, the long-term variation of some basic components of the Lamb scheme and the P-27 scheme is studied for the three-month winter season (December-February), which is the most relevant season for extreme discharges at Lobith. The section concludes with a discussion.

3.2.1 Literature review

Caspary and Bárdossy (1995) observed a strong increase in the annual peak discharges of the Enz river in Germany since the end of the 1970s, which was related to an increase in the number of days with a west-cyclonic circulation during the three-months winter season since the beginning of the 1970s. As noted before in Section 3.1.1, Bárdossy and Caspary (1990) already concluded that an increase in the frequency of mild and humid winters resulted from an increase in the frequency of zonal types in the months of December and January since the year 1973.

Hurrel (1995) examined the North Atlantic Oscillation (NAO) for the winter period. The NAO is associated with changes in the surface westerlies across the North Atlantic onto Europe. Hurrel based his index of the NAO on the difference of normalised sea level pressures between Lisbon (Portugal) and Stykkishólmur (Iceland) from 1864 through 1994. Especially since the 1970s Hurrel found an upward trend in the NAO-index. He also found that the NAO-index was correlated with winter precipitation in Europe. For Scandinavia the correlation coefficients were mainly positive whereas for Southern Europe negative values predominate. For the middle of Germany (Frankfurt am Main) and Paris the correlation coefficients were rather weak ($r = -0.19$). It was concluded that the causes for the significant decadal changes in the atmospheric circulation, and the NAO in particular, are not yet clear.

Other studies focused on systematic changes in precipitation. Systematic changes in precipitation may be indicative of long-term variation of the atmospheric circulation. For the Belgian part of the Meuse catchment, Delft Hydraulics (1994) and Demarée *et al.* (1994) did not find systematic changes in the observed precipitation during the winter half year in the period 1910–1993. The European climate change atlas of Schönwiese *et al.* (1993) gives for the Netherlands an upward trend in the winter (December-February) precipitation for the period 1891–1990. Vaessen and Visser (1991) studied the long-term variation of the precipitation at Hoofddorp for the period 1735–1989. For the years 1980–1989 they found 7 exceedances of the 87 percentage point of the three-months winter period precipitation. The probability of this event was estimated as 0.000053. For the precipitation at De Bilt a remarkable jump has been observed in the month of March (Zwart, 1993). For the period 1931–1960 the average precipitation in spring amounted to 144 mm, while for the period 1961–1990 this was 176 mm, a difference of 32 mm. The month of March was responsible for two thirds of this difference. For the three-months winter season the change in precipitation at De Bilt is less apparent. These findings for De Bilt are in line with a comparison between the seasonal averages for the periods 1981–1990 and 1951–1980 by the European Climate Support Network (ECSN, 1995). This comparison shows relatively high values for the spring amounts in the 1981–1990 decade for the Netherlands and the surrounding countries. The changes in winter precipitation are less apparent in this region.

3.2.2 Behaviour of the objective Lamb and P-27 classification schemes

Objective Lamb classification scheme

It is of interest to study the behaviour of the objective Lamb circulation during the past century and see whether or not the results agree with those found in the literature. For this purpose we centred the MSLP grid as close as possible over the Rhine catchment at 50°N-5°E. We did not study the behaviour of all weather types separately. Instead, we examined the behaviour of: (1) the average vorticity; (2) the average flow strength; (3) the frequency of the days belonging to one of the eight flow directions; (4) the average vorticity on days belonging to one of the eight flow directions; and (5) the average strength of the flow on days belonging to one of the eight flow directions.

A selection of the results is presented in Figures 5 to 9. All figures apply to the period 1881–1993 (more precisely, the winters of 1880/81 to 1992/93). Two smooth curves are added to the figures to filter out the high frequency variations: (1) a curve based on locally weighted regression smoothing with a span of 0.3 (30% of the data are used for each point); and (2) a curve based on the so-called supersmoother using local cross validation to determine the span. More information on the two smoothers can be found in Härdle (1990) and Hastie and Tibshirani (1990), respectively. Each series has been tested for stationarity using the maximum of the rescaled adjusted partial sums (Buishand, 1982).

The two smooth curves in Figure 5 show that the average vorticity in winter decreased since about 1975. The null hypothesis of a stationary record is rejected at the 10% level. A consequence of a decrease in average winter vorticity would be a decrease in precipitation. Figures 6 and 7 show the frequency of days with southerly and westerly flow directions, respectively. Although both series passed the test on the rescaled adjusted partial sums, it appears that there is a striking decrease in the frequency of days with southerly flow directions since about 1973 and a small increase in the frequency of days with westerly flow directions. Figure 8 shows the average vorticity on days with a westerly flow direction. This figure shows the same features as the average vorticity in Figure 5. Unlike the series in Figure 5, the series in Figure 8 is considered stationary. In the latter case, a larger change is needed to obtain a statistically significant result because of the smaller sample size. It appeared that the average vorticity for the southwesterly and northwesterly flow directions possessed the same features as for the westerly flow direction. Finally, Figure 9 shows the average strength of the flow on days with a westerly flow direction. Although the series in this figure must again be considered stationary, there is striking increase in the average strength of the flow since about 1975.

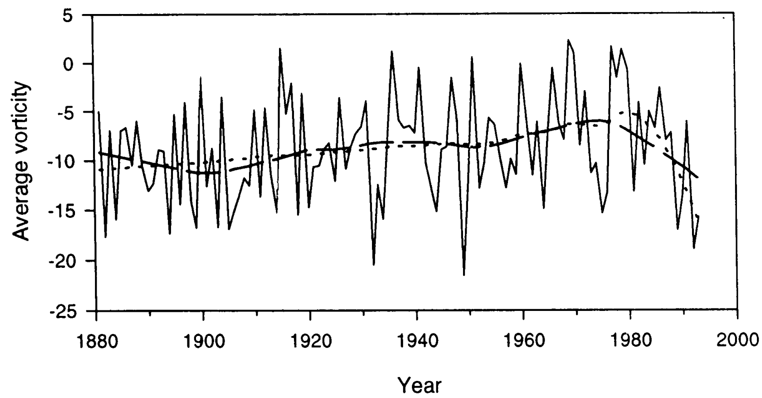


Figure 5: Average vorticity in winter (December-February) for the period 1881–1993 as calculated from the MSLP grid centred at 50°N-5°E. The long-dashed curve is a locally weighted regression smooth with a span of 0.3; the short-dashed curve is based on the supersmoother. The values along the vertical axis refer to the geostrophic vorticity expressed as hPa per 10° latitude at 50°N (10 units are equivalent to 0.65×10^{-5} s^{-1}).

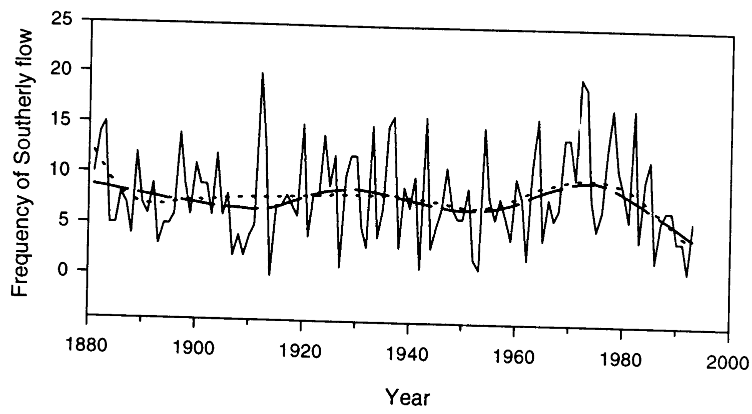


Figure 6: Frequency of days with southerly flow direction in winter (December-February) for the period 1881–1993 as calculated from the MSLP grid centred at 50°N-5°E. The long-dashed curve is a locally weighted regression smooth with a span of 0.3; the short-dashed curve is based on the supersmoother.

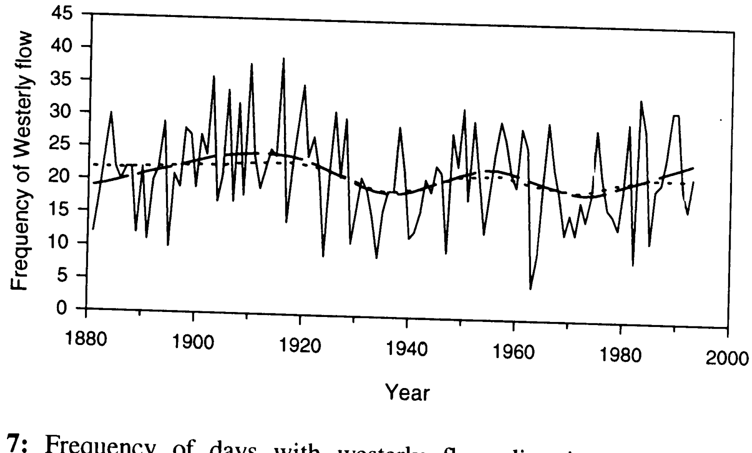


Figure 7: Frequency of days with westerly flow direction in winter (December-February) for the period 1881–1993 as calculated from the MSLP grid centred at 50°N-5°E. The long-dashed curve is a locally weighted regression smooth with a span of 0.3; the short-dashed curve is based on the supersmoother.

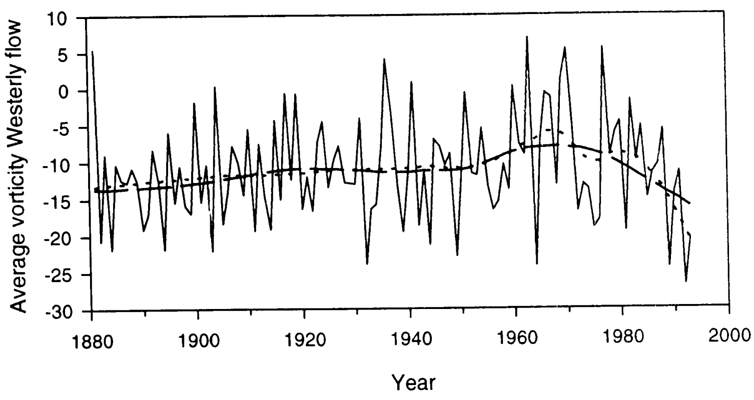


Figure 8: Average vorticity on days with a westerly flow direction in winter (December-February) for the period 1881–1993 as calculated from the MSLP grid centred at 50°N-5°E. The long-dashed curve is a locally weighted regression smooth with a span of 0.3; the short-dashed curve is based on the supersmoother. The values along the vertical axis refer to the geostrophic vorticity expressed as hPa per 10° latitude at 50°N (10 units are equivalent to $0.65 \times 10^{-5}\ s^{-1}$).

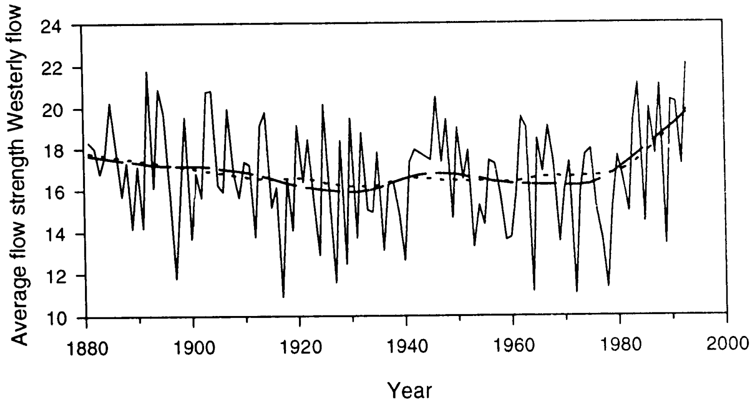


Figure 9: Average flow strength on days with a westerly flow direction in winter (December-February) for the period 1881–1993 as calculated from the MSLP grid centred at 50°N-5°E. The long-dashed curve is a locally weighted regression smooth with a span of 0.3; the short-dashed curve is based on the supersmoother. The flow units are geostrophic, expressed as hPa per 10° latitude at 50°N (1 unit is equivalent to 0.73 m/s).

Precipitation in the flat regions of the Rhine catchment is strongly associated with vorticity. For the hilly and mountainous regions the strength of the westerly flow becomes more important. Especially for the latter regions an increase in the number of humid winters as reported by Caspary and Bárdossy (1995) might be anticipated during the past twenty years.

P-27 classification scheme

The P-27 scheme was introduced in Section 3.1.2. In this section we study the time series of the three scores s_{1t}, s_{2t} and s_{3t}. As noted before, s_{1t} characterises the east-west component of the flow, s_{2t} the north-south component and s_{3t} the cyclonicity. We followed the same approach to describe the long-term variation as for the objective Lamb scheme.

The results for the period 1951–1993 are presented in Figure 10. It appears that s_{1t} shows a downward trend beginning roughly in 1970 while s_{3t} shows a downward trend during the whole period 1951–1993. The results for s_{2t} do not show special features. For both series, s_{1t} and s_{3t}, the null hypothesis of a stationary record is rejected at the 10% level. The downward trend in s_{1t} suggets an increase in the number of days with westerly circulation and/or an increase of the strength of the flow on days with westerly circulation. This roughly corresponds to the results for the Lamb scheme and the results from the literature. The results for s_{3t} indicate that there is a decrease in vorticity since 1970, which corresponds to the results for the Lamb scheme.

3.2.3 Discussion

The literature study and the results for the Lamb and P-27 classification schemes in this section show that especially in the last 20 years there have been some changes in the atmospheric circulation during the winter season. The statistical significance of these changes is, however, not very strong. For the two objective schemes considered here, the largest changes are just significant at the 10% level. Furthermore, it is difficult to say whether or not the observed changes are real trends that will continue in the future or that they belong to natural long-term fluctuations.

Also the effect of the changes in the atmospheric circulation on precipitation is not yet clear. For instance, for Germany an increase in the number of humid winters has been reported but for the Belgian part of the Meuse basin and large parts of the Netherlands this is not so obvious. Therefore, the geographical variation of changes in the mean winter precipitation needs further attention. Besides some indications

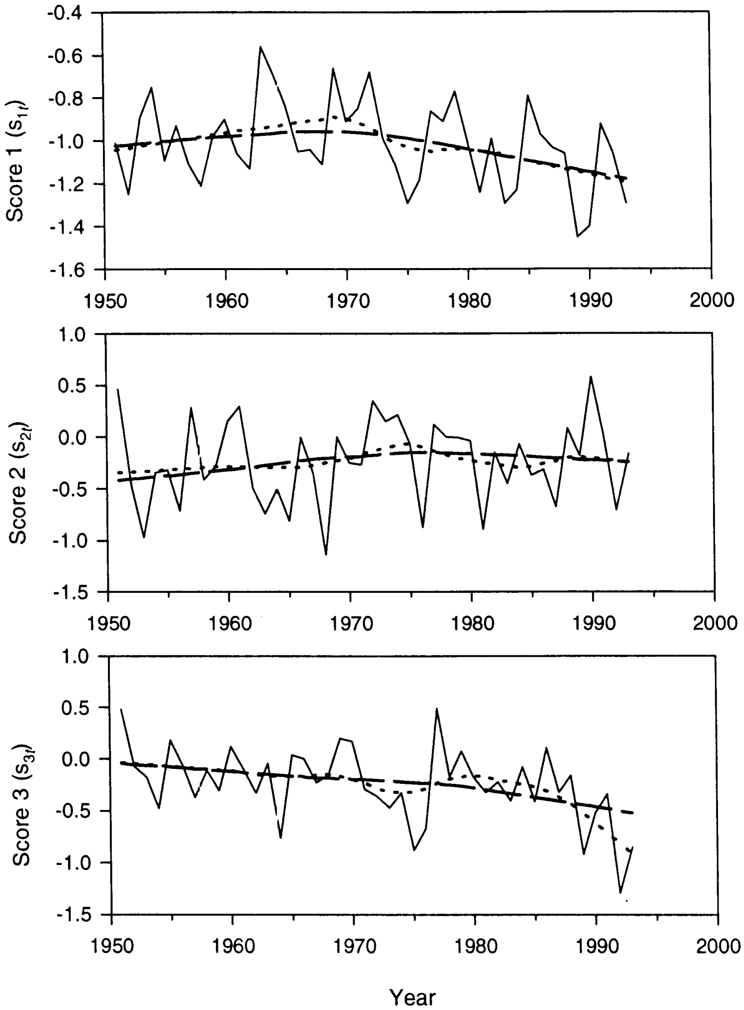


Figure 10: Average values of the scores s_{1t}, s_{2t} and s_{3t} for the P-27 scheme in winter (December-February) for the period 1951–1993. The long dashed curves are locally weighted regression smooths with a span of 0.79; the short dashed curves are based on the supersmoother.

of changes in winter precipitation, there is a somewhat stronger evidence of long-term variation during spring, in particular for the month of March.

From the above changes one may expect that fitting a stochastic precipitation model to the last 20 years of data will give somewhat different results than a model based on a longer period of observation. As long as there is no clear evidence of a persisting trend one should not limit the analysis to the last 20 years only. Modelling should be based on longer records and the need to incorporate extra long-term variation should be considered.

3.3 Simulation of weather types and circulation indices

Different approaches can be followed to simulate characteristics of the atmospheric circulation. In this section, we will distinguish between the simulation of weather types (categorical data) and the simulation of indices or amplitudes (continuous variables).

3.3.1 Simulation of weather types

One problem in the simulation of a time series of weather types is to preserve the autocorrelation of the historic record. The reproduction of the annual course is another problem.

Like the occurrence of wet and dry days, the succession of weather types can be described by Markov chains. The first-order Markov chain is defined in terms of one-step transition probabilities. These are the conditional probabilities that weather type j occurs on day t, given that the weather type on day $t-1$ was i, where i can be equal to j. For an atmospheric classification with 10 weather types, the number of transition probabilities runs up to 100. Bogárdi *et al.* (1993) and Wilby *et al.* (1994) used this Markov chain model to simulate sequences of weather types. In general a first-order Markov chain is not sufficient to describe the memory of the system. Two approaches to this problem are: (1) the use of high-order Markov chains; and (2) the use of semi-Markov processes.

Hughes *et al.* (1993) applied the first approach to extend a time series consisting of 5-years GCM generated weather state sequences. They used the lth order Raftery model that is defined by:

$$P(S_t|S_{t-1},\ldots,S_{t-l}) = \sum_{k=1}^{l} \lambda_k q_{S_{t-k},S_t} \tag{23}$$

where S_t is the weather state at time t. The λ_k's are a set of weights which must sum to 1 but are otherwise unrestricted. The q_{ij}'s are non-negative, such that:

$$\sum_{j=1}^{m} q_{ij} = 1, \qquad i = 1,\ldots,m \tag{24}$$

and

$$0 \le \sum_{k=1}^{l} \lambda_k q_{i_k j} \le 1, \qquad i_1,\ldots,i_l, j = 1,\ldots,m \tag{25}$$

with m the number of weather states. The first-order Raftery model is equivalent to the first-order Markov chain model with one-step transition probabilities q_{ij}. Through the particular form of $P(S_t | S_{t-1},\ldots,S_{t-k})$, a large reduction in the number of parameters is achieved when the order of the chain is greater than one.

The second approach has been used by e.g. Wilson *et al.* (1991) and Bárdossy and Plate (1991, 1992). Wilson *et al.* used the semi-Markov processes to describe both the succession of weather types and the precipitation occurrence while Bárdossy and Plate used the semi-Markov process to simulate a sequence of weather types. The semi-Markov process not only specifies the transition probabilities between the various states, but also the distribution of the duration that the process stays in a given state. This distribution can depend on the current state as well as the next. The semi-Markov process is thus a Markov chain with a randomly transformed time scale.

A time series of weather types usually shows a systematic annual cycle. The parameters of the models are therefore often estimated for each season or each month separately. This leads to a strong increase in the number of parameters and thus quite long records of weather types are needed to obtain accurate estimates of the various transition probabilities.

3.3.2 Simulation of circulation indices or PC amplitudes

If our point of departure is an objective classification scheme, the simulation of indices or amplitudes is a reasonable alternative to the simulation of weather types. In contrast to the latter, the simulation of indices or amplitudes deals with continuous variables. The number of variables is much less (usually three) than the number of weather types. Furthermore, the number of parameters that must be estimated from the data is much less than for the simulation of weather types with Markov chain models.

Provided that the indices are normally distributed, the sequences of indices or amplitudes can be generated by the multivariate AR(1) model of Section 2.2. For the P-27 scheme the lag-zero cross-correlation coefficients between the scores s_{1t}, s_{2t} and s_{3t} are small, but there might be significant lagged cross-correlations. It is then not allowed to take **A** to be diagonal as in contemporaneous models.

Kruizinga (1978) showed that the distributions of the scores s_{1t}, s_{2t} and s_{3t} used in the P-27 classification are close to the normal distribution. Transformations can be considered to make use of the multivariate AR(1) model in case of non-normality. There are also possibilities to build in long-term persistence.

4. Nonparametric approaches

Resampling from historical multisite data offers an alternative to generate daily precipitation amounts over a region. In contrast to the parametric models in Section 2, the nonparametric approach ensures that the spatial relationships between the 1-day amounts are preserved. The method was suggested recently in studies concerning the generation of realistic daily precipitation sequences associated with an altered atmospheric circulation (Hughes *et al.*, 1993; Zorita *et al.*, 1995). The linkage between precipitation and the atmospheric circulation is also useful for nonparametric generation of daily rainfall under present-day climate conditions because of the effect of the large-scale circulation on the autocorrelation of the daily precipitation amounts.

There are in fact two different methods to take the atmospheric circulation into account, namely (1) resampling from the historical observations conditional on the weather type; and (2) selection of analogues. The first method requires a classification of the circulation. Resampling is restricted to days with the same weather type as the day of interest. The analogue method finds the most similar day in the historical record with respect to the features of the atmospheric circulation and then sets the simulated precipitation amounts equal to the observed amounts on that day. Tracing the analogues requires a similarity measure of circulation patterns. In Zorita *et al.* (1995) the Euclidean distance between the amplitudes of the leading five principal components (PCs) of the MSLP field was used as a similarity measure.

Although the atmospheric circulation explains part of the temporal correlation this is not sufficient to preserve the complete autocorrelation structure of the daily precipitation amounts. Zorita *et al.* (1995) show that various resampling procedures fail to reproduce the distribution of the run-lengths of dry days. For the analogue method an improvement was achieved by comparing the MSLP field for the present day and the previous four days with all historical five-day segments. The generated precipitation amounts were set equal to the observed amounts on the fifth day of the most similar five-day segment in terms of PC amplitudes. From this improvement the authors recommend to consider the time-evolution of the MSLP field rather than the flow characteristics of a single day.

Another possibility to improve the reproduction of the autocorrelation structure is to include the precipitation occurrence or the precipitation amounts on the previous day in the resampling procedure. In Hughes *et al.*(1993) the precipitation amounts were resampled conditional on the weather type and the previous day's rain state (wet or dry). For the River Rhine it might be more convenient to discriminate between days below the median and those above the median amount, because the probability of no rain over the entire catchment is much smaller than 0.5. A further improvement in the representation of the temporal dependence is possible by generating the rain state with a separate parametric zero-one process or by generating the daily catchment average rainfall amount by a separate parametric model, e.g. the transformed normal AR(1) process in Section 2. In the latter case, the analogue method may provide an objective disaggregation of the simulated catchment average rainfall over the various subcatchments.

Temporal dependence controls the variances of the multi-day amounts (Section 2.3.1). A resampling procedure that fails to reproduce the temporal dependence will not be capable to describe the distribution of these amounts. Besides improving the reproduction of temporal dependence one could also think about resampling several days at-a-time. For instance, in case of a run of length D of a certain weather type one might resample from the runs in the historical observations of the same length. However, additional rules are then needed for long runs to avoid resampling from a limited number of runs.

Nonparametric resampling of multi-day amounts is a new technique that deserves further attention. The literature on this technique has put strong emphasis on the preservation of the distribution of the run-lengths of dry days. For our application it is, however, more relevant to study the upper tail of the distribution of the multi-day amounts. Although the resampling procedure reproduces the spatial dependence of the 1-day amounts this may not be true for the multi-day amounts.

5. Inclusion of daily temperatures

The methods in Sections 2 and 4 only provide daily precipitation. Snow often contributes to high river discharges at Lobith in the form of snowmelt runoff and by creating unfavourable antecedent conditions. An extension that allows for snow accumulation and snowmelt seems therefore desirable. Because these processes can be linked to the air temperature, we discuss in this section extensions of our stochastic precipitation models to generate the daily temperatures as well.

The daily temperature exhibits a much stronger autocorrelation than daily precipitation. Nonparametric resampling meets therefore serious difficulties. Although the distribution of daily temperatures is much closer to the normal distribution than that of daily precipitation, a transformation might be necessary for generating daily temperatures by parametric autoregressive models. The power transformation in Section 2.1, however, no longer applies because temperatures can be negative. The transformation can be based on the binormal distribution (Matyasovszky *et al.*, 1994).

The binormal distribution is a three-parameter distribution due to Toth and Szentimrey (1990). Both the left and right tail are represented by a normal distribution. The standard deviations of these two distributions are, however, unequal. The density of the binormal variable reads:

$$f(x) = \begin{cases} \dfrac{1}{\bar{\sigma}\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2}\dfrac{(x-M)^2}{\sigma_1^2}\right\}, & x \le M \\ \dfrac{1}{\bar{\sigma}\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2}\dfrac{(x-M)^2}{\sigma_2^2}\right\}, & x > M \end{cases} \tag{26}$$

where

$$\bar{\sigma} = (\sigma_1 + \sigma_2)/2 \tag{27}$$

and M is the mode of the distribution (most probable value). The distribution function $F(x)$ is given by:

$$F(x) = \begin{cases} \dfrac{\sigma_1}{\bar{\sigma}}\Phi\left(\dfrac{x-M}{\sigma_1}\right), & x \le M \\ \dfrac{\sigma_1 - \sigma_2}{2\bar{\sigma}} + \dfrac{\sigma_2}{\bar{\sigma}}\Phi\left(\dfrac{x-M}{\sigma_2}\right), & x > M \end{cases} \tag{28}$$

where Φ is the standard normal distribution function. The unknown parameters M, σ_1 and σ_2 can be estimated by the maximum likelihood method. Using the fitted distribution, the sequence $\{x_t\}$ of daily temperatures can be transformed as:

$$z_t = \Phi^{-1}\{F(x_t)\} \tag{29}$$

For each x_t the probability $F(x_t)$ is computed first and then the corresponding quantile in the standard normal distribution is determined via the inverse of Φ. This transformation is quite general. It can be used to transform any distribution to a standard normal distribution.

The multivariate AR(1) process in Section 2.2 can be extended to a coupled daily rainfall-temperature model. Part of the vector $\mathbf{z}_t$ in Equation (11) then refers to the transformed daily temperatures at time t and the other part to the precipitation amounts. Such a model generates daily precipitation and temperature simultaneously, taking their mutual correlation into account. A problem with this approach is the large number of correlation coefficients. Besides the correlation between the daily precipitation and the temperature at the same subcatchment there is also a correlation between the daily precipitation amount over a subcatchment and the temperature at surrounding subcatchments.

Another approach is to generate daily precipitation first and then the daily temperatures conditional on precipitation occurrence. Richardson (1981) followed this approach to supplement generated point precipitation data with other weather variables. His method can also be applied when multisite temperature data are required. The conditioning on precipitation occurrence implies that the parameters of the temperature distribution (i.e. M, σ_1 and σ_2 in case of the binormal distribution) depend on the rain state. These parameters must further be conditioned on the atmospheric circulation when circulation characteristics are used to generate the precipitation amounts. The standardised (transformed) temperatures can be generated by the multivariate AR(1) process.

The conditional simulation of temperature on precipitation occurrence does not require the correlation coefficients between daily precipitation and temperature. The method can also be used when the daily precipitation data have been obtained by a resampling technique. It does not necessarily preserve the autocorrelation structure of the daily temperatures. Katz (1995) even notes that Richardson's model may not necessarily produce the proper lag-one autocorrelation coefficient of the daily temperatures. For daily January temperatures at Denver, Colorado, over a 20-year period, he obtained a theoretical value of 0.630 from the fitted model, being much lower than the value of 0.742 from the data.

For the design discharges at Lobith the relatively simple conditional simulation of temperature on precipitation occurrence might be attractive because the effect of precipitation dominates. The total seasonal amount at temperatures below zero (or a small positive threshold) is an important characteristic to verify the model's capability to reproduce snow accumulation.

6. Climate change applications

This study is focused on statistical techniques to generate multisite daily rainfall for the Rhine catchment for present climate conditions. However, conditions might change in the future because of, for instance, greenhouse gas induced global warming. Therefore, also some attention is given to climate change applications.

Impact assessment of a potential future global warming on river discharges is a new application of statistical time series simulation. The effect of increasing greenhouse gases in the atmosphere is not restricted to temperature alone. Higher temperatures have an influence on precipitation because the capacity of the atmosphere to hold water vapour increases with increasing temperature. The changes in temperature may also be accompanied by systematic changes in the atmospheric circulation. Higher temperatures further affect the discharges of the River Rhine because of their impact on snowfall and evapotranspiration.

Statistical time series generation makes it possible to obtain realistic climate time series or scenarios for the situation of a global warming. Two different approaches have been suggested to include the large-scale information from the GCMs: (1) scenario production by adjusting the parameters in the time series model to the GCM predicted changes in the precipitation and other variables (Burlando and Rosso, 1991; Wilks, 1992); and (2) conditional simulation on the large-scale atmospheric circulation patterns produced by the GCM (Bárdossy and Plate, 1992; Zorita *et al.*, 1995). The latter approach is often referred to as statistical downscaling.

A weak point in the first approach is the use of GCM predicted changes in precipitation. As an alternative the changes in precipitation may be obtained from statistical relations with flow characteristics and temperature. It is important that the selected relations are invariant under climate change. Matyasovszky *et al.* (1993) adjusted the parameters of the daily precipitation distribution, using the changes in the average height of the 500 hPa level. When there is no change in the large-scale circulation representative precipitation records for a warmer climate can be obtained by multiplying the observed amounts by a temperature dependent factor based on the regression of the precipitation amounts on temperature and MSLP or other relevant flow characteristics (Buishand and Klein Tank, 1996). The number of rain days remains unchanged in that method.

Statistical downscaling assumes that the GCM realistically reproduces the large-scale circulation. The length of the generated time series is the same as that of the GCM run (often not more than 10 years). Most publications only consider the link between precipitation and the atmospheric circulation. The effect of the increasing temperature on precipitation is thus ignored. The link between precipitation and temperature requires therefore further attention in this method. The generation of realistic multisite daily temperature data differs from that in Section 5, because the GCM already provides rather realistic temperatures over a grid of about 250×250 km^2. Further, conditional simulation of temperature on precipitation occurrence is rather unnatural when a scenario for a global warming is required. Finally, statistical downscaling has suffered from the fact that systematic changes in the large-scale circulation in GCM simulations with increased atmospheric greenhouse gas concentrations are smaller than the differences between the GCM control run and the observed climate.

All time series generation techniques in Sections 2 and 4 have prospects for climate change applications. However, for the situation of a global warming extensions are needed to account for the temperature effect on precipitation. This holds in particular for nonparametric resampling techniques and statistical downscaling. As an alternative to statistical downscaling, the effects of the systematic changes in the temperature and atmospheric circulation can be incorporated by changing the parameters in the statistical time series model. The link to the daily temperature and pressure fields of the driving GCM is lost in that approach.

7. Conclusions and recommendations

Several techniques have been proposed to generate rainfall sequences. Only a few of these can be used to obtain simultaneous daily values at several sites in a region or over several subcatchments of a large river basin. Promising results have been reported about the performance of a transformed multivariate AR(1) process. Nonparametric resampling from historical data is another possibility to generate multivariate daily values. Although these two methods do not give new information about the distribution of the 1-day amounts, they provide different temporal patterns of extreme events and other spatial distributions of large multi-day rainfall over the catchment than those that have been observed in the past. The use of such synthetic data in combination with a hydrological/hydraulic model, does not only provide the peak discharges, but also the duration of high river discharges. However, the generated precipitation data can only have an additional value for the determination of design discharges when the model is capable to reproduce the statistical properties of extreme events derived from the historical data. For the Rhine discharges at Lobith the extreme-value distribution of multi-day amounts and the spatial association of large amounts during winter are important. There is a serious lack of insight into the quality of the two earlier mentioned rainfall generators with respect to these properties.

The atmospheric circulation partly controls the probability of rain, the amount of rain on wet days, the persistence of daily rainfall and the spatial patterns of the amounts. Nonparametric resampling techniques always take the atmospheric circulation into account. There is also a growing interest to link parametric rainfall models to the atmospheric circulation. Several classifications of the atmospheric circulation have been developed. These explain a reasonable part of the variances of the monthly amounts. Time series modelling of flow indices or PC amplitudes is an attractive alternative to the generation of sequences of weather types by means of transition probabilities (Markov chains, semi Markov processes) because of parameter parsimony and the possibility of incorporating long-term variations of circulation characteristics.

Transformed multivariate autoregressive processes are also suitable for modelling multisite temperatures. Conditional simulation of temperature on precipitation occurrence is a simplification that requires much less parameters than a coupled daily rainfall-temperature model. It can also be used when the daily rainfall data have been generated by a nonparametric resampling technique. The generation of precipitation sequences in the situation of an altered climate requires some modifications. In particular the link between precipitation and temperature has to be examined further when scenarios for impact assessment of a potential global warming are required.

The construction of a rainfall generator for the Rhine catchment is a unique exercise, which may meet several difficulties. Further development of nonparametric resampling techniques and model evaluation with respect to the properties of multi-day extremes need first priority. The following five phases can be discriminated.

(A) Further evaluation of the multivariate AR(1) model

The literature refers to multisite point rainfall instead of multivariate daily average amounts over various subcatchments. Coupling with an existing classification of circulation types should be considered. Parameter estimation needs some attention. It is expected that a sound estimation procedure combined with a proper choice of the transformation parameter λ will result in a reasonable reproduction of the distribution of extreme 1-day amounts and the decay of the autocorrelation function. The distribution of extreme multi-day amounts and the association of large amounts remains a crucial point. The evaluation can be restricted to the German part of the catchment ($\approx$ 100 000 km^2). This subproject may not take more than 15 months. It is assumed that the rainfall data for Germany are available at the beginning of the project. The final result of this phase is a statistical model for generating daily rainfall over the German subcatchments of the River Rhine, conditional on a historical series of weather types.

(B) Further development of resampling techniques

The literature learns that resampling techniques have difficulties with reproducing the autocorrelation structure of the daily amounts. Several extensions are possible to achieve a better reproduction of temporal dependence. Further development of resampling techniques can also be restricted to the German part of the catchment first. As for the parametric approach in Phase (A), verification should consider the extreme-value distribution and spatial dependence of multi-day amounts. This subproject will take about 18 months, where again it is assumed that the rainfall data for Germany are available at the beginning of the project. The outcome of this phase is an alternative technique for generating daily rainfall over the German subcatchments of the River Rhine, conditional on observed circulation characteristics, that might reproduce the statistical properties of multi-day rainfall better than the model under Phase (A).

(C) Statistical simulation of characteristics of the atmospheric circulation

Statistical simulation of time series of flow indices or PC amplitudes is a new development. Weather types can be derived from the components of these time series. A weather type classification is, however, not needed when the daily precipitation amounts are generated by resampling from analogues. A normalising transformation of flow indices may be necessary. The daily average MSLP or 500 hPa heights could be included in the search for analogues. This subproject will take about 9 months. Its final result is a statistical technique for generating characteristics of the atmospheric circulation, making it possible to obtain synthetic rainfall sequences over longer time-horizons than the observed records.

(D) Further development of the multivariate AR(1) model

Possibilities to reduce the number of parameters should be investigated. Especially the structure of the lag-zero cross-correlation matrix is of interest. The entire upstream area of Lobith should now be considered. Conditional simulation of temperature on precipitation occurrence should be examined. At this stage it is desirable to test the sensitivity of generated extreme river discharges to the model parametrization. This subproject will take about 12 months. The development of a coupled daily rainfall-temperature model will require another 4 months. This subproject leads to a refined and extended version of the model obtained under Phase (A), that can provide simultaneous records of daily temperature and precipitation over the entire upstream area of Lobith.

(E) Statistical simulation of multisite rainfall by nonparametric resampling

The most promising technique from Phase (B) should be extended to the entire upstream area of Lobith. Minor modification may be necessary. This subproject will take about 3 months and will result in an alternative technique to generate multisite daily rainfall over the Rhine catchment. Concurrent temperatures can be generated with the conditional multivariate AR(1) model developed under Phase (D).

Phases (A) and (B) should be tackled first. After these phases, the meteorological data for subcatchments outside Germany should be available. The need to study long-term variation should be considered by the end of Phase (C). It is desirable that a hydrological/hydraulic model is already available at the beginning of Phases (D) and (E). Statistical time series simulation for climate change impact assessment is seen as a separate project.

The estimated durations hold for a competent junior scientist. Supervision will require 8 to 12 months for the whole project.

Acknowledgements

We thank B.W.A.H. Parmet (RIZA) and H.R.A. Wessels (KNMI) for comments on earlier drafts of the report. The UK Meteorological Office gridded MSLP data were kindly provided by P.D. Jones (Climatic Research Unit, University of East Anglia, Norwich).

References

Bárdossy, A., *Stochastische Modelle zur Beschreibung der raum-zeitlichen Variabilität des Niederschlages*, Mitteilungen 44, Institut für Hydrologie und Wasserwirtschaft Universität Karlsruhe, Karlsruhe, Germany, 153 pp., 1993 (in German).

Bárdossy, A., *Modelle zur Abschätzung der regionalen hydrologischen Folgen einer Klimaänderung*, Mitteilungen 47, Institut für Hydrologie und Wasserwirtschaft Universität Karlsruhe, Karlsruhe, Germany, 90 pp., 1994 (in German).

Bárdossy, A., and H.J. Caspary, Detection of climate change in Europe by analyzing European circulation patterns from 1881 to 1989, *Theor. Appl. Climatol.*, **42**, 155–167, 1990.

Bárdossy, A., and E.J. Plate, Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence, *J. Hydrol.*, **122**, 33–47, 1991.

Bárdossy, A., and E.J. Plate, Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, **28**, 1247–1259, 1992.

Bárdossy, A., L. Duckstein, and I. Bogárdi, Fuzzy rule-based classification of atmospheric circulation patterns, *Int. J. Climatol.*, **15**, 1087–1097, 1995.

Bhuiya, R.K., and V. Yevjevich, *Effects of truncation on dependence in hydrological time series*, Hydrology paper No. 31, Colorado State University Fort Collins, Colorado, 1968.

Bogárdi, I., I. Matyasovszky, A. Bárdossy, and L. Duckstein, Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns, *J. Geophys. Res.*, **98**(D9), 16,653–16,667, 1993.

Brandsma, T., and T.A. Buishand, *Analysis and modelling of the Durham (UK) precipitation data and comparison of circulation classification schemes*, Memorandum KO-95-04, KNMI, De Bilt, the Netherlands, 10 pp., 1995 (unpublished document).

Bras, R.L., and I. Rodríguez-Iturbe, *Random Functions and Hydrology*, Addison-Wesley, Reading Mass., 559 pp., 1995.

Briffa, K.R., P.D. Jones, and P.M. Kelly, Principal component analysis of the Lamb catalogue of daily weather types: part 2, seasonal frequencies and update to 1987, *Int. J. Climatol.*, **10**, 549–563, 1990.

Buishand, T.A., *Stochastic modelling of daily rainfall sequences*, Meded. Landbouwhogeschool Wageningen 77-3, Wageningen, the Netherlands, 212 pp., 1977a.

Buishand, T.A., *De variantie van de gebiedsneerslag als functie van puntneerslagen en hun onderlinge samenhang*, Meded. Landbouwhogeschool Wageningen 77-10, Wageningen, the Netherlands, 12 pp., 1977b (in Dutch).

Buishand, T.A., Some methods for testing the homogeneity of rainfall records, *J. Hydrol.*, **58**, 11–27, 1982.

Buishand, T.A., Bivariate extreme-value data and the station-year method, *J. Hydrol.*, **69**, 77–95, 1984.

Buishand, T.A., and T. Brandsma, Further comparison between the objective Lamb scheme and the Grosswetterlagen, Memorandum KO-95-05, KNMI, De Bilt, the Netherlands, 18 pp., 1995 (unpublished document).

Buishand, T.A., and A.M.G. Klein Tank, Regression model for generating time series of daily precipitation amounts for climate change impact studies, *Stochastic Hydrology and Hydraulics* (Accepted), 1996.

Burlando, P., and R. Rosso, Extreme storm rainfall and climatic change, *Atmos. Res.*, **27**, 169–189, 1991.

Caspary, H.J., and A. Bárdossy, Markieren die Winterhochwasser 1990 und 1993 das Ende der Stationarität in der Hochwasserhydrologie infolge von Klimaänderungen?, *Wasser & Boden*, **47**, 18–24, 1995 (in German).

Chandler, R., V. Isham, A. Kakou, and P. Northrop. Spatial-temporal rainfall prosesses: stochastic models and data analysis, In: *Sixth international meeting on statistical climatology, June 19–23, 1995, Galway, Ireland*, 431–434, 1995.

Cowpertwait, P.S.P., Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resour. Res.*, **27**, 1431–1438, 1991.

Cowpertwait, P.S.P., A generalized point process model for rainfall, *Proc. R. Soc. Lond. A*, **447**, 23–37, 1994.

Cowpertwait, P.S.P., A generalized spatial-temporal model of rainfall based on a clustered point process, *Proc. R. Soc. Lond. A* (Accepted), 1995.

Cox, D.R., and V. Isham, Stochastic models of precipitation, In: *Statistics for the Environment 2: Water Related Issues*, edited by V. Barnett and K.F. Turkman, 3–18, Wiley, Chichester, 1994.

David, H.A., *Order Statistics, 2nd edition*, Wiley, New York, 360 pp., 1981.

Delft Hydraulics, Onderzoek watersnood Maas, deelrapport 4: Hydrologische aspecten, Delft Hydraulics, Emmeloord, the Netherlands, 1994 (in Dutch).

Delft Hydraulics, and EAC-RAND, *Toetsing uitgangspunten rivierdijkversterkingen, Deelrapport 2: Maatgevende belastingen*, Delft Hydraulics, Emmeloord, and European American Center for Policy Analysis (EAC-RAND), Delft, the Netherlands 1993 (in Dutch).

Demarée, G.R., S. Derasse, and D. Gellens, Hoogwaterstanden en wateroverlast van de Belgische Maas te Visé, KMI, Brussel, Belgium, 55 pp., 1994 (in Dutch).

ECSN, *Climate of Europe: recent variation, present state and future prospects*, European Climate Support Network, KNMI, De Bilt, the Netherlands, 72 pp., 1995.

Gaver, D.P., and P.A.W. Lewis, First-order autoregressive gamma sequences and point processes, *Adv. Applied Probability*, **12**, 727–745, 1980.

Gerstengarbe, F.W., and P.C. Werner, *Katalog der Grosswetterlagen Europas nach Paul Hess und Helmuth Brezowski 1881-1992*, Berichte des Deutschen Wetterdienstes nr. 113, Offenbach am Main, Germany, 249 pp., 1993.

Gregory, J.M., T.M.L. Wigley, and P.D. Jones, Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals, *Clim. Dyn.*, **8**, 299–310, 1993.

Haltiner, J.P., and J.D. Salas, Development and testing of a multivariate, seasonal ARMA(1,1) model, *J. Hydrol.*, **104**, 247–272, 1988.

Härdle, W., *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 333 pp., 1990.

Hastie, T.J., and R.J. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, 335 pp., 1990.

Hess, P., and H. Brezowsky, *Katalog der Grosswetterlagen Europas*, Ber. Dtsch. Wetterdienstes 113, vol. 15, 2nd ed., Deutsche Wetterdienst, Offenbach am Main, Germany, 1969 (in German).

Hipel, K.W., and A.I. Mc Leod, *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, Amsterdam, 1013 pp., 1994.

Hughes, J.P., and P. Guttorp, Incorporating spatial dependence and atmospheric data in a model of precipitation, *J. Appl. Meteorol.*, **33**, 1503–1515, 1994.

Hughes, J.P., D.P. Lettenmaier, and P. Guttorp, A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation, *Water Resour. Res.*, **29**, 3303–3315, 1993.

Hulme, M., K.R. Briffa, P.D. Jones, and C.A. Senior, Validation of GCM control simulations using indices of daily airflow types over the British Isles, *Clim. Dyn.*, **9**, 95–105, 1993.

Hurrel, J.W., Decadal trends in the North Atlantic oscillation: regional temperatures and precipitation, *Science*, **269**, 676–679, 1995.

Hutchinson, M.F., Stochastic space-time weather models from ground-based data, *Agric. For. Meteorol.*, **73**, 237–264, 1995.

Hutchinson, M.F., C.W. Richardson, and P.T. Dyke, Normalization of rainfall across different time steps, In: *Management of irrigation and drainage systems, 21–23 July 1993, Park City, Utah*, 432–439, Irrigation and drainage division, ASCE, US Department of Agriculture, 1993.
Jenkinson, A.F., and F.P. Collison, *An initial climatology of gales over the North Sea*, Synoptic Climatology Branch Memorandum no. 62, Meteorological Office, Bracknell, 1977 (unpublished document).
Jones, P.D., and P.M. Kelly, Principal component analysis of the Lamb catalogue of daily weather types: part 1, annual frequencies, *J. Climatol.*, **2**, 147–157, 1982.
Jones, P.D., M. Hulme, and K.R. Briffa, A comparison of Lamb circulation types with an objective classification scheme, *Int. J. Climatol.*, **13**, 655–663, 1993.
Katz, R.W., Role of stochastic weather generators in climate impact assessment, , In: *Ninth conference on applied climatology, January 15–20, 1995, Dallas, Texas*, 269–272, AMS, Boston, Mass., 1995.
Katz, R.W., and M.B. Parlange, Effect of an index of atmospheric circulation on stochastic properties of precipitation, *Water Resour. Res.*, **29**, 2335–2344, 1993.
Klemeš, V., Probability of extreme hydrometeorological events - a different approach, In: *Extreme Hydrological Events: Precipitation, Floods and Droughts* (Proceedings of the Yokohama Symposium, July, 1993), 167–176, IAHS Publ. no. 213, 1993.
Kruizinga, S., *Objectieve classificatie van dagelijkse 500 mbar patronen*, Scientific report W.R. 78–8, KNMI, De Bilt, the Netherlands, 11 pp. + 28 figures, 1978 (in Dutch).
Kruizinga, S., Objective classification of daily 500 mbar patterns, In: *preprints Sixth Conference on Probability and Statistics in Atmospheric Sciences, October 9–12, 1979, Banff, Alberta, Canada*, 126–129, AMS, Boston, Mass., 1979.
Lamb, H.H., British Isles weather types and a register of the daily sequence of circulation patterns, 1861–1971, *Geophysical Memoir*, **116**, HMSO, London, 85 pp., 1972.
Leytham, K.M., A joint rank test for assessing multivariate normality in hydrologic data, *Water Resour. Res.*, **23**, 2311–2317, 1987.
Marshall, R., Statistical analysis of storm and daily rainfall data, Ph.D. dissertation, Department of Civil Engineering, University of Bristol, UK, 1977 (unpublished document).
Matalas, N.C., Mathematical assessment of synthetic hydrology, *Water Resour. Res.*, **3**, 937–945, 1967.
Matyasovszky, I., I. Bogárdi, A. Bárdossy, and L. Duckstein, Space-time precipitation reflecting climate change, *Hydrol. Sci. J.*, **38**, 539–558, 1993.
Matyasovszky, I., I. Bogárdi, and L. Duckstein, Comparison of two general circulation models to downscale temperature and precipitation under climate change, *Water Resour. Res.*, **30**, 3437–3448, 1994.
Mejía, J.M., and I. Rodríguez-Iturbe, Correlation links between normal and log normal processes, *Water Resour. Res.*, **10**, 689–690, 1974.
Muthén, B., Moments of the censored and truncated bivariate normal distribution, *British Journal of Mathematical and Statistical Psychology*, **43**, 131–143, 1990.
Onof, C., and H.S. Wheater, Improved fitting of the Bartlett-Lewis rectangular pulse model for hourly rainfall, *Hydrol. Sci. J.*, **39**, 663–680, 1994.
Richardson, C.W., *A model of stochastic structure of daily precipitation over an area*, Hydrology paper No. 91, Colorado State University, Fort Collins, Colorado, 1977.
Richardson, C.W., Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resour. Res.*, **17**, 182–190, 1981.
Rodríguez-Iturbe, I., D.R. Cox, and V. Isham, Some models for rainfall based on stochastic point processes, *Proc. R. Soc. Lond. A*, **410**, 269–288, 1987.
Rodríguez-Iturbe, I., D.R. Cox, and V. Isham, A point process model for rainfall: further developments, *Proc. R. Soc. Lond. A*, **417**, 283–298, 1988.
Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado, 1980.

Schönwiese, C.D., J. Rapp, T. Fuchs, and M. Denhard, *Klimatrend-Atlas Europa 1891–1990*, Johann Wolfgang Goethe-Universität, Berichte des Zentrums für Umweltforschung, Nr. 20, Frankfurt am Main, Germany, 1993 (in German).

Schubert, S., A weather generator based on the European Grosswetterlagen, *Clim. Res.*, **4**, 191–202, 1994.

Toth, Z., and T. Szentimrey, The binormal distribution: A distribution for representing asymmetrical but normal-like weather elements, *J. Climate*, **3**, 128–136, 1990.

Vaessen, R.J., and H. Visser, *Analyse van klimaatgerelateerde tijdreeksen: een eerste verkenning en inventarisatie*, report 92252-MOF 90-3423, KEMA, Arnhem, the Netherlands, 68 pp., 1991 (in Dutch).

Wateren-de Hoog, B. van der, and H. Middelkoop, *Flooding of the River Rhine and atmospheric circulation patterns*, report GEOPRO 1992.02, Geographical Institute, Utrecht University, Utrecht, the Netherlands, 1992.

Wigley T.M.L., and P.D. Jones, England and Wales precipitation: a discussion of recent changes in variability and an update to 1985, *J. Climatol.*, **7**, 231–246, 1987.

Wilby, R.L., Stochastic weather type simulation for regional climate change impact assessment, *Water Resour. Res.*, **30**, 3395–3404, 1994.

Wilby, R.L., Simulation of precipitation by weather pattern and frontal analysis, *J. Hydrol.*, **173**, 91–109, 1995.

Wilby, R.L., B. Greenfield, and C. Glenny, A coupled synoptic-hydrological model for climate change impact assessment, *J. Hydrol.*, **153**, 265–290, 1994.

Wilks, D.S., Adapting stochastic weather generation algorithms for climate change studies, *Clim. Change*, **22**, 67–84, 1992.

Wilson, L.L., D.P. Lettenmaier, and E.F. Wood, Simulation of daily precipitation in the Pacific Northwest using a weather classification scheme, *Surv. Geophys.*, **12**, 127–142, 1991.

Wilson, L.L., D.P. Lettenmaier, and E. Skyllingstad, A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation, *J. Geophys. Res.*, **97** (D3), 2791–2809, 1992.

Woolhiser, D.A., Modelling daily precipitation—Progress and problems, In: *Statistics in the Environmental and Earth Sciences*, edited by A. Walden and P. Guttorp, 71–89, Edward Arnold, London, 1992.

Woolhiser, D.A., T.O. Keefer, and K.T. Redmont, Southern oscillation effects on daily precipitation in the southwestern United States, *Water Resour. Res.*, **29**, 1287–1295, 1993.

Yarnal, B., *Synoptic Climatology in Environmental Analysis: A Primer*, Belhaven Press, London and Florida, 195 pp., 1993.

Young, G.K., Discussion of 'Mathematical assessment of synthetic hydrology' by N.C. Matalas, *Water Resour. Res.*, **4**, 681–683, 1968.

Zorita, E., J.P. Hughes, D.P. Lettenmaier, and H. von Storch, Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, *J. Climate*, **8**, 1023–1042, 1995.

Zwart, B., Het klimaat verandert, *Zenit*, Februari, 90–94, 1993 (in Dutch).