KNMI-publication; 186-I

KNMI

Rainfall Generator for the Rhine Basin

Single-site generation of weather variables by nearest-neighbour resampling

Theo Brandsma, T. Adri Buishand

Koninklijk Nederlands Meteorologisch Instituut

De Bilt, 1997

KNMI-publication; 186-I

De Bilt, 1997

KNMI
PO Box 201
3730 AE De Bilt
Wilhelminalaan 10
The Netherlands
Telephone +31 30 220 69 11
Telefax +31 30 221 04 07

Authors: Theo Brandsma and T. Adri Buishand

UDC: 556.162
551.579.1
551.501.45
(282.243.1)

ISBN: 90-369-2124-4

KNMI-publication; 186-I

Rainfall Generator for the Rhine Basin

Single-site generation of weather variables by nearest-neighbour resampling

Theo Brandsma, T. Adri Buishand

Work performed under contract RI-2041 to Ministry of Transport, Public Works and Water Management,
Institute for Inland Water Management and Waste Water Treatment RIZA,
PO Box 17, 8200 AA Lelystad (The Netherlands), telephone +31 320 29 84 11, telefax +31 320 24 92 18

De Bilt, 1997

Table of contents

Summary

This report presents the first phase of a project for the development of a rainfall generator for the Rhine basin. The request for this generator arose from the need to study the likelihood of extreme river discharges in the Netherlands, using a hydrological/hydraulic model. This first phase deals with single-site generation of weather variables by nearest-neighbour resampling for the stations Essen, Kahler Asten, Trier, Frankfurt, Bamberg, Freudenstadt and Stuttgart in the German part of the basin. Daily precipitation and temperature data for these stations were made available for the period 1961–1990. The ultimate aim of the project is to generate simultaneous records of daily average rainfall amounts, together with daily temperatures, over about 30 subcatchments, each with an area of about 5000 km^2, to study the effects of unprecedented weather situations.

The essence of the nearest-neigbour resampling technique is that the variables for a new day are sampled with replacement from a selected set of historical data (the nearest neighbours or analogues). The method needs a feature vector $\mathbf{D}_t$ to find the nearest neighbours of the day of interest (day t). For unconditional resampling, $\mathbf{D}_t$ contains variables that characterize the state of the weather on day t–1, e.g., daily precipitation, temperature and atmospheric circulation indices. Conditional simulation on the atmospheric flow is also possible with the nearest-neighbour resampling technique. Circulation indices for day t are then included in $\mathbf{D}_t$. A (weighted) Euclidean distance measure is used to determine the k nearest neighbours of $\mathbf{D}_t$. From the k nearest neighbours, one day is resampled with a predefined probability. In case of unconditional simulation, the successor in the historical record to the resampled day becomes the new day, whereas in case of conditional simulation, precipitation and temperature are set equal to the values of the resampled day.

The data for Stuttgart are used to study the influence of k, the composition of $\mathbf{D}_t$, and some other options in the resampling procedure. For the intended application of the rainfall generator, the most important criteria for judgement are the ability to reproduce the autocorrelation structure and the distributions of multi-day winter (October-March) maximum precipitation amounts of the historical records. Additionally, cross-validation (leaving out one observation at a time and predicting that observation from the remaining observations) is a means to determine the composition of $\mathbf{D}_t$ and to find an optimum value for k.

The autocorrelation for Stuttgart is most sensitive to k and the variables included in $\mathbf{D}_t$. For precipitation, it is necessary to include both weather variables and circulation indices in $\mathbf{D}_t$ to reproduce the decay of the autocorrelation coefficients with increasing lag. Comparison of standard deviations of monthly precipitation sums is a sensible test for the latter. The autocorrelation of daily temperatures is systematically underpredicted. This underprediction is strongly reduced if the circulation indices are removed from $\mathbf{D}_t$. The value of k should not be too large (usually ≤ 20 for a historical record of 30 years). Cross-validation confirms the choice of the variables in $\mathbf{D}_t$.

The success of reproducing the multi-day winter (October-March) precipitation maxima of Stuttgart is strongly related to that of the autocorrelation properties. The main conclusion is, that both the weather variables and circulation indices

must be considered in $\mathbf{D}_i$. Furthermore, there is no clear preference to either the conditional or unconditional method.

For applications to the seven stations in the Rhine basin, we selected the most promising conditional and unconditional method. The distributions of the multi-day winter precipitation maxima are generally well reproduced. However, the median of the N-day maxima is systematically underpredicted (up to about 8% for the 20-day maxima) by both methods. In a similar study with data for De Bilt (the Netherlands) over different periods, the simulated means of the six highest maxima were lower than those observed (up to about 15% for 20-day maxima in both methods). Systematic underprediction can have different causes. For the 10-day and 20-day maxima it can be due to an inadequate reproduction of long-term memory. Further, resampling techniques are in principle unable to generate larger 1-day amounts than the largest historical values.

Unlike the resampled 1-day values, the multi-day values can exceed the observed maxima as a result of a different succession of historical days. Unprecedented maxima were generally found in 300-year simulations for the seven stations. It is expected that the simulations can be extended to much longer periods than the 300 years used so far. However, it is then desirable to extend the historical time series beyond the present 30-year period, to reduce the effect of sampling variability of observed rainfall.

The daily temperatures were used to determine snow accumulation and melt. Despite the systematic underprediction of the autocorrelation, the distributions of the total winter snow accumulation and the N-day snowmelt maxima were well reproduced.

We conclude that the nearest-neigbour method is a promising technique for simulating daily precipitation and temperature time series. The results justify the development of a multi-site extension. An important point in such an extension is the composition of the feature vector. Furthermore, special attention is needed for testing the reproduction of the spatial association of large N-day amounts.

1. Introduction

The Rhine is the most important river in the Netherlands. The river flows through several countries (Figure 1). Large parts of its basin are situated in Switzerland, Germany, France and the Netherlands. Protection against flooding is a point of continuous concern. In the Netherlands, safety standards are laid down in the Act of Safety against flooding. The safety situation is evaluated every five years. A design water level with a mean recurrence time of 1250 years is used for rivers in the non-tidal area. The design water level is mainly determined by the design discharge, i.e. the discharge that in a given year is exceeded with probability 1/1250. For the Rhine, an important piece of information forms the discharge record at Lobith, where the river enters the Netherlands (Figure 1). Several probability distributions have been fitted to the discharge maxima of that record. The large return period 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 basin into account (Delft Hydraulics and EAC-RAND, 1993). For this purpose, a hydrological/hydraulic model for the whole basin has to be developed. This model needs a statistical model to generate the spatial and temporal distribution of precipitation over the basin 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.

KNMI carried out a feasibility study into the possibilities of a rainfall generator (Buishand and Brandsma, 1996). In that study, the statistical techniques for generating daily rainfall sequences were reviewed. For a multi-site application in a large catchment like the Rhine basin, two, quite different, alternatives were discussed: (1) parametric time series modelling of the observed daily precipitation using a transformed multivariate AR(1) process, and (2) nonparametric resampling from historic data. Although some promising results have been reported for the two methods, there is a serious lack of knowledge about the reproduction of properties of extreme rainfall. Some of these properties, like the extreme-value distributions of multi-day amounts and the spatial association of large amounts during winter, are important for the peak discharges of the Rhine in the Netherlands.

Both the parametric multivariate time series model and the nonparametric resampling technique, allow for a linkage with the atmospheric circulation. This linkage has received much interest in order to improve the reproduction of the persistence of daily rainfall and to assess the effects of systematic changes in the atmospheric circulation, e.g. resulting from increased atmospheric greenhouse gas concentrations. For the Rhine basin application, it is also important to extend the generation of daily precipitation with that of daily temperature in order to account for the effect of snow and frozen soils on high river discharges.

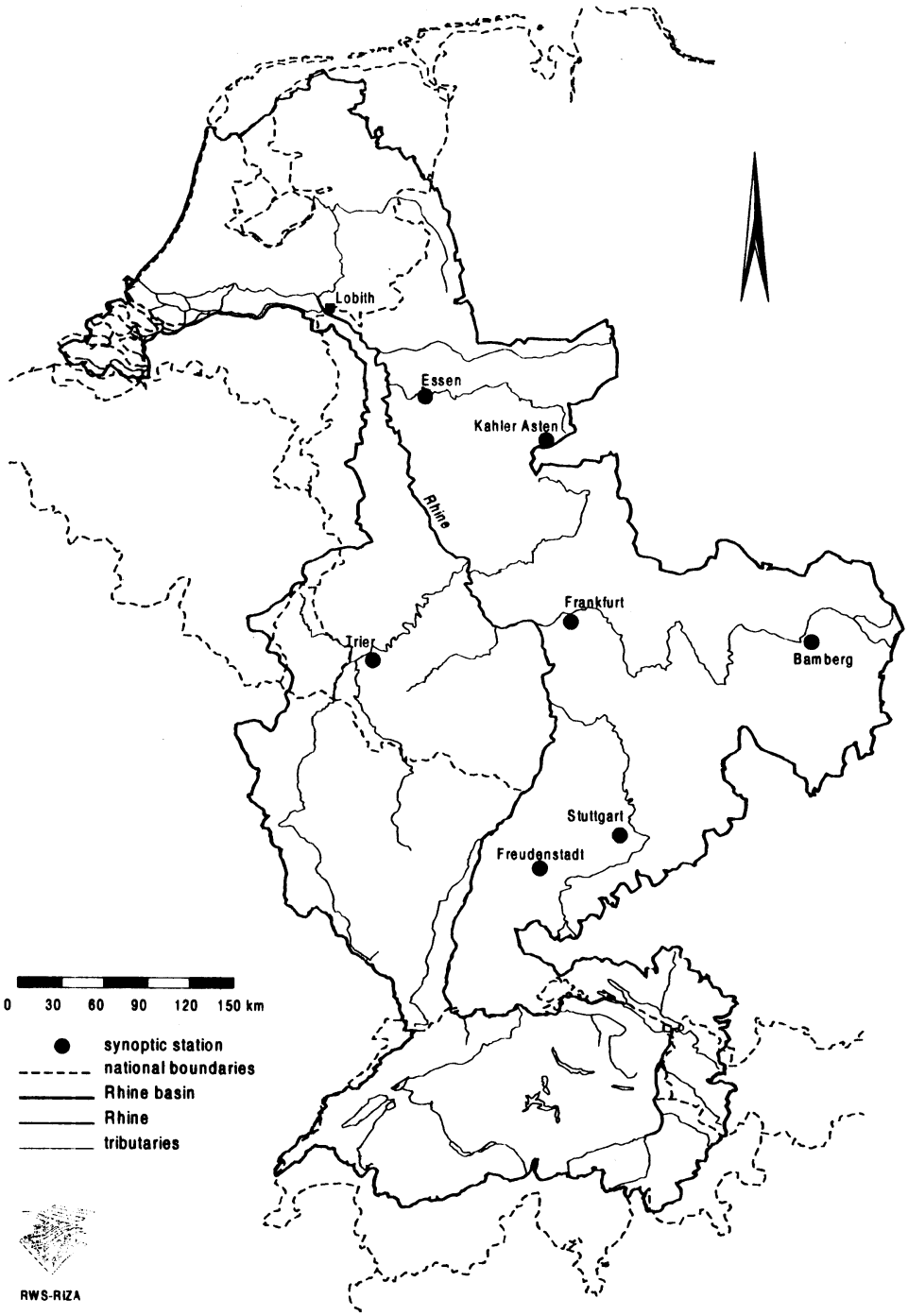

Figure 1: Location of Lobith in the Netherlands and the seven German stations used in this study in the drainage basin of the River Rhine.

In this report, nonparametric resampling is worked out further using data from Essen, Kahler Asten, Trier, Frankfurt, Bamberg, Freudenstadt and Stuttgart (Figure 1) in the German part of the Rhine basin (about 2/3 of the upstream catchment area). The main reason for this approach was that a study of the reproduction of relevant single-site properties already gives a good indication of the prospects of multi-site non-parametric resampling. This in contrast with parametric time series modelling, for which the multi-site extension requires a rather strong assumption of multivariate normality of the transformed daily data to use the AR(1) process. Such an additional assumption is not necessary in the

non-parametric approach, where the spatial association between the 1-day amounts is already automatically preserved.

The emphasis in this study, is on the reproduction of the autocorrelation structure and the distributions of multi-day maximum precipitation amounts. Snow accumulation and melt are also discussed. Although data for several sites are analysed, multi-site generation is not considered in the present report.

Standard resampling methods (permutation techniques, bootstrap) assume that the data are independent. These methods are not appropriate for resampling daily time series of hydrological and meteorological data, because of autocorrelation. Some mechanism is therefore needed to incorporate the dependence structure. This can be accomplished in different ways. One can discriminate between: (1) Markov modelling using a classification of representative states; (2) resampling from analogues or nearest neighbours; and (3) the moving blocks bootstrap. In the first approach, resampling is restricted to those days with the same representative state as the day of interest. Yakowitz (1979) followed the method to obtain synthetic sequences of daily river flows. A more recent application is the generation of daily precipitation conditional on the atmospheric circulation in Hughes *et al.* (1993) and Conway *et al.*(1996). To generate the precipitation for a given day t, the resampling procedure considered days with the same circulation type as that given for day t and the same wet/dry status as that generated for day $t - 1$. The conditioning on the wet/dry status of the previous day, is done to improve the reproduction of the autocorrelation properties of the generated precipitation amounts. A drawback of the method is that continuous variables are partitioned into discrete categories. Moreover, additional simulation of temperature requires an intractable growth of the number of categories to preserve the autocorrelation structure of the daily temperatures.

The analogue method has a long history in weather forecasting (Kruizinga and Murphy, 1983). Its use for generating synthetic precipitation sequences is, however, rather new. Zorita *et al.* (1995) used the analogue method to generate multi-site daily precipitation in a climate change study. The generated amounts for a given day were set equal to the observed amounts for the day in the historical record with the most similar sea-level pressure field. In a later paper (Cubasch *et al.*, 1996), the conditioning was done on both sea-level pressure and 700 hPa temperature. Apart from these recent applications in climate change studies, there is an independent development of the use of analogues in the hydrological literature. Lall and Sharma (1996) discussed a nearest-neighbour bootstrap for generating hydrological time series. Resampling is done from the successors to the historic k nearest neighbours rather than taking the observed precipitation for the closest neighbour only as in Zorita *et al.* (1995) and Cubasch *et al.* (1996). The use of the nearest-neighbour method for generating daily multivariate weather data (precipitation, maximum and minimum temperature, solar radiation, humidity and wind speed) has been demonstrated in Rajagopalan and Lall (1995). Atmospheric flow characteristics were not considered in that study. The emphasis was on the seasonal distributions of the daily values and their lag 0 and lag 1 cross-correlation coefficients (lag 1 autocorrelation, inclusive). The reproduction of the distributions of the multi-day maximum precipitation amounts has, however, not been verified.

The moving blocks bootstrap is a recent extension of the bootstrap for resampling stationary time series data (Künsch, 1989). The method considers all possible

blocks of a given length L, where L should be large enough that dependence between observations more than L time units apart can be neglected. A bootstrap time series is formed by resampling from these blocks and pasting them together. For $L = 1$, the method reduces to the standard bootstrap technique for resampling independent data. Vogel and Shallcross (1996) studied the value of the moving blocks bootstrap for generating annual flows in a reservoir design application. For univariate time series, Wilks (1997) developed some useful rules for the choice of L. The extension to multivariate data with different strengths of autocorrelation (like daily precipitation and temperature) is, however, not obvious. It is also unclear how the moving blocks bootstrap must be used for conditional simulation on the observed atmospheric flow.

At the beginning of the project, the nearest-neighbour method was thought to be the most promising nonparametric technique for generating daily precipitation and temperature in the Rhine basin. Section 2 provides the necessary background of the method. In Section 3 we present results for the seven German stations. The method is evaluated in Section 4. Finally, Section 5 gives an outlook for the next phase of the project.

2. Nearest-neighbour method

2.1 Time series generation by resampling

The principle of the nearest-neighbour resampling method is simple. The generation of precipitation and temperature on day t requires a feature vector $\mathbf{D}_t$ to find analogue situations in the historic data. In the method of Rajagopalan and Lall (1995) for generating multivariate daily weather data, $\mathbf{D}_t$ contains the values of the weather variables generated for day t–1 (method 1 in Figure 2). The k nearest neighbours (k-NN) of $\mathbf{D}_t$, in terms of Euclidean distance, are abstracted from the historic record. Let $t(j)$, $j=1,\ldots,k$ be the times associated with these nearest neighbours, such that the distance of $\mathbf{D}_{t(j)}$ to $\mathbf{D}_t$ increases with increasing j. The vector of weather variables following $\mathbf{D}_{t(j)}$, the successor to $\mathbf{D}_{t(j)}$, is denoted as $\mathbf{x}_{t(j)}$.

One of the successors of the k-NN is sampled using a discrete probability distribution or kernel $\{p_j\}$. For the uniform kernel, p_j is given by:

$$p_j = 1/k, \qquad j=1,\ldots,k \tag{1}$$

In Lall and Sharma (1996) the following decreasing kernel was recommended:

$$p_j = \frac{1/j}{\sum_{j=1}^{k} 1/j}, \qquad j=1,\ldots,k \tag{2}$$

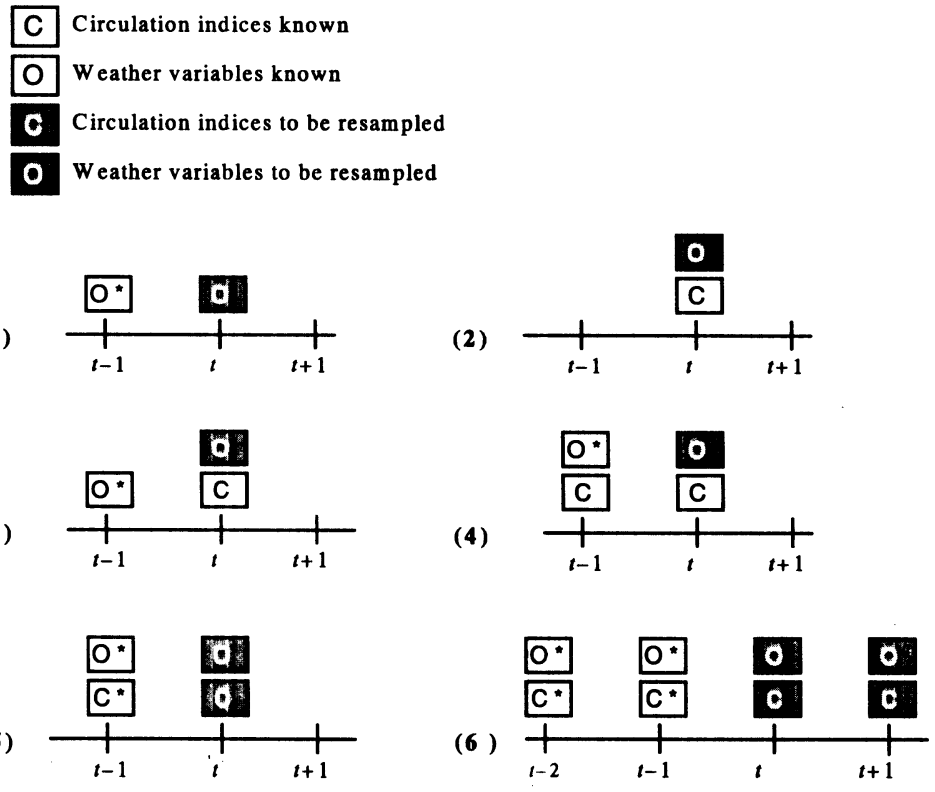

Figure 2: Six methods for the generation of new variables using different compositions of the feature vector $\mathbf{D}_t$ (open squares). The asterisks indicate that the corresponding variables are resampled values from the previous time step(s).

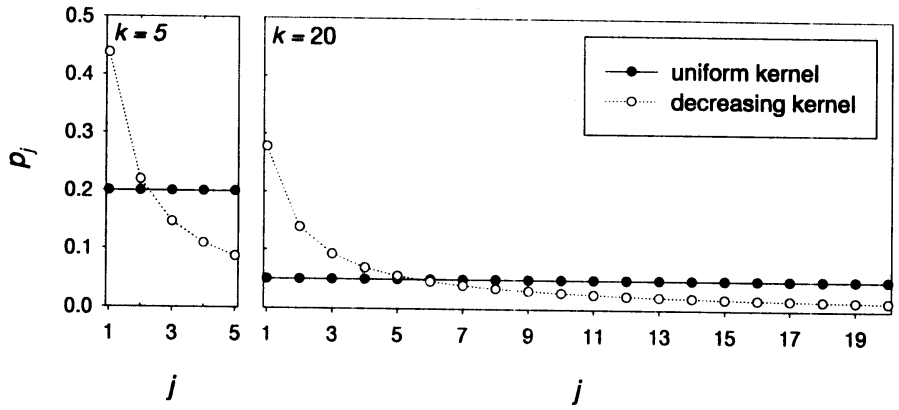

Figure 3: Resampling probability p_j as a function of the index j of the sorted Euclidian distances for $k = 5$ and $k = 20$ for both the uniform and decreasing kernels.

Figure 3 shows the two kernels for $k = 5$ and $k = 20$. The decreasing kernel gives relatively high probability mass to the closest neighbours, whereas the uniform kernel assigns the same probability to all neighbours.

We distinguish between conditional and unconditional simulation. Methods 2–4 in Figure 2, are examples of conditional simulation of weather variables given the atmospheric circulation. $\mathbf{D}_t$ then contains circulation indices on day t and possibly also the values of weather variables that were generated for previous days (methods 3 and 4 in Figure 2). Resampling occurs from the observed precipitation and temperatures on the days $t(j)$ in the nearest neighbourhood. Method 1 is an example of unconditional simulation, where $\mathbf{D}_t$ only contains the weather variables generated for the previous day. Methods 5 and 6 are extended versions of unconditional resampling. In method 5, $\mathbf{D}_t$ also includes the circulation indices generated for the previous day and, additionally, method 6 considers resampling of multiple days.

To account for the systematic annual cycle in the various weather variables, the search for the k-nearest neighbours of the feature vector is restricted to days in a specified moving window of width W_{mw} days, centred at the day of interest (see Figure 4). For example, for $W_{mw} = 61$ days and a historical time series of 30 years the Euclidean distances for a specific day are calculated for 61x30 = 1830 days.

A further reduction of seasonal variation can be achieved by working with standardised variables. In Rajagopalan and Lall (1995) standardisation was done by subtracting the calendar day's mean m_d and dividing by the calendar day's sample standard deviation s_d:

$$\tilde{x}_t = (x_t - m_d)/s_d, \qquad t = 1,\ldots,n;\ d = (t-1) \bmod 365 + 1 \tag{3}$$

where x_t and $\tilde{x}_t$ are the original and standardised variable, respectively, for day t, and n is the total number of days in the time series. For variables with a normal or almost normal distribution, $\tilde{x}_t$ usually takes values between –3 and +3. However, for daily precipitation the range of $\tilde{x}_t$ is quite different. For a dry day,

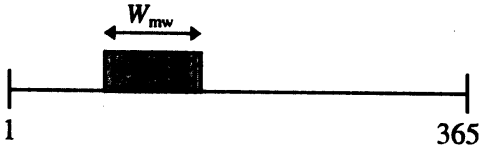

Figure 4: Moving window.

$\tilde{x}_t = -m_d / s_d \approx -0.5$ at a lowland station in the Rhine basin, whereas for days with heavy precipitation $\tilde{x}_t \approx 10$.

For daily precipitation, resampling of negative standardised values from adjacent days may result in negative precipitation amounts. Thus equation (3) does not give the most appropriate standardisation for that variable. In hydrology, division by the mean is a popular method to standardise non-negative variables. For daily precipitation, we therefore considered:

$$\tilde{x}_t = x_t / m_{d,\text{wet}} \tag{4}$$

where $m_{d,\text{wet}}$ is the calendar day's mean precipitation for wet days. For dry days $\tilde{x}_t = 0$ and for the most extreme wet days $\tilde{x}_t$ is about 10. Two alternatives to the standardisation in equation (4) were briefly examined: (1) $\tilde{x}_t = x_t / m_d$; and (2) $\tilde{x}_t = \sqrt{x_t / m_d}$. The use of m_d, instead of $m_{d,\text{wet}}$ in the first alternative, increases the range of $\tilde{x}_t$. Precipitation will then receive more weight in the search to the nearest neighbours. The square root transformation in the second alternative, reduces the weight of the extreme amounts. The right tail of the resulting distribution of $\tilde{x}_t$, is rather close to that of the normal distribution.

To reduce the effect of sampling variability, we used smooth approximations of m_d, $m_{d,\text{wet}}$ and s_d instead of the raw values (see further Section 3.1).

Through the standardisation, the elements v_{ti} of the feature vector $\mathbf{D}_t$ are dimensionless quantities. The weighted Euclidean distance between two vectors $\mathbf{D}_t$ and $\mathbf{D}_u$ is given by:

$$\delta_{tu} = \sqrt{\sum_{i=1}^{q} w_i (v_{ti} - v_{ui})^2} \tag{5}$$

where q is the number of variables in $\mathbf{D}_t$ and $\mathbf{D}_u$, and w_i is the weight associated with the ith variable. Unless stated otherwise, the weights w_i will be set equal to 1.

The final simulated value $x_{t,\text{sim}}$ for day t is obtained from the standardised resampled value $\tilde{x}_{t(j)}$ by inverting equation (3) or (4):

$$x_{t,\text{sim}} = m_d + s_d \tilde{x}_{t(j)} \tag{6}$$

$$x_{t,\text{sim}} = m_{d,\text{wet}} \tilde{x}_{t(j)} \tag{7}$$

Because $\tilde{x}_t \geq 0$ in equation (4), the simulated value $x_{t,\text{sim}}$ in equation (7) cannot be negative. This is not true for equation (6). Figure 5 presents a flow diagram of the full resampling procedure for the unconditional methods 1 and 5 of Figure 2.

Now it becomes clear that there are various options in the nearest-neighbour method, and that the results may depend on factors like:

- The number k of nearest neighbours used for resampling.
- The width W_{mw} of the running window.
- The type of kernel used for attaching probabilities to the k-NN.
- The composition of the feature vector (Figure 2).
- The use of weights in the calculation of distances.
- The method used to standardise the variables in the feature vector.

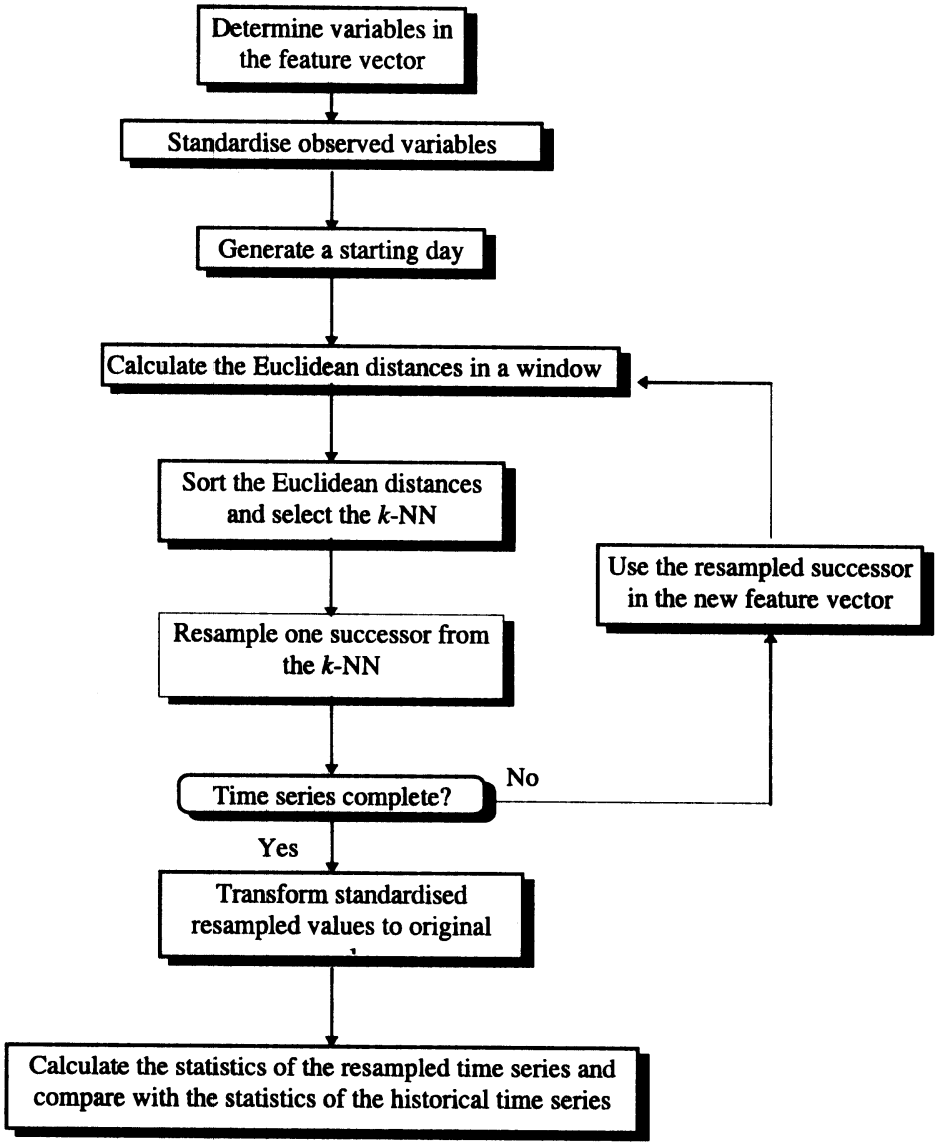

Figure 5: Flow diagram for methods 1 and 5 (see Figure 2).

2.2 Prediction and cross-validation

Lall and Sharma (1996) suggested the use of cross-validation to determine the number k of nearest neighbours and the variables in the feature vector. Cross-validation works by leaving out one observation at a time and predicting that observation from the remaining observations. The verification set thus always consists of one observation and the calibration set is formed by the other observations. The cross-validation score CV for a particular resampling model is the mean-squared prediction error over all possible predictions.

The k-NN prediction of a value x_t from all observations in a record is obtained as:

$$\hat{x}_t = \sum_{j=1}^{k} p_j x_{t(j)} \tag{8}$$

The error e_t of this prediction is:

$$e_t = x_t - \hat{x}_t = x_t - \sum_{j=1}^{k} p_j x_{t(j)} \tag{9}$$

However, if we derive the k-NN from all observations, the closest neighbour $\mathbf{D}_{t(1)}$ of $\mathbf{D}_t$ is $\mathbf{D}_t$ itself. Consequently, $x_{t(1)} = x_t$ and equation (9) can be rewritten as:

$$e_t^* = \frac{e_t}{1-p_1} = x_t - \sum_{j=2}^{k} p_j^* x_{t(j)} \tag{10}$$

where $p_j^* = p_j / (1-p_1)$ is the conditional probability that the value $x_{t(j)}$ is drawn when $x_{t(1)}$ is left out. The sum on the right-hand side of equation (10) gives the prediction $\hat{x}_t^*$ of x_t after deletion of $x_{t(1)}$. The error e_t^* corresponds with the deletion or jackknife residual in multiple regression (Green and Silverman, 1994).

The cross-validation score CV is now given by:

$$CV = \frac{1}{n_\ell} \sum_{t=\ell+1}^{n} (e_t^*)^2 = \frac{1}{n_\ell} \sum_{t=\ell+1}^{n} \frac{e_t^2}{(1-p_1)^2} \tag{11}$$

where ℓ is the number of time lags in the feature vector, and $n_\ell = n - \ell$ (the total number of possible predictions in a sequence of length n).

The CV-score in equation (11) applies to a single weather variable. A total CV-score for all weather variables of interest can be obtained as:

$$CV = \frac{1}{n_\ell} \sum_{t=\ell+1}^{n} \mathbf{r}_t^{*T} \mathbf{W} \mathbf{r}_t^* \tag{12}$$

where $\mathbf{r}_t^*$ is the vector of prediction errors at time t after deletion of the observation at that time, and $\mathbf{W}$ is a diagonal weight matrix.

In this study, the CV-scores are expressed in terms of the prediction errors for the standardised variables. CV gives the proportion of unexplained variance if the variable is standardised by equation (3). The weights on the diagonal of $\mathbf{W}$ should correspond with those in equation (5) for the Euclidean distances.

3. Results

3.1 Data

Daily precipitation and temperature data were made available by the Deutscher Wetterdienst via the 'International Commission for the Hydrology of the Rhine Basin' (CHR) for 34 synoptic stations in the German part of the Rhine basin for the period 1961–1990. For the study described in this report, we only analysed the data for Essen, Kahler Asten, Trier, Frankfurt, Bamberg, Freudenstadt and Stuttgart (Figure 1). Figure 6 presents some relevant precipitation and temperature characteristics of these stations, together with the station elevation. Note that the mean precipitation at the two highest stations (Kahler Asten and Freudenstadt), is about twice that at the other stations.

To incorporate atmospheric flow characteristics, we considered daily mean sea level pressure (MSLP) data from the UK Meteorological Office on a 5° latitude by 10° longitude grid. These data extend back to December 1880. For a grid centred at the Rhine basin (see Figure 7), we calculated three daily air-flow indices: (1) total shear vorticity Z; (2) strength of the westerly flow W; and (3) strength of the southerly flow S (see also Jones *et al.*, 1993). These three indices form the elements of the vector **C** in Figure 2.

Before resampling, the data have been standardised using the smoothed values of the calendar day's mean, m_d ($m_{d,\mathrm{wet}}$), and standard deviation, s_d, as described in Section 2.1. Figure 8 presents m_d and s_d for Z, W, S and T, and $m_{d,\mathrm{wet}}$ for P, together with the smooth approximations (P and T values apply to Stuttgart). The smooth curves are based on Friedman's supersmoother (Härdle, 1990). Before calculating the smooths, we repeated the values for $d = 336,\ldots,365$ for $d < 1$ and the values for $d = 1,\ldots,30$ for $d > 365$ to harmonize the smoothed values at the beginning and end of the year. For most variables, Figure 8 shows a large sampling variability. The use of the smooth approximations is, therefore, desirable. Note further that the largest standard deviations of the flow indices (vorticity, strength of the flow) are found in winter. The mean westerly flow is also relatively large in that season. The largest mean wet-day precipitation amounts are found in summer, which is due to the influence of convection (summer showers).

3.2 Definition of test cases

At the end of Section 2.1, we mentioned six factors that may influence the characteristics of the simulated data. The sensitivity of the autocorrelation properties and extreme N-day precipitation amounts to these factors were studied for 25 test cases as defined in Table 1. The first digit of each case number corresponds to the resampling method in Figure 2. We restricted the sensitivity study to the data from Stuttgart only. After the sensitivity study, we decided to use cases 4.1 and 5.2 to further explore the reproduction of extreme-value properties for the other stations. With the exception of Section 3.6.3, the resampled daily time series have a length of 30 years each. In Section 3.6.3 we consider long duration time series with a length of 300 years.

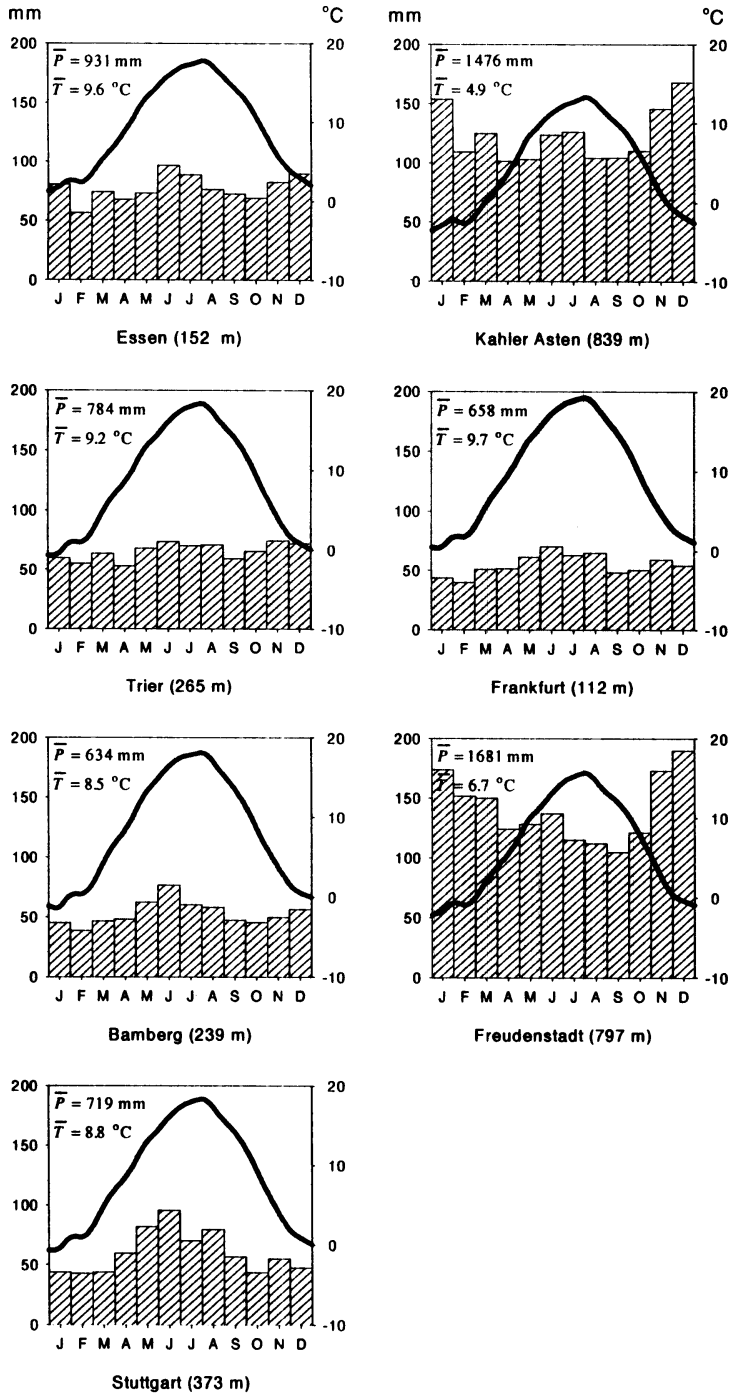

Figure 6: Precipitation and temperature characteristics of the seven stations (1961–1990) together with the elevation above mean sea level. The bars and lines give the mean monthly precipitation and mean calendar-day temperature, respectively. $\overline{P}$ and $\overline{T}$ in the various panels denote annual mean values for precipitation and temperature, respectively.

3.3 Autocorrelation for Stuttgart

Figure 9 compares the lag 1, 2 and 3 autocorrelation coefficients of daily precipitation and temperature for the historical data, with those for cases 2.1, 3.5, 4.1, and 5.2 (Table 1). For the historical data the standard error is also presented. Both the autocorrelation coefficients and their standard errors were estimated with the jackknife method of Buishand and Beersma (1993). For temperature (all 3 lags) and precipitation (lag 1), it is immediately seen that the autocorrelation is

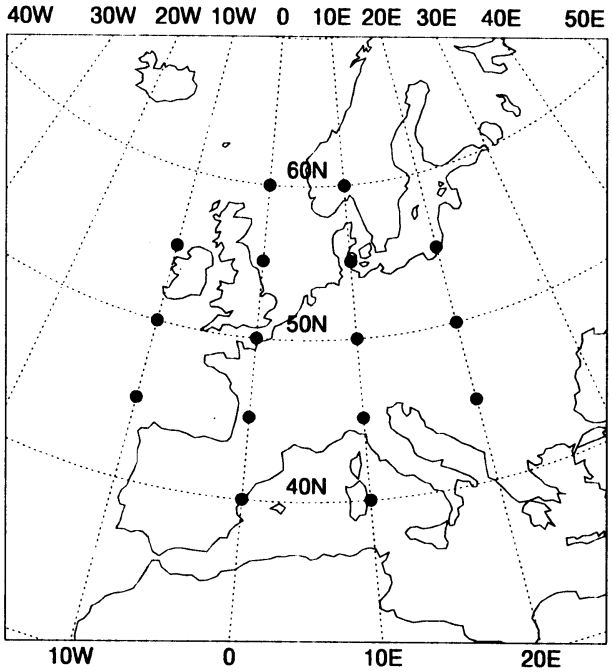

Figure 7: Grid points of mean sea-level pressure used for the calculation of the air-flow indices over the Rhine basin.

seriously underestimated in case 2.1, that uses only $\tilde{\mathbf{C}}_t$ to find the nearest neighbours. The reproduction of the autocorrelation is much improved by including $\tilde{P}^*_{t-1}$ and $\tilde{T}^*_{t-1}$ in $\mathbf{D}_t$ (case 3.5), although some bias still remains. A further improvement, for precipitation, is achieved by considering $\tilde{\mathbf{C}}_{t-1}$ in $\mathbf{D}_t$ as well (case 4.1). Case 5.2, that extends Rajagopalan and Lall's (1995) method for unconditional simulation, has similar performance.

Because a graphical comparison of all cases is both space consuming and difficult, we examined the mean difference (MD) and mean absolute difference (MAD) between the estimated autocorrelation coefficients of the historic and simulated time series:

$$\mathrm{MD} = \overline{R}_H - \overline{R}_S \tag{13}$$

$$\mathrm{MAD} = \frac{1}{12}\sum_{m=1}^{12}\left|R_{H,m} - R_{S,m}\right| \tag{14}$$

where $R_{H,m}$ and $R_{S,m}$ represent the estimated autocorrelation coefficients of a given lag, for the historical and simulated time series, respectively, for the mth month; $\overline{R}_H$ and $\overline{R}_S$ are the annual averages of $R_{H,m}$ and $R_{S,m}$, respectively. MD checks for a positive systematic difference between $R_{H,m}$ and $R_{S,m}$ over the year. The possibility of systematic negative differences over the year, i.e. too much autocorrelation in the simulated data, can be ruled out a priori for the nonparametric resampling techniques used here. MAD is not sensitive to the sign of the differences. It verifies whether the month-to-month fluctuations of the differences between $R_{H,m}$ and $R_{S,m}$ have the right order of magnitude for two time series with the same autocorrelation properties.

The expected value of MAD is given in Table 2. The table also presents the standard errors of MD and MAD. Details about the derivation of these statistical properties are given in Appendix A.

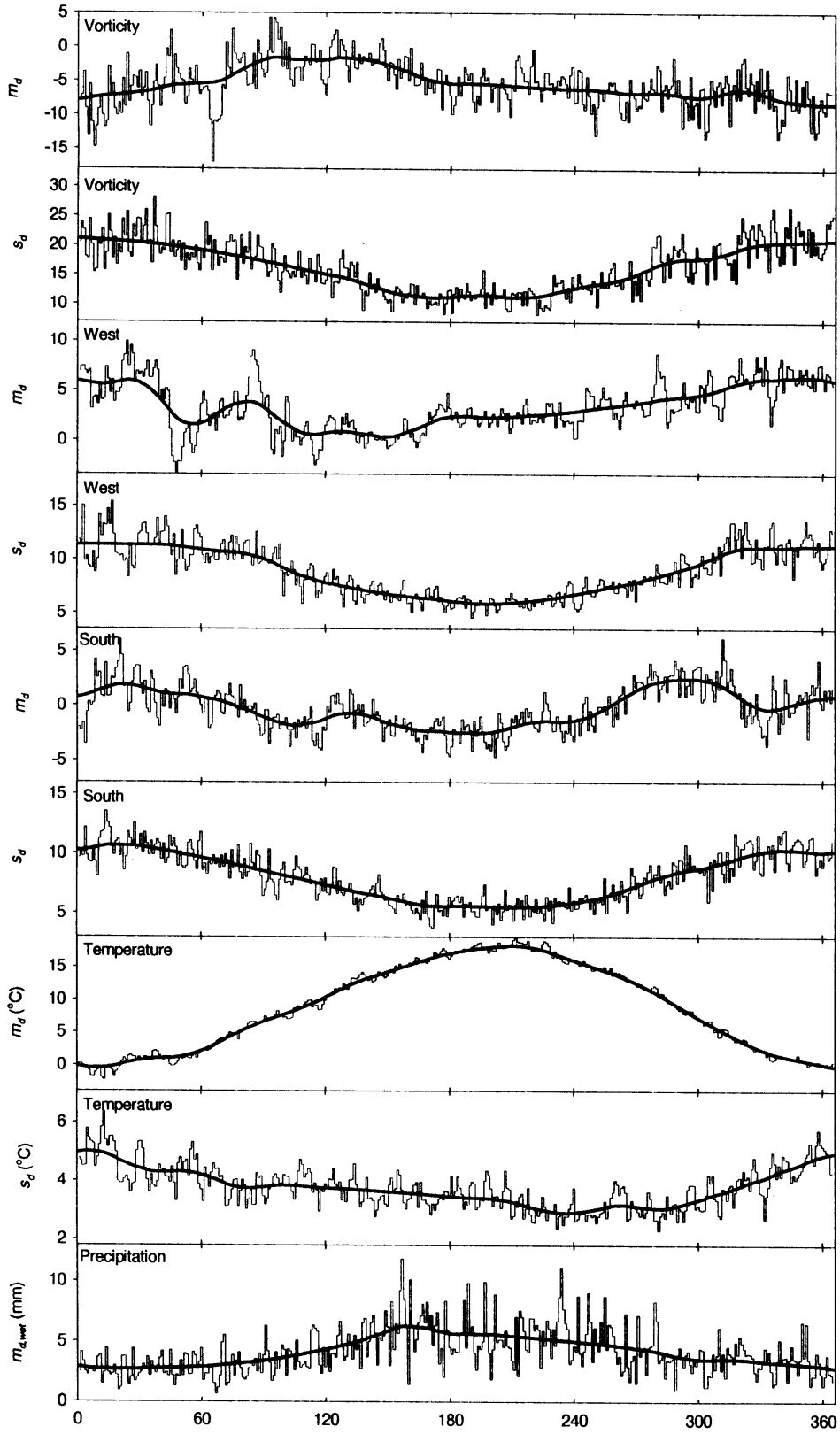

Figure 8: Values of m_d and s_d for Z, W, S, T and $m_{d,\mathrm{wet}}$ for P together with the smooth approximations (P and T values apply to Stuttgart) as a function of calendar day d for the period 1961–1990. The smooth curves are based on the supersmoother (Härdle, 1990).

case	W_{mw} (days)	k	w_i	$\mathbf{D}_t$	kernel	$\tilde{P}_t$
1.1	61	20	(1,1)	$(\tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
2.1	61	20	(1)	$(\tilde{\mathbf{C}}_t)$	uniform	$P_t / m_{d,\text{wet}}$
3.1	61	20	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.2	61	40	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.3	61	5	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.4	61	10	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
3.5	61	20	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
3.6	61	40	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
3.7	121	20	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.8	31	20	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.9	61	20	(1/3,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.10	61	20	(2/3,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.11	61	5	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.12	61	5	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$(P_t - m_d) / s_d$
3.13	61	5	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	P_t / m_d
3.14	61	5	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$\sqrt{P_t / m_d}$
3.15	31	10	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
3.16	121	40	(1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
4.1	61	20	(1/4,1/4,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{\mathbf{C}}_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
4.2	61	20	(1,1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{\mathbf{C}}_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$
4.3	61	20	(1,1,1,1)	$(\tilde{\mathbf{C}}_t, \tilde{\mathbf{C}}_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
5.1	61	10	(1,1,1)	$(\tilde{\mathbf{C}}^*_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
5.2	61	20	(1,1,1)	$(\tilde{\mathbf{C}}^*_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
5.3	61	40	(1,1,1)	$(\tilde{\mathbf{C}}^*_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	decreasing	$P_t / m_{d,\text{wet}}$
6.1	61	20	(1,1,1,1,1,1)	$(\tilde{\mathbf{C}}^*_{t-2}, \tilde{P}^*_{t-2}, \tilde{T}^*_{t-2}, \tilde{\mathbf{C}}^*_{t-1}, \tilde{P}^*_{t-1}, \tilde{T}^*_{t-1})$	uniform	$P_t / m_{d,\text{wet}}$

Table 1: Definition of test cases. The w_i values for the circulation apply to all three components of the vector $\tilde{\mathbf{C}}$: $\tilde{Z}, \tilde{W}$ and $\tilde{S}$. The asterisks in the feature vector $\mathbf{D}_t$, indicate that the corresponding variables are resampled values from the previous time step(s).

	MD*100			MAD*100		
	lag1	lag2	lag3	lag1	lag2	lag3
P (mean)	0	0	0	5.32	4.32	4.02
P (se)	1.97	1.60	1.56	1.19	0.96	0.94
T (mean)	0	0	0	2.20	4.14	5.37
T (se)	0.80	1.51	1.96	0.49	0.91	1.18

Table 2: Mean of MD and MAD, together with their standard errors (se), for autocorrelation of daily precipitation (P) and temperature (T) if there would be no differences in the autocorrelation structure of the observed and simulated data for Stuttgart.

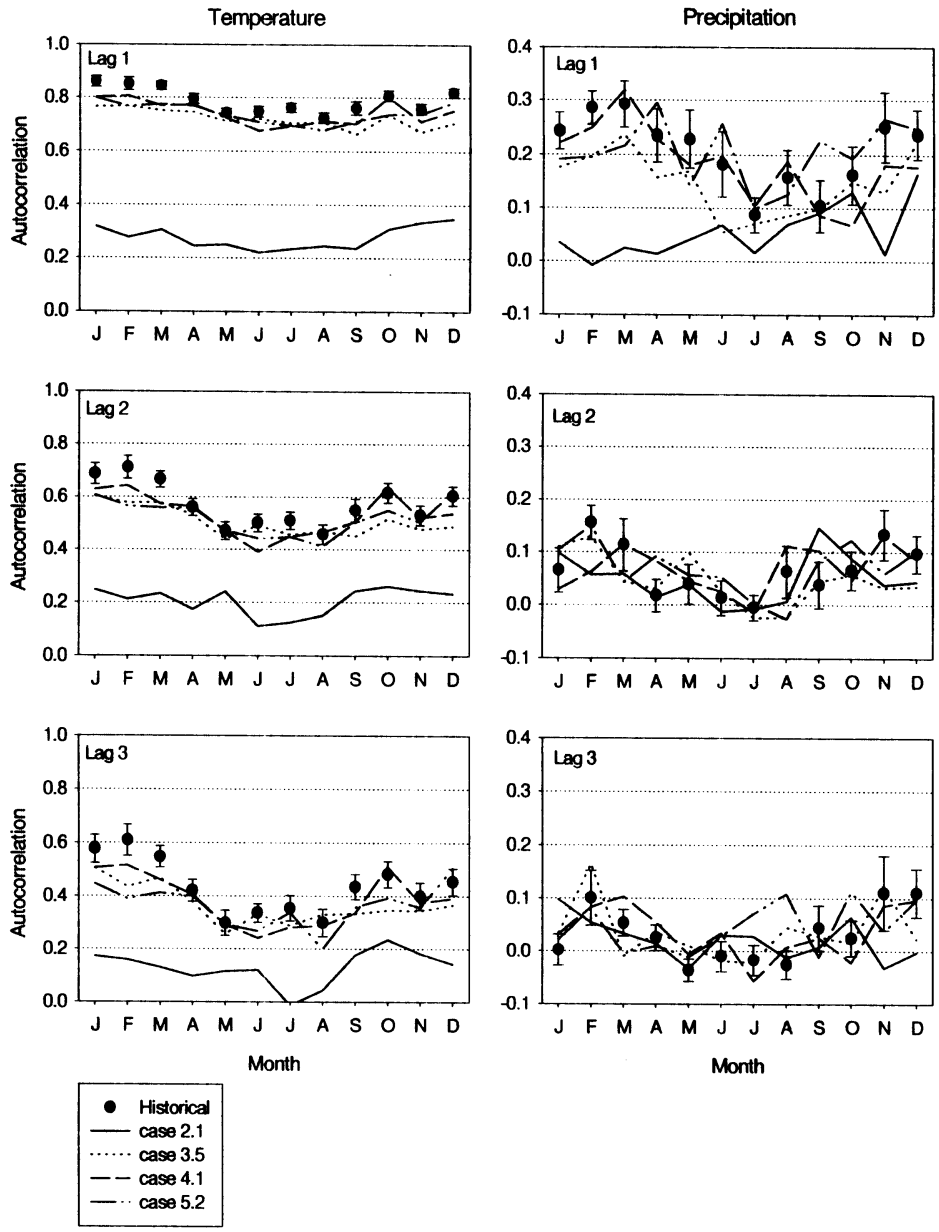

Figure 9: Lag 1, 2 and 3 autocorrelation coefficients of daily precipitation and temperature at Stuttgart for each month for the cases 2.1, 3.5, 4.1 and 5.2.

Tables 3 and 4 present the values of MD and MAD for precipitation and temperature, respectively. Despite the relatively short record length of 30 years in each test case, the results give a good indication of the sensitivity to the factors mentioned at the end of Section 2.1. For a one-sided test (MD) at the 5% level, the values may not deviate more than 1.65 times the standard error from the expected value, and for a two-sided test (MAD) not more than twice the standard error.

Table 3 shows that the lag 1 autocorrelation for precipitation is well reproduced for about half of the cases. For these cases, MD is not significant and the variation of the differences between $R_{H,m}$ and $R_{S,m}$ corresponds, according to the values of MAD, with that expected for two times series with the same autocorrelation properties. For temperature, Table 4 shows that the lag 1 autocorrelation is nearly always significantly underpredicted. To a certain extent, this underprediction can be alleviated by making appropriate use of the results of

the sensitivity study discussed below. The six factors mentioned at the end of Section 2.1 are considered separately.

Sensitivity to k

The sensitivity to k becomes clear by comparing cases 3.1, 3.2 and 3.3 (uniform kernel) and cases 3.4, 3.5 and 3.6 (decreasing kernel). For both precipitation and temperature, smaller values of k tend to result in a better reproduction of the autocorrelation properties. For cases 5.1, 5.2 and 5.3 (decreasing kernel) the best results are obtained if $k = 20$.

Sensitivity to W_{mw}

The sensitivity to W_{mw} is not obvious. Comparison of cases 3.1, 3.7 and 3.8 (uniform kernel) shows for precipitation a weak preference to $W_{mw} = 61$ days (case 3.1), while for temperature $W_{mw} = 121$ days (case 3.7) yields the best results. Because our interest is mainly in the reproduction of precipitation properties, we chose to use $W_{mw} = 61$ days in subsequent test cases.

Sensitivity to kernel type

The sensitivity to the type of kernel can be illustrated by comparing cases 3.1 and 3.2 (uniform kernel) with cases 3.5 and 3.6 (decreasing kernel). For the same value of k, the decreasing kernel improves the reproduction of the autocorrelation for temperature while the effect of the kernel type is not clear for precipitation. However, this does not mean that the decreasing kernel is better than the uniform kernel. For instance, comparison of cases 3.3 and 3.5 shows that improvement obtained by using the decreasing kernel with $k = 20$, can also be obtained (and even be better) for the uniform kernel with $k = 5$. Therefore, if we are free to vary

	MD*100			MAD*100		
Case	lag1	lag2	lag3	lag1	lag2	lag3
1.1	2.28	3.53	2.02	5.26	4.14	3.93
2.1	15.05	1.93	1.11	15.05	4.64	5.02
3.1	4.96	1.63	1.80	5.96	4.53	2.89
3.2	6.97	4.36	2.88	7.17	5.35	4.68
3.3	0.93	0.12	-0.78	4.80	4.95	3.23
3.4	1.52	0.12	0.92	5.61	3.32	5.16
3.5	5.98	2.00	-0.37	5.98	4.23	3.54
3.6	3.54	1.97	2.22	5.87	4.82	4.73
3.7	3.41	3.55	0.81	6.28	4.64	4.02
3.8	4.83	2.11	0.49	6.78	5.34	4.92
3.9	3.02	2.28	0.65	5.92	4.08	3.60
3.10	5.88	1.48	2.03	6.22	4.14	4.97
3.11	2.33	0.16	-0.08	3.97	5.04	3.37
3.12	2.72	1.61	2.00	4.84	3.87	5.61
3.13	1.24	1.64	0.77	5.53	4.57	4.30
3.14	3.41	1.52	1.72	6.02	5.01	3.30
3.15	7.24	2.00	0.81	7.42	2.83	3.94
3.16	5.44	3.04	0.87	8.51	5.00	3.94
4.1	2.23	-0.15	-0.46	3.70	3.10	3.15
4.2	3.21	0.86	0.76	4.85	3.87	3.85
4.3	4.01	-0.02	-0.92	6.71	5.50	4.31
5.1	2.76	0.14	0.55	4.05	4.00	4.82
5.2	0.13	-0.34	-1.32	5.50	4.35	5.22
5.3	2.38	0.27	1.06	3.51	4.30	4.63
6.1	2.71	-1.41	-0.47	3.81	3.88	3.86

Table 3: Values of MD and MAD for autocorrelation of daily precipitation at Stuttgart for the cases defined in Table 1.

k, it is relatively unimportant which type of kernel is chosen.

Sensitivity to the use of weights w_i

For the circulation indices, the sensitivity to the changes in the weights w_i in equation (5), can be seen by comparing case 3.1 with cases 3.9, 3.10 and 1.1 (uniform kernel). Note that method 3 with w_i for circulation set to zero, is identical to method 1 of Figure 2. For precipitation, the results are quite insensitive to a reduction of w_i, while for temperature there is an obvious improvement for smaller w_i.

Sensitivity to the method for standardising precipitation

The sensitivity to the method for standardising precipitation can be illustrated by comparing cases 3.11–3.14 (decreasing kernel). The autocorrelation is quite insensitive to the standardisation method. For both precipitation and temperature, the default method for standardising precipitation (case 3.11) performs slightly better than the other methods.

Sensitivity to the resampling method

The six methods of Figure 2 can be compared by considering the cases 1.1, 2.1, 3.1, 4.2, 5.2, and 6.1. For conditional resampling (cases 2.1, 3.1, 4.2), method 2 (only circulation indices in $\mathbf{D}_t$) is inferior to methods 3 and 4, for both precipitation and temperature. Method 4 performs slightly better than method 3 for precipitation. Both methods produce daily temperatures with too little autocorrelation. For unconditional resampling (cases 1.1, 5.2, 6.1), method 1 (no circulation indices in $\mathbf{D}_t$) reproduces best the autocorrelation coefficients of daily temperature. Methods 5 and 6 (circulation indices included in $\mathbf{D}_t$) work quite well

	MD*100			MAD*100		
Case	lag1	lag2	lag3	lag1	lag2	lag3
1.1	1.43	-2.06	-2.8	1.91	3.56	4.47
2.1	51.17	36.52	30.4	51.17	36.52	30.40
3.1	8.53	7.81	7.65	8.53	7.96	7.65
3.2	12.04	11.41	11.45	12.04	11.41	11.54
3.3	4.25	2.73	3.80	4.25	3.37	4.57
3.4	4.28	2.57	2.52	4.28	3.13	3.38
3.5	6.93	6.61	6.89	6.93	6.79	7.40
3.6	8.48	7.80	8.16	8.48	8.25	10.50
3.7	6.49	4.31	4.11	6.49	5.94	6.30
3.8	7.08	6.06	6.28	7.08	7.00	8.11
3.9	5.35	3.53	3.86	5.35	4.28	4.96
3.10	8.56	8.19	8.52	8.56	8.91	9.33
3.11	3.89	2.46	3.18	3.89	3.87	5.11
3.12	5.70	3.96	3.12	5.70	4.84	4.65
3.13	5.46	3.91	4.29	5.46	4.60	5.57
3.14	4.98	2.70	3.10	4.98	3.90	5.37
3.15	7.25	5.53	5.63	7.25	5.66	6.18
3.16	9.43	9.28	9.78	9.43	9.28	9.78
4.1	4.67	4.05	5.07	4.67	4.58	5.54
4.2	8.38	7.94	7.51	8.38	7.94	8.88
4.3	5.94	5.44	6.04	5.94	6.18	7.26
5.1	4.55	6.30	9.17	4.81	7.03	10.13
5.2	4.89	5.28	7.10	4.89	5.57	7.96
5.3	5.60	7.66	10.56	5.60	7.87	10.69
6.1	6.90	6.09	6.39	6.90	6.94	7.43

Table 4: Values of MD and MAD for autocorrelation of daily temperature at Stuttgart for the cases defined in Table 1.

for precipitation, but, like conditional simulation on the atmospheric flow, they have difficulties with the autocorrelation properties of daily temperature. It further turns out that the autocorrelation coefficients of the simulated vorticity indices are too low in these methods. The systematic departure from the observed lag 1 autocorrelation coefficient is about 0.05, which is comparable to that for the simulated daily temperatures. The autocorrelation coefficients of the simulated vorticity indices also show a too rapid decay with increasing lag.

Discussion

The sensitivity study in this section shows that the autocorrelation properties of daily precipitation and temperature are most sensitive to the value of k and the variables included in $\mathbf{D}_t$. The value of k should not be too large (usually ≤ 20 for a historical record of 30 years) for an adequate reproduction of autocorrelation properties. For conditional simulation on the atmospheric circulation, it is of major importance to include the weather variables on day $t-1$ in $\mathbf{D}_t$. For unconditional simulation, the inclusion of circulation indices degrades the reproduction of autocorrelation properties of the daily temperatures. This should be ascribed to the smaller influence of temperature on day $t-1$ on the selection of the nearest neighbours.

3.4 Standard deviation of monthly precipitation for Stuttgart

The standard deviation of monthly precipitation sums is strongly determined by the autocorrelation structure. For an arbitrary $N>1$, the following relation exists between the variance V_N of the N-day amounts and the autocorrelation coefficients:

$$V_N = \sigma_x^2 \left[N + 2\sum_{l=1}^{N-1} (N-l)\rho_{xx}(l) \right] \tag{15}$$

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

From equation (15) it follows that the positive autocorrelation, as observed in daily climate data, leads to relatively large values of V_N. Underestimation of the autocorrelation coefficients, will then naturally result in monthly precipitation sums with too little variation. In contrast to the comparisons for each individual lag in Section 3.3, a test on the standard deviation of monthly precipitation sums considers all lags simultaneously. Such a test is useful to discover a too rapid decay of $\rho_{xx}(l)$ with increasing l in the simulated data.

To compare the historical data with the simulated data, we calculated the mean standard deviation $\bar{s} = \sum_{m=1}^{12} s_m / 12$ of the monthly precipitation sums, where s_m is the standard deviation for the mth month. For the historical data $\bar{s} = 30.4$ mm with and estimated standard error $\mathrm{se}(\bar{s}) = 1.34$ mm. The latter was obtained using a bootstrap procedure (see Appendix B).

case	$\bar{s}$ (mm)	case	$\bar{s}$ (mm)	case	$\bar{s}$ (mm)	case	$\bar{s}$ (mm)	case	$\bar{s}$ (mm)
1.1	25.5	**3.4**	26.2	**3.9**	27.0	**3.14**	26.2	**4.3**	29.8
2.1	26.1	**3.5**	28.0	**3.10**	27.7	**3.15**	24.3	**5.1**	28.3
3.1	27.2	**3.6**	24.4	**3.11**	28.9	**3.16**	28.0	**5.2**	27.2
3.2	25.0	**3.7**	26.2	**3.12**	27.9	**4.1**	31.3	**5.3**	29.1
3.3	27.7	**3.8**	27.5	**3.13**	28.2	**4.2**	28.6	**6.1**	27.4

Table 5: Means of the standard deviations of monthly precipitation sums at Stuttgart for the cases defined in Table 1. For the historical data of Stuttgart: $\bar{s} = 30.40$ mm and $se(\bar{s}) = 1.31$ mm.

Table 5 presents $\bar{s}$ for each case. The table shows that 15 out of the 25 cases have values lower than the observed value of $\bar{s}$ minus twice[1] the standard error. It is noteworthy that case 2.1 performs better than cases 1.1, 3.2, 3.6 and 3.15, although for case 2.1 the reproduction of the lag 1 autocorrelation coefficient is worst of all (Table 3). Clearly, the strong emphasis on circulation on day t in $\mathbf{D}_t$ in case 2.1 results in a relatively good reproduction of higher order autocorrelation coefficients. Nevertheless, the monthly standard deviation is significantly underpredicted. The relatively poor performance of case 1.1, indicates that the autocorrelation of the resampled daily precipitation is too weak if there are no circulation indices in $\mathbf{D}_t$. Table 3 shows a non-significant underprediction of the autocorrelation coefficients for the first three lags for case 1.1. This contributes to the significant underprediction of the monthly standard deviations.

3.5 Cross-validation

Cross-validation is an objective method that can be used to find an optimum value for the number k of nearest neighbours and to determine the composition of the feature vector. For each of the six methods in Figure 2, we calculated the *CV*-score as a function of k for $W_{mw} = 61$ days, $w_i = 1$ for all i, and for a uniform kernel. Initial calculations showed that the *CV*-score is relatively insensitive to changes in W_{mw} and w_i. Furthermore, the type of kernel only affects the value of k for which the *CV*-score reaches its minimum, not the minimum itself.

Figure 10 presents the *CV*-scores of both P and T as a function of k for the six methods. Note that the asterisks of Figure 2 are removed, all variables in $\mathbf{D}_t$ are historical. Because the conditioning on the atmospheric circulation reduces the *CV*-score, it would not be fair, for simulation purposes, to directly compare the conditional methods (2,3,4) and unconditional methods (1,5,6). Therefore, we make a distinction between unconditional and conditional methods in the discussion of the results.

Two general features are apparent from Figure 10. The first feature is that the *CV*-score for precipitation is much larger than for temperature. On the one hand, this can be explained by the nature of daily precipitation, which is much more random than temperature. As a result, the *CV*-score for T is much more sensitive to the variables included in $\mathbf{D}_t$. On the other hand, the relative prediction errors for precipitation are relatively high through the use of a different standardisation

[1] The observed and simulated monthly precipitation sums are practically uncorrelated for the unconditional method, whilst there is a correlation coefficient of about 0.3 for the conditional method. In the first case, the standard deviation of the difference between the observed and simulated values of $\bar{s}$ should be about $1.4 \times se(\bar{s})$ and in the second case $1.2 \times se(\bar{s})$. For a one-sided test at the 5% level, these values should be multiplied by 1.65. The criterion of $2 \times se(\bar{s})$ is then quite reasonable. The condition that $\bar{s}$ must be approximately normally distributed is satisfied because it is an average of weakly skewed sample standard deviations.

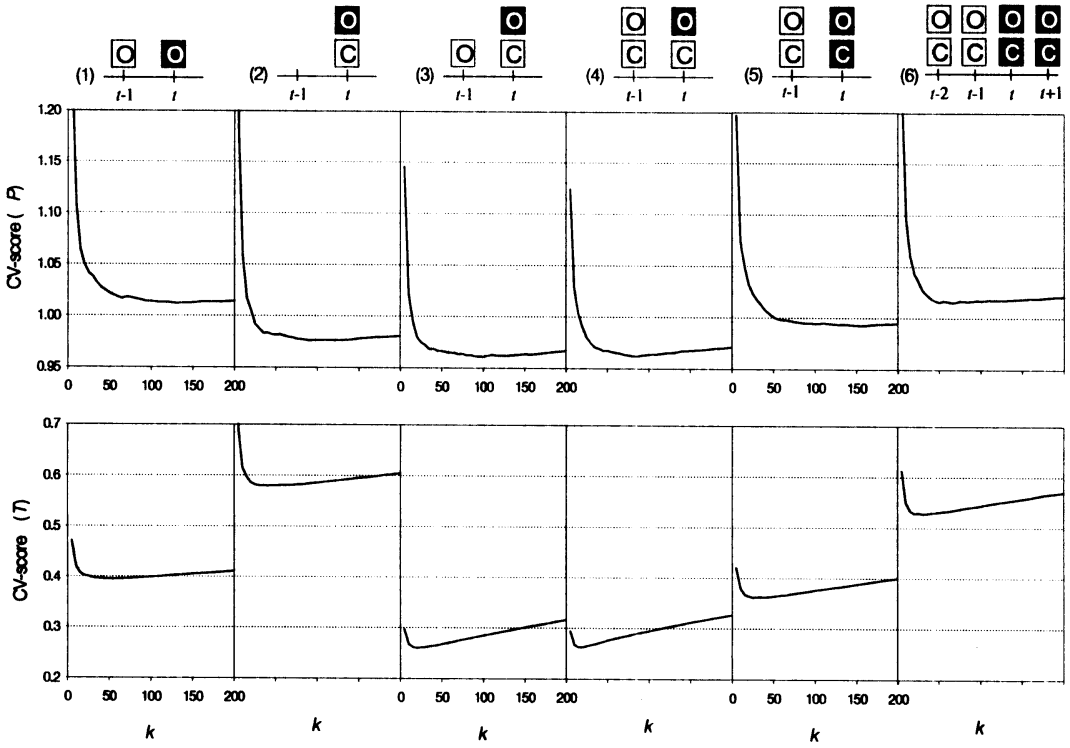

Figure 10: *CV*-score as a function of k for observed daily weather variables at Stuttgart for all cases in Figure 2 (uniform kernel; $W_{mw} = 61$ days; $w_i = 1$ for all i).

method. It is also seen that for temperature the *CV*-score has a more pronounced minimum than for precipitation. The second feature is that for $k < 25$ the *CV*-score decreases with increasing k as a result of the decrease of the variance of the predicted value $\hat{x}_t^*$ with k (see Appendix C). Therefore, k may not be too small for prediction purposes. However, for time series simulation it is important that the simulated data have the correct autocorrelation structure. The decrease of $\mathrm{var}(\hat{x}_t^*)$ with increasing k is then not relevant. The comparisons in Sections 3.3 and 3.4 showed that the autocorrelation properties of daily rainfall are well preserved if k is much smaller than the optimum value $k \approx 25$ for prediction purposes. These small values of k also resulted in the smallest underprediction of the autocorrelation coefficients of the daily temperatures.

For the conditional methods, starting from method 2, the inclusion of $\tilde{P}_{t-1}$ and $\tilde{T}_{t-1}$ in $\mathbf{D}_t$ (method 3), slightly improves the *CV*-score of P whereas the *CV*-score of T becomes much better. The inclusion of $\tilde{T}_{t-1}$ in $\mathbf{D}_t$ is thus of major importance. A further inclusion of $\tilde{\mathbf{C}}_{t-1}$ in $\mathbf{D}_t$ has little influence on the *CV*-scores.

For the unconditional methods, starting from method 1, the inclusion of $\tilde{\mathbf{C}}_{t-1}$ in $\mathbf{D}_t$ (method 5) slightly improves the *CV*-scores of both P and T. For block resampling (method 6), especially the *CV*-score of T becomes worse. This can be explained by the relatively large prediction errors in that method for the second day of the block, day $t + 1$ in Figure 2. In contrast to the other methods, the prediction for that day is based on observations for days $t - 2$ and $t - 1$ instead of day t.

3.6 *N*-day winter precipitation maxima

3.6.1 Stuttgart

For Stuttgart, we abstracted the N-day (N = 1,4,10,20) winter (October-March) precipitation maxima from the historical record and all simulated cases. Figures 11 and 12 compare the Gumbel plots of the observed 1-day and 10-day winter precipitation maxima with those for the cases 2.1, 3.5, 4.1 and 5.2. For both N = 1 and N = 10, cases 4.1 and 5.2 compare well with the historical data. Case 3.5 especially underpredicts the highest 10-day maxima, whereas for case 2.1 the underprediction of these maxima is more systematic as a result of the systematic underprediction of the autocorrelation. For the latter case, the relative magnitude of the underprediction of the percentiles of the 10-day winter maxima (15 to 20%) is comparable to that for the standard deviation of the monthly precipitation sums in winter (22%).

Although resampling techniques are in principle unable to generate larger daily values than the largest historical value, the simulated 1-day winter maxima in cases 4.1 and 5.2 in Figure 11 are larger than the historical maxima. This can be due to: (1) the use of a slowly seasonally varying mean to standardise the observations before resampling; and (2) the use of the moving window (Figure 4), which allows for resampling of days outside the boundaries of the winter half year.

For an objective verification of the reproduction of the N-day maxima distributions for N = 1,4,10 and 20, we considered the following three quantities:

1. The maximum of the N-day winter maxima (highest N-day amount in the record).

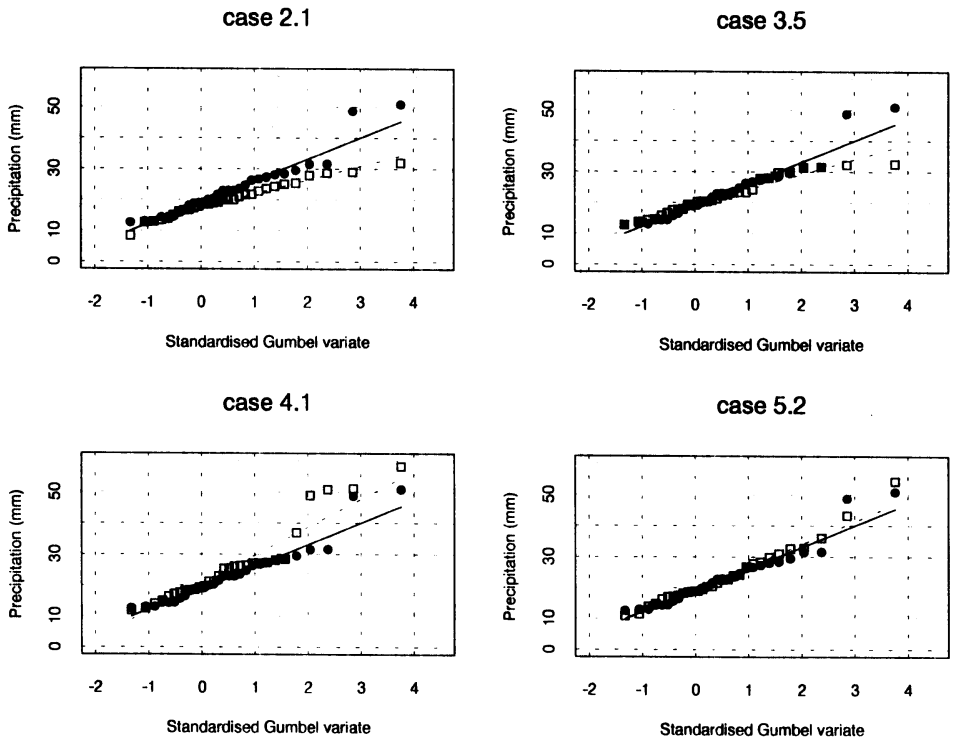

Figure 11: Gumbel-plots of 1-day winter (October-March) maxima for observed precipitation at Stuttgart (solid dots, solid lines) and simulated precipitation for cases 2.1, 3.5, 4.1 and 5.2 (open squares, dashed lines). The solid and dashed lines are probability-weighted moment fits (Landwehr *et al.*, 1979) to the historical and simulated data, respectively. The winter maxima are plotted using the median plotting position.

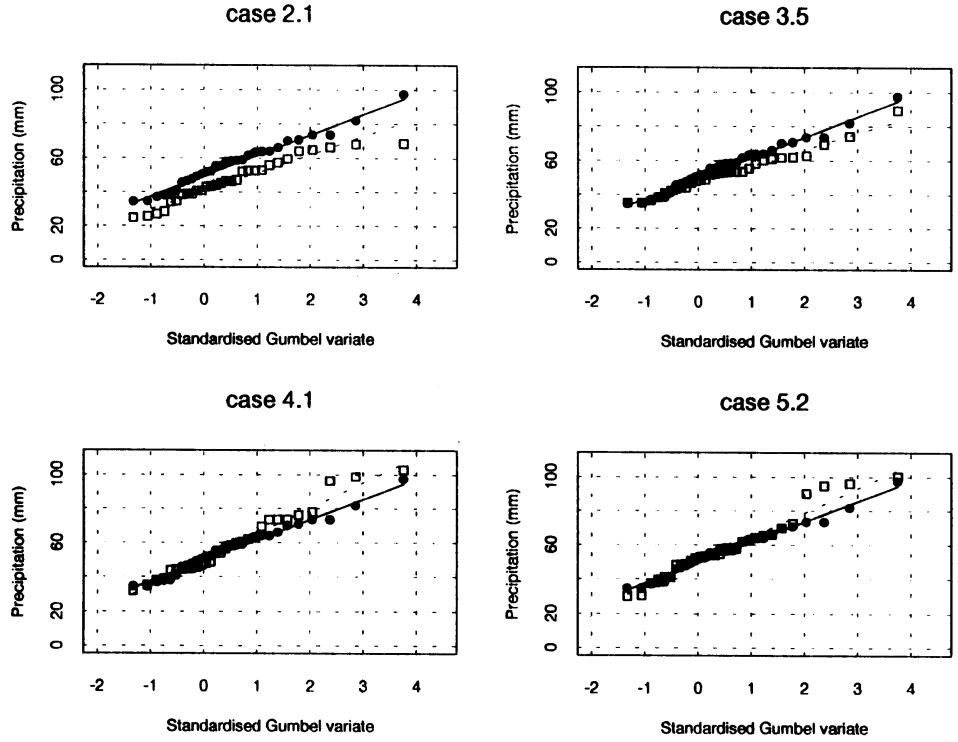

Figure 12: Gumbel-plots of 10-day winter (October-March) maxima for observed precipitation at Stuttgart (solid dots, solid lines) and simulated precipitation for cases 2.1, 3.5, 4.1 and 5.2 (open squares, dashed lines). The solid and dashed lines are probability-weighted moment fits (Landwehr *et al.*, 1979) to the historical and simulated data, respectively.

2. The upper quintile mean of the N-day winter maxima. For 30 years of data this is the mean of the 6 largest winter maxima. This mean value has an average return period of 12.5 years (Appendix D).
3. The median of the N-day winter maxima.

Table 6 presents the three quantities for the historical data of Stuttgart together with the simulations. Standard errors are also given. These standard errors assume that the N-day maxima follow the Gumbel distribution (see Appendix D for details). The differences between the simulated and observed values in Table 6 are in most cases less than twice the standard error. Surprisingly, the strongest indication of a systematic departure is found for the median, which is mostly lower in the simulated data. The largest discrepancies in Table 6 are mainly the result of an inadequate resampling procedure. It is therefore useful to study the effects of the various factors that determine the resampling procedure, as given at the end of Section 2.1, on the properties of the N-day precipitation maxima. This is done below for the same cases as in Section 3.3.

Sensitivity to k

Comparison of cases 3.1, 3.2 and 3.3 (uniform kernel) and cases 3.4, 3.5 and 3.6 (decreasing kernel), shows, except for case 3.2, no obvious preference to a particular value of k. For case 3.2, especially for large N, the results are worse. The autocorrelation coefficients and the standard deviation of monthly precipitation were also not satisfactorily reproduced for that case. Apparently, the combination of a uniform kernel with a large number of neighbours ($k = 40$) is not appropriate.

	maximum (mm)				upper quintile mean (mm)				median (mm)			
case	N=1	N=4	N=10	N=20	N=1	N=4	N=10	N=20	N=1	N=4	N=10	N=20
hist	**50.8**	**69.8**	**97.2**	**128.4**	**36.7**	**58.4**	**77.6**	**107.8**	**22.2**	**38.6**	**56.4**	**79.9**
(se)	*8.9*	*11.8*	*15.4*	*19.7*	*3.9*	*5.2*	*6.7*	*8.6*	*1.8*	*2.4*	*3.1*	*4.0*
1.1	33.6	51.3	80.0	115.2	29.2	48.2	66.5	96.9	19.8	35.3	47.9	67.8
2.1	31.9	59.5	68.6	96.9	27.9	49.7	65.3	90.7	19.5	30.1	45.9	70.0
3.1	32.9	63.0	82.6	105.9	30.5	55.4	73.9	100.3	23.9	37.5	51.8	69.0
3.2	48.8	60.9	88.1	107.0	35.9	52.5	71.0	87.0	20.2	33.1	49.4	66.9
3.3	34.3	72.3	87.3	118.6	30.9	51.8	71.4	102.0	19.4	36.6	52.4	69.7
3.4	43.9	70.9	90.1	120.8	31.7	56.5	80.6	109.4	18.1	34.4	49.3	65.7
3.5	32.5	59.1	89.3	127.1	31.2	49.2	70.0	101.3	20.9	35.6	52.6	76.1
3.6	49.5	71.0	85.7	112.7	39.9	59.1	72.9	104.5	23.0	42.1	57.0	74.2
3.7	45.0	50.2	71.9	104.5	30.6	46.3	65.9	90.4	19.8	32.1	48.1	68.5
3.8	59.3	87.8	107.9	126.6	40.2	65.2	82.1	109.2	22.4	37.2	53.7	70.3
3.9	47.7	66.2	87.4	131.9	31.8	57.9	80.3	107.1	19.3	40.7	56.7	77.9
3.10	50.9	65.2	92.7	126.9	40.9	59.8	76.4	103.3	25.8	38.9	56.9	76.8
3.11	54.7	70.3	103.5	124.0	35.0	59.2	78.4	99.6	19.7	36.5	54.5	73.3
3.12	51.0	78.7	108.0	139.8	36.4	61.9	85.8	115.0	22.0	38.3	58.1	79.2
3.13	29.6	69.4	92.7	108.4	28.6	56.2	79.1	95.0	22.4	36.4	52.5	67.0
3.14	51.3	65.6	97.7	121.6	33.6	53.7	76.7	100.8	18.7	34.5	50.3	71.2
3.15	26.7	56.9	104.5	137.6	24.2	44.7	68.9	93.0	18.5	29.8	44.4	64.1
3.16	39.0	70.4	97.6	121.2	29.3	57.0	80.7	104.5	19.5	35.7	54.5	78.1
4.1	58.2	70.7	102.8	159.6	45.7	65.1	87.6	120.6	24.1	36.0	54.9	74.2
4.2	51.4	69.0	99.9	130.0	35.4	56.9	79.1	111.0	22.0	36.6	48.8	70.2
4.3	50.1	56.5	88.5	145.6	42.4	54.7	74.4	106.2	21.5	35.9	56.0	72.9
5.1	56.4	78.4	88.9	113.7	36.4	54.3	75.3	105.2	22.8	36.3	51.0	72.0
5.2	54.3	81.7	100.2	123.1	38.3	65.2	87.2	110.8	21.0	38.4	54.1	77.5
5.3	57.7	102.7	118.8	144.4	39.5	70.6	93.4	124.1	20.2	36.9	51.1	76.1
6.1	57.4	85.7	90.3	94.4	39.2	59.8	74.9	89.9	20.9	38.3	49.9	73.3

Table 6: Maximum, upper quintile mean and median of the N-day winter (October-March) maxima for Stuttgart and for the cases defined in Table 1.

Sensitivity to W_{mw}

Comparison of cases 3.1, 3.7 and 3.8 (uniform kernel) shows that the results for W_{mw} = 121 days are inferior to those for W_{mw} = 31 and 61 days.

Sensitivity to kernel type

Comparison of cases 3.1 and 3.2 (uniform kernel) with cases 3.5 and 3.6 (decreasing kernel) shows no clear preference to the type of kernel.

Sensitivity to the use of weights w_i

For the circulation indices, the sensitivity to the changes in the weights w_i in equation (5), can be seen by comparing case 3.1 with cases 3.9, 3.10 and 1.1 (uniform kernel). Cases 3.9 and 3.10 with w_i = 1/3 and 2/3, respectively, perform somewhat better than case 3.1 for which w_i = 1. Setting w_i = 0 (case 1.1) deteriorates the reproduction of extreme rainfall amounts.

Sensitivity to the method for standardising precipitation

Comparison of cases 3.11–3.14 (decreasing kernel) shows that the results for case 3.13 are somewhat worse than for the other cases.

Sensitivity to the resampling method

The six methods of Figure 2 can be compared by considering the cases 1.1, 2.1, 3.1, 4.2, 5.2 and 6.1. Method 2 (only circulation indices in $\mathbf{D}_i$) is inferior to all other methods. Method 1 (no circulation indices in $\mathbf{D}_i$) also performs rather

poorly. Apparently, precipitation needs to be considered together with the circulation variables. The results of the remaining four cases, indicate that for conditional simulation the inclusion of $\tilde{\mathbf{C}}_{t-1}$ in $\mathbf{D}_t$ (case 4.2) in addition to $\tilde{\mathbf{C}}_t$, $\tilde{P}^*_{t-1}$ and $\tilde{T}^*_{t-1}$ (case 3.1) leads to a slight improvement. For unconditional simulation, there seems no advantage in resampling more than one day at a time (compare cases 5.2 and 6.1).

Discussion

The sensitivity study in this section, shows that for the reproduction of the statistical features of the N-day winter maxima, the resampling method is of major importance. The best results are obtained if the search for nearest neighbours considers the most recent resampled weather variables together with the circulation indices. For both the conditional and unconditional method, there is an indication that the median of the N-day maxima is underestimated (especially for large values of N).

3.6.2 Stuttgart and the other stations

We selected the cases 4.1 and 5.2 in Table 1, to examine the behaviour of the N-day winter precipitation maxima for all seven stations. Case 4.1 is a conditional method that gave good results for the autocorrelation structure and N-day winter maxima at Stuttgart. Case 5.2 is the unconditional method that best reproduces the lag 1 autocorrelation coefficient of daily rainfall and the distribution of N-day winter maxima. The feature vector of these cases contains both resampled weather variables and circulation indices. For each station, we generated four new time series of 30 years each.

Tables 7 and 8 present the results for cases 4.1 and 5.2, respectively. The bottom line gives the percentage difference between the historical time series and the average of the four simulations, averaged over all stations. Comparison of both tables shows no obvious preference to one of the two cases. Both the maximum and the upper quintile mean are well reproduced. The average percentage difference is small for these quantities and the magnitude of the differences between the observed and simulated values for the individual stations is generally not more than that expected from the presented standard errors. Quite remarkable are the extremely high 10-day and 20-day maximum amounts in the first simulated record for Freudenstadt in Table 7 (case 4.1a), which exceed the observed maxima by more than 200 mm. The median of the N-day winter maxima tends to be underestimated by the simulated data, in particular for $N = 20$. This is consistent with the earlier discussed results for Stuttgart in Table 6. The average percentage differences of –8.5 and –8.3% in Tables 7 and 8, respectively, are large compared with the value of 2.8% for the standard error of the average percentage deviation of the sample median from the true median, as derived from a bootstrap simulation in Appendix B.

To investigate whether the relatively high underprediction of the median also occurred in other periods, we carried out the same kind of simulation for De Bilt in the Netherlands. Three separate periods were considered: 1906–1935, 1936–1965 and 1966–1993. Table 9 presents the results for cases 4.1 and 5.2. The values for each case are averages of 4 simulations. The bottom lines give the percentage difference between the historical time series and the average of the simulations, averaged over the three periods. The table shows that the underprediction of the median is negligible for the unconditional method, but not

	maximum (mm)				upper quintile mean (mm)				median (mm)			
	N=1	*N*=4	*N*=10	*N*=20	*N*=1	*N*=4	*N*=10	*N*=20	*N*=1	*N*=4	*N*=10	*N*=20
Essen	**38.3**	**83.8**	**103.7**	**180.6**	**36.6**	**67.6**	**98.1**	**150.5**	**27.1**	**47.2**	**79.5**	**113.5**
se	*6.7*	*11.3*	*14.7*	*22.6*	*2.9*	*4.9*	*6.4*	*9.9*	*1.4*	*2.3*	*3.0*	*4.6*
4.1a	37.5	78.7	118.4	166.5	36.7	66.7	98.6	151.0	26.7	48.1	72.4	104
4.1b	36.4	75.9	121.7	174.4	34.2	66.6	104.4	141.5	28.7	47.4	79.7	110.8
4.1c	38.3	79.3	114.8	167.1	35.3	65.5	104.2	150.3	24.3	52.1	77.4	116.9
4.1d	38.4	62.5	103.7	153.9	35.8	59.2	91.4	131.6	30.0	48.8	73.9	103.7
Kahl.	**66.4**	**147**	**232.9**	**352.3**	**58.5**	**118.5**	**201.5**	**308.9**	**44.9**	**94.4**	**139.9**	**214.2**
se	*10.8*	*18.0*	*34.8*	*57.6*	*4.7*	*7.8*	*15.2*	*25.1*	*2.2*	*3.6*	*7.0*	*11.6*
4.1a	70.1	153.4	260.6	367.3	60.3	132.1	201.3	290.7	40.2	85.7	130.9	190.7
4.1b	54.2	110.6	191.8	294.5	52.5	105.9	172.5	251.4	40.2	88.0	129.6	197.3
4.1c	70.1	139.6	262.7	318.7	58.1	133.0	210.1	301.6	39.2	89.6	156.8	222.3
4.1d	74.2	163.1	319.4	440.1	61.0	119.7	209.7	293.5	38.9	78.2	128.4	198.0
Trier	**44.0**	**87.6**	**113.5**	**134.4**	**36.6**	**68.6**	**104.3**	**130.9**	**26.4**	**46.8**	**73.5**	**102.1**
se	*7.8*	*13.4*	*19.8*	*22.2*	*3.4*	*5.9*	*8.6*	*9.7*	*1.6*	*2.7*	*4.0*	*4.5*
4.1a	33.6	81.2	119.2	154.2	30.7	65.9	102.4	141.5	20.7	43.7	67.3	94.2
4.1b	38.8	73.9	125.7	150.0	34.3	59.0	95.1	124.0	23.6	45.8	66.3	91.1
4.1c	37.3	77.8	95.9	138.1	32.7	65.7	88.9	128.9	22.9	43.6	65.6	95.8
4.1d	46.3	79.2	118.6	156.2	36.8	72.7	100.8	136.8	24.7	47.8	66.6	95.2
Fran.	**34.6**	**94.9**	**112.9**	**125.8**	**31.8**	**62.7**	**89.1**	**115.4**	**24.2**	**41.7**	**60.6**	**81.9**
se	*6.5*	*12.5*	*17.5*	*20.1*	*2.8*	*5.4*	*7.6*	*8.8*	*1.3*	*2.5*	*3.5*	*4.1*
4.1a	30.8	65.1	88.7	109.0	28.9	50.5	72.7	98.8	21.7	38.3	56.0	75.6
4.1b	32.3	57.6	95.3	145.0	30.8	54.7	81.1	108.3	21.1	38.7	59.8	82.6
4.1c	34.1	66.3	76.6	130.2	29.3	49.2	72.1	99.9	20.5	36.7	55.0	76.3
4.1d	32.2	68.8	105.1	138.8	30.7	61.0	84.3	111.5	24.2	38.0	57.9	77.8
Bamb.	**54.0**	**60.8**	**87.9**	**145.9**	**36.1**	**56.7**	**74.7**	**108.5**	**20.6**	**35.4**	**55.9**	**79.6**
se	*8.2*	*10.7*	*12.6*	*20.8*	*3.6*	*4.7*	*5.5*	*9.1*	*1.7*	*2.2*	*2.5*	*4.2*
4.1a	57.2	74.9	78.6	110.9	37.2	52.9	68.4	102.4	19.3	34.6	49.5	70.9
4.1b	55.1	66.4	83.3	121.3	36.1	55.9	75.6	99.9	20.0	36.1	51.2	72.4
4.1c	52.7	64.3	93.3	123.1	34.6	52.6	76.9	98.2	21.9	35.0	49.7	72.1
4.1d	44.6	59.4	78.1	106.6	33.5	52.3	67.6	98.3	19.5	35.9	54.0	75.0
Freu.	**112.6**	**246.4**	**295.7**	**443.9**	**105.8**	**213.3**	**276.4**	**407.5**	**76.1**	**150.5**	**209.7**	**297.4**
se	*24.3*	*46.8*	*54.6*	*78.5*	*10.6*	*20.4*	*23.8*	*34.3*	*4.9*	*9.5*	*11.0*	*15.9*
4.1a	109.5	262.6	541.3	611.3	90.5	209.7	334.4	437.6	66.7	134.6	204.8	267.4
4.1b	124.0	259.7	331.2	426.5	103.5	199.4	290.9	390.9	64.2	133.5	193.4	261.7
4.1c	103.9	226.9	325.1	449.0	92.8	185.5	286.6	389.8	64.0	107.8	154.6	215.2
4.1d	102.7	231.6	421.2	510.5	89.2	194.6	312.2	395.0	64.2	131.3	189.3	256.6
Stut.	**50.8**	**69.8**	**97.2**	**128.4**	**36.7**	**58.4**	**77.6**	**107.8**	**22.2**	**38.6**	**56.4**	**79.9**
se	*8.9*	*11.8*	*15.4*	*19.7*	*3.9*	*5.2*	*6.7*	*8.6*	*1.8*	*2.4*	*3.1*	*4.0*
4.1a	56.8	79.4	102.8	109.4	44.0	62.3	82.1	98.3	21.2	34.7	50.2	69.6
4.1b	50.4	76.1	89.4	110.3	36.1	59.3	69.9	98.0	23.0	34.6	49.1	73.3
4.1c	50.4	57.7	104.6	141.6	37.3	54.7	79.5	105.6	23.1	32.5	50.5	61.6
4.1d	46.9	81.6	103.7	118.0	32.2	60.2	82.5	109.7	20.1	36.1	54.0	75.0
% diff	-3.3	-6.3	5.5	-1.1	-4.2	-4.5	-1.7	-4.9	-7.4	-6.3	-7.4	-8.5

Table 7: Maximum, upper quintile mean and median of the *N*-day winter (October-March) maxima for seven stations for case 4.1 (historical data and four runs of 30 years each) . The bottom line gives the percentage difference between the historical time series and the average of the four simulations, averaged over all stations.

	maximum (mm)				upper quintile mean (mm)				median (mm)			
	N=1	*N*=4	*N*=10	*N*=20	*N*=1	*N*=4	*N*=10	*N*=20	*N*=1	*N*=4	*N*=10	*N*=20
Essen	**38.3**	**83.8**	**103.7**	**180.6**	**36.6**	**67.6**	**98.1**	**150.5**	**27.1**	**47.2**	**79.5**	**113.5**
se	*6.7*	*11.3*	*14.7*	*22.6*	*2.9*	*4.9*	*6.4*	*9.9*	*1.4*	*2.3*	*3.0*	*4.6*
5.2a	38	79.2	113.2	155.5	34.9	65.8	100.4	144.0	24.4	43.1	72.2	108.7
5.2b	38.5	94.9	141.4	181.7	37.1	67.8	104.1	147.0	26.2	50.0	77.2	108.8
5.2c	36.7	76.6	157.0	241.8	36.0	67.6	105.9	158.5	26.9	51.7	81.9	112.6
5.2d	38.5	81.4	119.2	173.2	37.1	67.5	100.7	154.7	26.7	50.4	72.1	103.5
Kahl.	**66.4**	**147.0**	**232.9**	**352.3**	**58.5**	**118.5**	**201.5**	**308.9**	**44.9**	**94.4**	**139.9**	**214.2**
se	*10.8*	*18.0*	*34.8*	*57.6*	*4.7*	*7.8*	*15.2*	*25.1*	*2.2*	*3.6*	*7.0*	*11.6*
5.2a	63.7	129.9	206.1	292.7	59.2	119.5	188.1	275.9	44.2	89.9	142.0	213.7
5.2b	59.3	148.2	254.0	423.7	56.3	128.7	210.8	303.0	40.7	93.5	140.9	191.5
5.2c	66.8	130.3	215.4	301.6	60.5	117.5	193.1	268.7	45.8	91.2	145.3	206.0
5.2d	66.4	159.6	241.6	408.0	57.2	134.7	212.8	299.9	43.0	98.9	150.6	193.2
Trier	**44**	**87.6**	**113.5**	**134.4**	**36.6**	**68.6**	**104.3**	**130.9**	**26.4**	**46.8**	**73.5**	**102.1**
se	*7.8*	*13.4*	*19.8*	*22.2*	*3.4*	*5.9*	*8.6*	*9.7*	*1.6*	*2.7*	*4.0*	*4.5*
5.2a	37.1	77.0	112.4	150.6	33.1	69.3	102.7	135.7	23.2	48.7	75.9	98.1
5.2b	42.9	74.4	115.3	163.9	36.9	65.2	96.4	138.6	29.0	51.9	73.5	102.3
5.2c	47.8	73.8	115.4	158.1	37.4	64.8	98.3	136.2	22.2	44.7	65.1	93.5
5.2d	43.3	71.4	102.4	152.1	33.4	57.1	87.7	124.4	21.5	46.5	63.6	87.9
Fran.	**34.6**	**94.9**	**112.9**	**125.8**	**31.8**	**62.7**	**89.1**	**115.4**	**24.2**	**41.7**	**60.6**	**81.9**
se	*6.5*	*12.5*	*17.5*	*20.1*	*2.8*	*5.4*	*7.6*	*8.8*	*1.3*	*2.5*	*3.5*	*4.1*
5.2a	32.7	66.9	108.5	136.1	30.7	55.0	82.3	106.2	21.3	36.1	56.4	81.3
5.2b	33.9	90.5	125.9	136.6	30.6	54.3	78.5	107.9	20.6	36.0	53.0	79.2
5.2c	41.4	64.3	114.2	173.0	32.1	55.4	82.7	122.1	21.7	37.5	53.4	71.6
5.2d	38.4	115.2	158.1	173.8	34.3	79.2	106.0	130.2	21.8	39.0	58.9	78.7
Bamb.	**54.0**	**60.8**	**87.9**	**145.9**	**36.1**	**56.7**	**74.7**	**108.5**	**20.6**	**35.4**	**55.9**	**79.6**
se	*8.2*	*10.7*	*12.6*	*20.8*	*3.6*	*4.7*	*5.5*	*9.1*	*1.7*	*2.2*	*2.5*	*4.2*
5.2a	53.0	79.7	107.5	161.3	39.3	57.2	87.5	127.4	19.5	41.0	56.4	75.5
5.2b	31.5	63.2	86.4	106.9	29.9	54.0	77.6	100.5	21.1	37.2	53.0	71.4
5.2c	51.1	63.3	87.8	115.1	37.6	56.8	78.7	104.1	20.5	35.0	51.9	71.2
5.2d	50.5	56.9	78.0	137.4	31.0	50.3	68.4	110.7	19.1	35.7	47.5	65.0
Freu.	**112.6**	**246.4**	**295.7**	**443.9**	**105.8**	**213.3**	**276.4**	**407.5**	**76.1**	**150.5**	**209.7**	**297.4**
se	*24.3*	*46.8*	*54.6*	*78.5*	*10.6*	*20.4*	*23.8*	*34.3*	*4.9*	*9.5*	*11.0*	*15.9*
5.2a	116.2	275.3	400.1	451.1	106.4	226.4	313.6	431.9	62.9	123.0	178.1	243.3
5.2b	112.4	233.4	356.8	476.1	94.0	194.8	297.7	404.9	65.1	124.6	186.5	238.2
5.2c	116.2	296.2	380.1	472.7	100.4	221.5	320.1	388.2	61.3	131.6	193.5	242.2
5.2d	111.5	245.9	343.9	445.3	97.6	199.2	288.2	369.7	72.5	133.6	205.6	256.6
Stut.	**50.8**	**69.8**	**97.2**	**128.4**	**36.7**	**58.4**	**77.6**	**107.8**	**22.2**	**38.6**	**56.4**	**79.9**
se	*8.9*	*11.8*	*15.4*	*19.7*	*3.9*	*5.2*	*6.7*	*8.6*	*1.8*	*2.4*	*3.1*	*4.0*
5.2a	46.5	70.0	99.4	111.2	38.6	59.8	84.8	103.3	20.7	36.7	51.3	78.1
5.2b	41.8	79.7	82.7	125.6	32.7	56.7	69.8	101.0	22.8	36.8	48.3	68.9
5.2c	53.3	62.0	100.5	119.9	35.4	56.4	90.3	111.9	23.4	41.5	56.2	79.3
5.2d	55.7	83.2	99.4	134.8	37.1	63.1	82.5	109.3	20.8	38.7	55.1	71.5
% diff	-2.4	-1.5	8.9	4.8	-2.4	-1.2	2.2	-0.7	-6.6	-2.3	-5.3	-8.3

Table 8: Maximum, upper quintile mean and median of the *N*-day winter (October-March) maxima for seven stations for case 5.2 (historical data and four runs of 30 years each). The bottom line gives the percentage difference between the historical time series and the average of the four simulations, averaged over all stations.

	maximum (mm)				upper quintile mean (mm)				median (mm)			
	N=1	*N*=4	*N*=10	*N*=20	*N*=1	*N*=4	*N*=10	*N*=20	*N*=1	*N*=4	*N*=10	*N*=20
Bilt1	**46.2**	**68.2**	**111.9**	**160.1**	**33.0**	**60.0**	**94.5**	**136.2**	**20.3**	**38.7**	**62.0**	**87.4**
se	*7.2*	*12.7*	*18.4*	*26.6*	*3.2*	*5.5*	*8.0*	*11.6*	*1.5*	*2.6*	*3.7*	*5.4*
4.1	38.6	60.4	93.2	117.9	29.6	51.9	79.4	108.4	20.1	35.1	54.1	79.5
5.2	43.4	63.5	98.9	138.4	31.7	53.2	83.3	119.0	19.8	38.1	59.8	89.8
Bilt2	**56.2**	**109.6**	**120.3**	**187.5**	**42.9**	**73.3**	**108.8**	**162.8**	**23.2**	**42.1**	**65.2**	**95.3**
se	*10.5*	*17.2*	*24.6*	*36.5*	*4.6*	*7.5*	*10.7*	*15.9*	*2.1*	*3.5*	*5.0*	*7.4*
4.1	55.4	91.5	128	160	39.1	69.8	97.3	132	21.3	38.5	62.1	88.3
5.2	45.1	74.2	99.5	139	34.1	61.3	87.7	122.2	22.1	41.1	63.3	90.1
Bilt3	**41.9**	**79.7**	**112.2**	**158.3**	**34.6**	**69.5**	**100.5**	**149.9**	**24.8**	**48.6**	**76.7**	**109.8**
se	*6.9*	*13.4*	*16.8*	*27.2*	*3.0*	*5.8*	*7.3*	*11.9*	*1.4*	*2.7*	*3.4*	*5.5*
4.1	44.3	79.3	120.1	165.8	34.9	66	96.9	137.4	24.6	44.6	68.6	96.9
5.2	62.7	82.2	114	173.4	41	69.2	100.2	147.1	26.1	48.9	74.7	109.5
% diff (4.1)	-4.1	-9.5	-1.1	-12.1	-6.1	-7.8	-10	-15.9	-3.3	-8.7	-9.4	-9.3
% diff (5.2)	7.9	-12.0	-9.1	-10.0	-2.0	-9.4	-10.5	-13.1	-0.7	-1.1	-3.1	-0.9

Table 9: Maximum, upper quintile mean and median of the *N*-day winter (October-March) maxima for De Bilt for cases 4.1 and 5.2 (historical data and averages of four runs). Bilt1, Bilt2, and Bilt3 refer to the periods 1906–1935, 1936–1965, and 1966–1993, respectively. The bottom lines give the percentage difference between the historical time series and the average of the four simulations, averaged over the three periods.

for the conditional method. Both methods, however, show an underprediction of the upper quintile mean of the *N*-day winter maxima, up to about 15% for $N = 20$. The largest differences are found for the first two periods 1906–1935 and 1936–1965. For the third period, 1966–1993, the differences between the observed and simulated upper quintile means are small. This is in agreement with our results for the German stations for more or less the same years.

From the comparisons with the observed *N*-day precipitation amounts, we have to conclude that the percentiles of the *N*-day winter maxima are too small in the simulated time series. The fact that the departures tend to increase with increasing *N*, suggests that long-term memory is inadequately reproduced by the model.

3.6.3 Long duration simulations

The ultimate goal of the rainfall generator is the simulation of unprecedented extreme rainfall situations over the Rhine basin. An interesting question is, therefore, how far the rainfall generator can provide higher *N*-day amounts than those observed. Some 30-year runs in Tables 7 and 8 already showed larger maxima than those observed in the historical record. To investigate this further, we carried out a 300-year simulation for all seven stations. A conditional method would need an additional model for generating circulation indices for that purpose. Consequently, the results in this section refer to case 5.2 (unconditional) only.

Figure 13 compares for Kahler Asten and Frankfurt the Gumbel plots of the 1-day and 10-day winter precipitation maxima for the historical record with the 300-year simulation for case 5.2. For reasons discussed in Section 0, the largest 1-day winter precipitation maxima of the 300-year simulations exceed the historical values, in particular at Frankfurt.

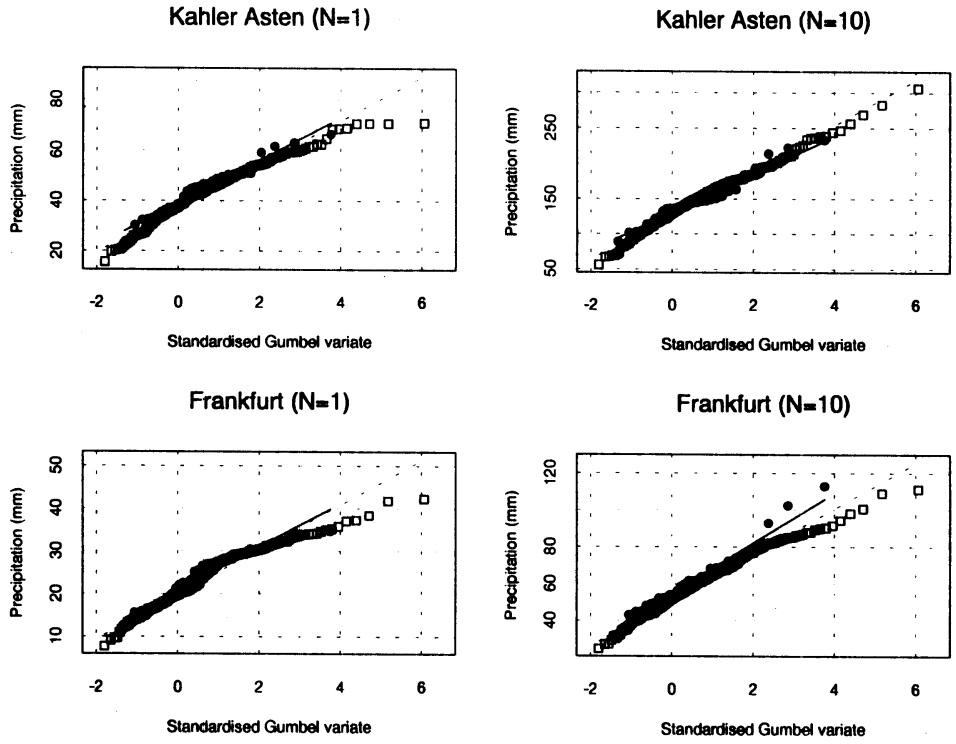

Figure 13: Gumbel-plots of 1-day and 10-day winter (October-March) maxima for observed precipitation at Kahler Asten and Frankfurt (solid dots, solid lines) and simulated precipitation for a 300-year run for case 5.2 (open squares, dashed lines). The solid and dashed lines are probability-weighted moment fits (Landwehr *et al.*, 1979) to the historical and simulated data, respectively.

For the peak discharges of the Rhine in the Netherlands, the 1-day precipitation maxima are of minor importance than e.g. the 10-day maxima. The Gumbel plot of the 10-day winter maxima in the 300-year simulation for Kahler Asten in Figure 13, shows values up to 74 mm larger than the largest historical value. The simulated 10-day maxima nicely follow the Gumbel distribution. For most of the other stations we obtained similar results. An exception is the 300-year simulation for Frankfurt (Figure 13), which contains no 10-day maxima larger than the largest historical value. The fact that the three largest observed 10-day winter maxima lie above the straight line for the Gumbel distribution in Figure 13, contributes to this phenomenon. On the other hand, the largest 10-day values of cases 5.2b, 5.2c and 5.2d in Table 8 for Frankfurt exceed the historic maximum.

3.7 Snowmelt

So far, we only considered the N-day winter precipitation maxima without distinguishing between rainfall and solid precipitation. However, large winter discharges of the Rhine may be partly influenced by snowmelt. In this section, we therefore explore the ability of the nearest-neighbour method to reproduce snowmelt for the seven stations in Section 3.6.2. Historical estimates and simulated values of snowmelt are derived from daily precipitation and temperature.

We assume that for $T < 0$ precipitation accumulates on the surface as snow. To calculate the N-day snowmelt maxima, we first transformed the snow into snowmelt using the degree days method. In this method, the amount of snowmelt on a certain day is proportional to the temperature excess (number of degrees Celsius above freezing point on that day), of course as long as there is solid

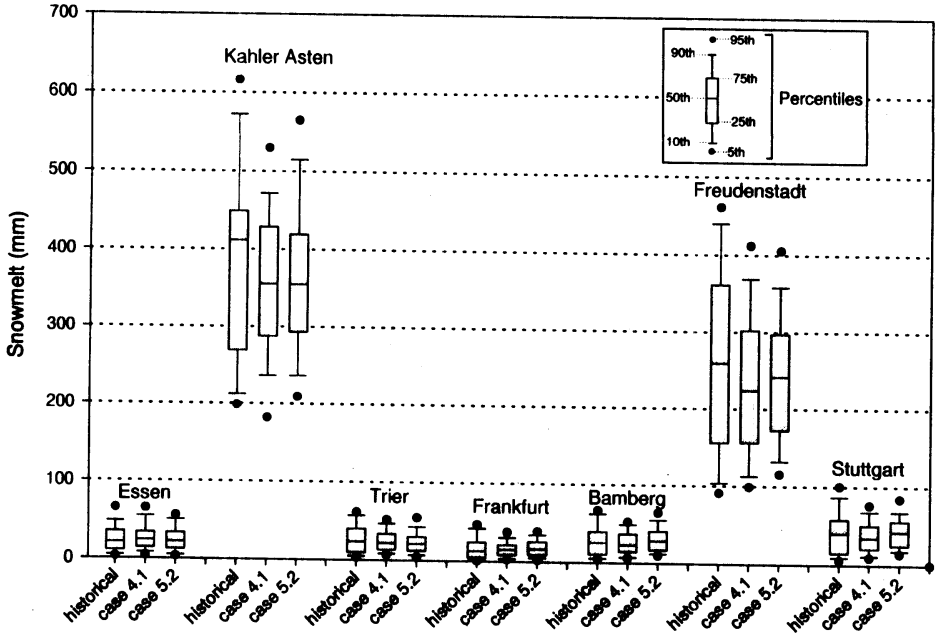

Figure 14: Boxplot of the accumulated winter (October-March) snowmelt for each station for the historical data and for cases 4.1 and 5.2. Each box for the simulations applies to a concatenated series of four 30-year runs.

precipitation stored on the surface. The constant of proportionality is known as the degree days factor (mm/°C). We set this factor equal to 4, which is an average of the values found in the literature (Linsley et al., 1988; Gray and Prowse, 1993).

Figure 14 compares the historical snow accumulation in winter (October-March) for each station with the simulations of Section 3.6.2. Each box for the simulations applies to the concatenated four 30-year runs. The figure shows that the correspondence between the historical and simulated values is quite good, though for the two highest stations, Kahler Asten and Freudenstadt, there is some tendency to underpredict the median and higher percentiles. The contrast between these two stations and the other stations is striking. For Kahler Asten and Freudenstadt, the mean historical snowmelt in the winter half year amounts to 26.0% and 15.7% of the mean annual precipitation, respectively. For the other stations this ranges only between 2.6% for Essen and Frankfurt and 6.0% for Stuttgart. Because of their higher location, Kahler Asten and Freudenstadt have also some snowmelt outside the winter half year. This amounts to 3.1% and 1.4% of the mean annual precipitation for these stations, respectively.

Figure 15 compares for Kahler Asten and Stuttgart the Gumbel plots of the 4-day and 20-day snowmelt maxima for the historical record with a 30-year simulation of case 5.2. The figure shows that the historical data are close to the straight line of the Gumbel distribution. The agreement between the historical and simulated snowmelt is good except for the 20-day snowmelt maxima of Kahler Asten. The difference between the upper quintile mean of the historical and simulated 20-day snowmelt maxima for Kahler Asten is, however, not significant.

Analogous to Section 3.6.2, we calculated the maximum, upper quintile mean and median for the N-day snowmelt maxima. The mean percentage differences between the observed and simulated upper quintile means turned out to be negative. For $N = 10$ and 20, case 5.2 underpredicts the upper quintile mean by 7.1% and 1.9%, respectively, the corresponding values for case 4.1 are 7.7% and

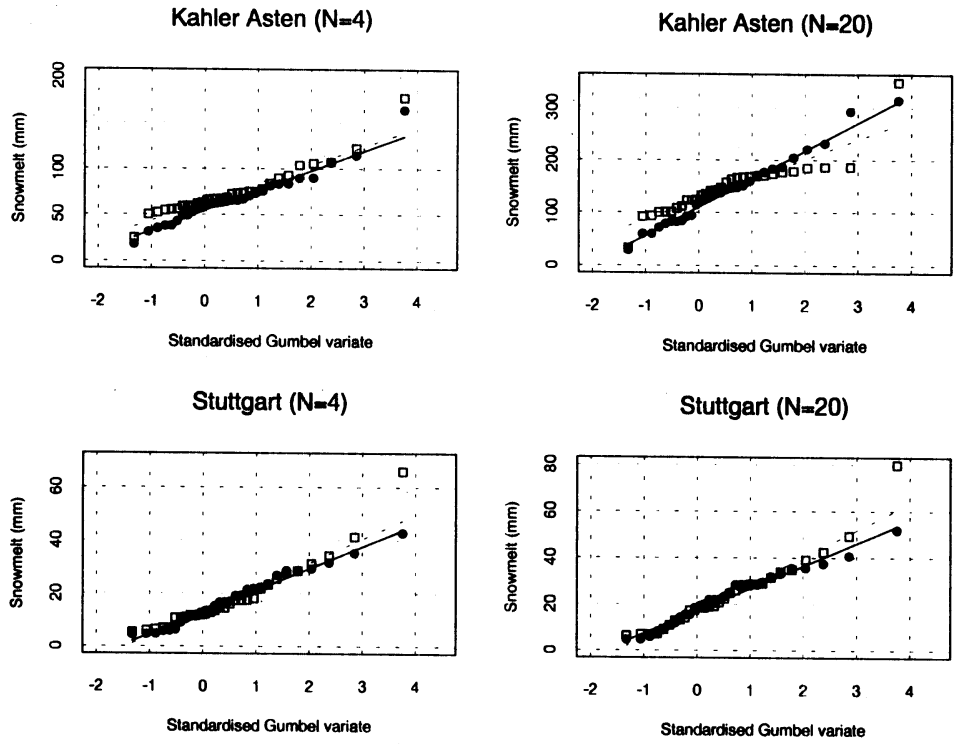

Figure 15: Gumbel-plots of 4-day and 20-day historical winter (October-March) snowmelt maxima at Kahler Asten and Stuttgart (solid dots, solid lines) and simulated snowmelt maxima for a 30 year run of case 5.2 (open squares, dashed lines). The solid and dashed lines are probability-weighted moment fits (Landwehr et al., 1979) to the historical and simulated data, respectively.

3.4%. In contrast to Section 3.6.2, there is no systematic underprediction of the median.

It should be noted that underprediction of the *N*-day snowmelt maxima is of minor importance compared to underprediction of *N*-day precipitation maxima, because the contribution of snowmelt to the peak discharges in the Netherlands is small compared to the contribution of rainfall.

The fact that the autocorrelation of the simulated temperatures is too low seems to be irrelevant for the distribution of snowmelt maxima. The influence of snow and frost on the antecedent conditions of flooding may be more sensitive to the temperature autocorrelation. These sensitivities may require some attention in a future stage of the project.

4. Discussion and conclusions

The results in this report show that the nonparametric nearest-neighbour method is a promising technique for simulating daily precipitation and temperature time series.

The emphasis in this study for the Rhine basin, is on the reproduction of extreme N-day precipitation amounts. This requires that not only the lag 1 autocorrelation coefficient is preserved but also the higher order autocorrelation coefficients (see equation (15)). The results in Section 3.3 show that the autocorrelation for Stuttgart is most sensitive to the number k of nearest-neighbours and the variables included in $\mathbf{D}_t$. The value of k should not be too large. Conditional simulation on the atmospheric flow needs the resampled temperature and precipitation for day $t-1$ in $\mathbf{D}_t$ to reproduce the autocorrelation properties. The effect of the use of circulation indices in $\mathbf{D}_t$ in case of unconditional simulation is for temperature not the same as for precipitation. For the autocorrelation of daily temperatures, it is beneficial to leave out the circulation indices (case 1.1), while for precipitation circulation indices have to be included to reproduce the decay of the autocorrelation coefficients with increasing lag. Comparison of standard deviations of monthly precipitation sums appeared to be a sensible test for the latter.

The results for cross-validation in Section 3.5, confirm the choice of the variables in $\mathbf{D}_t$. The number k of nearest neighbours that minimizes the CV-score may, however, be too large for time series simulation. For resampling from a historical record of 30 years, k should usually be 20 or less.

The results for the N-day winter (October-March) maxima of Stuttgart in Section 0 are, of course, a reflection of the results for the autocorrelation properties and the standard deviation of the monthly precipitation sums. The main conclusion is, that both the weather variables and circulation indices must be considered in $\mathbf{D}_t$. Furthermore, there is no clear preference to either the conditional or unconditional method.

The application of the two most promising methods (4 and 5 in Figure 2) to the seven stations in the Rhine basin in Section 3.6.2, shows that the most important properties of the N-day winter maxima are generally well reproduced. It turns out, however, that the median of the N-day maxima is systematically underpredicted (up to about 8% for $N = 20$). This holds both for the conditional method 4 and the unconditional method 5. For data from De Bilt over different periods, the unconditional method reproduces the median quite well. The conditional method shows, however, a similar underprediction as for the German stations. Moreover, the upper quintile means for De Bilt are considerably underestimated in both methods (up to about 15% for $N = 20$).

There are a number of explanations for the above-mentioned systematic underestimation of the percentiles of the distributions of the N-day winter maxima. For $N = 1$ the highest winter maximum can only exceed the observed maximum as a result of sampling of standardised values and the use of a moving window. This explains the negative percentage difference for the highest 1-day precipitation amount in winter in Tables 7 and 8. It may also have a slight effect on the extremes for $N > 1$. However, the fact that the largest underestimation of

extreme-value properties is found for the longest duration ($N = 20$ days) suggests that the temporal dependence structure is not adequately reproduced. Unfortunately, small departures from the observed autocorrelation properties generally pass statistical tests. It is therefore difficult to judge whether information for day $t - 2$, $t - 3$, ... should be included in $\mathbf{D}_t$ to improve the reproduction of long-term persistence. Failures to reproduce extreme-value properties may become more apparent in a multi-site extension with more data.

The results in Section 3.6.3 for the 300-year simulations show that unprecedented extreme rainfall situations can be obtained. As for the 30-year simulations, the multi-day maxima in the 300-year simulation generally follow the Gumbel distribution. It should be stressed that simulated multi-day values larger than the observed maximum are mainly the result of a different succession of historical days. It is expected that the simulations can be extended to much longer periods than the 300 years used so far. However, resampling from only 30 years of observed data can become very questionable. For De Bilt, the extreme-value statistics of observed precipitation in Table 9 show a considerable variation over 30-year periods (about 20% for the N-day median and N-day upper quintile mean), causing also differences between the simulated values based on different 30-year periods It is therefore desirable to extend the historical time series beyond the present 30-years period.

The results for snowmelt in 3.7, indicate that it is possible to obtain reasonable values for the N-day snowmelt maxima and the total winter snowmelt, despite the systematic underprediction of the autocorrelation coefficients of temperature (Section 3.3). Besides snowmelt, it may be important to consider frozen soils as well. Frozen soils may have a large effect on the runoff coefficient of a basin.

In the present study, no clear preference to either the conditional or unconditional method appeared. However, for future simulations it is important to simulate time series much longer than the length of the historical time series. As shown in Section 3.6.3, this requirement is much easier met with the unconditional method, because that method resamples the circulation indices along with the weather variables. For the conditional method, we are restricted to the length of the time series of historical MSLP (December 1880–present). To simulate longer time periods with the conditional method, we need a separate model for the simulation of circulation indices. It is likely that such a model better reproduces the autocorrelation properties of these indices than the unconditional methods presented here. In particular, a significant improvement is expected for vorticity. Furthermore, potential climate change applications of a rainfall generator, generally require conditional simulation on atmospheric flow patterns.

5. Outlook for the future

The results of this study justify the development of a multi-site extension. At first, this extension can be restricted to the German part of the Rhine basin. Rather than using daily point rainfall data from synoptic stations, daily average rainfall over subcatchments of about 5000 km^2 will be considered (about 20 subcatchments in Germany).

Multi-site generation by nearest-neighbour resampling is, in fact, rather straightforward. The composition of the feature vector $\mathbf{D}_t$ needs, however, special attention. A considerable growth of the dimension of $\mathbf{D}_t$ may not be desirable. Principal component analysis offers the possibility to describe the temperature field by a small number of amplitudes or scores, just as we used three air-flow indices to characterise the MSLP field. The strong non-normality of daily precipitation may hamper a principle component analysis. Daily rainfall over large regions could be considered instead. A relatively large weight could be attached to the most important regions for high river discharges at Lobith.

The reproduction of the spatial association of large N-day amounts in the winter season is of particular interest. A natural test seems the reproduction of the cross-correlation coefficients of the N-day winter maxima. However, a correlation coefficient measures the strength of linear dependence between two variables. Specific dependence functions, based on the number of joint exceedances of high thresholds, may be more suitable to describe the reproduction of the spatial association of extreme N-day amounts (Buishand and Brandsma, 1996). Regional estimation techniques may be needed to reduce sampling variability. Because the reproduction of properties of extreme precipitation is of main importance for the design discharge at Lobith, temperature could be omitted in the first stage of the multi-site extension.

Acknowledgements

We thank B.W.A.H. Parmet and H.C. van Twuiver (RIZA) 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). The daily precipitation and temperature data for the German stations were made available by the Deutcher Wetterdienst via the 'International Commission for the Hydrology of the Rhine Basin' (CHR).

References

Balakrishnan, N., and P.S. Chan, Order statistics from extreme value distribution, I: tables of means, variances and covariances, *Commun. Statist.–Simula.*, **21**: 1199–1217, 1992.

Bartlett, M.S., On the theoretical specification and sampling properties of autocorrelated time series, *J. Roy. Statist. Soc.*, **B8**: 27–41, 1946.

Buishand, T.A., Extreme rainfall estimation by combining data from several sites, *Hydrol. Sci. J.*, **36**: 345–365, 1991.

Buishand, T.A., and J.J. Beersma, Jackknife tests for differences in autocorrelation between climate time series, *J. Climate*, **6**: 2490–2495, 1993.

Buishand, T.A., and T. Brandsma, *Rainfall Generator for the Rhine catchment: a feasibility study*, Technical report TR-183, KNMI, De Bilt, 54 pp., 1996.

Conway, D., R.L. Wilby, and P.D. Jones, Precipitation and air flow indices over the British Isles, *Clim. Res.*, **7**: 169–183, 1996.

Cubasch, U., H. von Storch, J. Waszkewitz, and E. Zorita, Estimates of climate change in Southern Europe derived from dynamical climate model output, *Clim. Res.*, **7**: 129–149, 1996.

Dales, M.Y., and D.W. Reed, Regional flood and storm hazard assessment, Report No. 102, IH, Wallingford, 159 pp., 1989.

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).

Dupriez, G.L., and G. Demarée, *Totaux Pluviométriques sur des Périodes Continues de 1 à 30 Jours: I. Analyse de 11 Séries Pluviométriques de plus de 80 Ans*, Miscellanea Série A, No. 8, Institut Royal Météorologique de Belgique, Bruxelles, 154 pp., 1988.

Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, 92 pp., 1982.

Efron, B, and J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.

Gray, D.M., and T.D. Prowse, *Snow and Floating Ice*, Chapter 7 in Handbook of Hydrology (D.R. Maidment, ed.), McGraw-Hill, New York, 1993.

Green, P.J., and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*, Monographs on Statistics and Applied Probability 58, Chapman & Hall, London, 182 pp., 1994.

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

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.

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.

Kruizinga, S., and A.H. Murphy, Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands, *Mon. Wea. Rev.*, **111**: 2244–2254, 1983.

Künsch, H., The Jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, **17**: 1217–1241, 1989.

Lall, U., and A. Sharma, A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, **32**: 679–693, 1996.

Landwehr, J.M., N.C. Matalas, and J.R. Wallis, Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, **15**: 1055–1064, 1979.

Leone, F.C., L.S. Nelson, and R.B. Nottingham, The folded normal distribution, *Technometrics*, **3**: 543–550, 1961.

Linsley, R.K., M.A. Kohler, and J.L.H. Paulhus, *Hydrology for Engineers*, McGraw-Hill, London, 492 pp., 1988.

Rajagopalan, B., and U. Lall, A nearest neighbor bootstrap for resampling daily precipitation and other weather variables, Working paper WP-95-HWR-UL/013, 1995.

Vogel, R.M., and A.L. Shallcross, The moving blocks bootstrap versus parametric time series models, *Water Resour. Res.*, **32**: 1875–1882, 1996.

Wilks, D.S., Resampling hypothesis tests for autocorrelated fields, *J. Climate*, **10**: 65–82, 1997.

Yakowitz, S.J., A nonparametric Markov model for daily river flow, *Water Resour. Res.*, **15**: 1035–1043, 1979.

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.

Appendices

A. Statistical properties of MD and MAD

In this appendix we derive the first two moments of MD and MAD for two independent time series with the same autocorrelation properties. Thus, for all lags:

$$E(R_{H,m}) = E(R_{S,m}) \tag{A1}$$

The variance of an estimated autocorrelation coefficient R depends mainly on the record length and the autocorrelation properties of the underlying process (Bartlett, 1946). Therefore, for equal record lengths we may assume here that:

$$\text{var}(R_{H,m}) = \text{var}(R_{S,m}) = V_m = \sigma_m^2 \tag{A2}$$

To derive the variances of MD and MAD we further assume that the estimated autocorrelation coefficients for different calendar months are independent.

Equation (A1) implies E(MD)=0. For the variance of MD we obtain:

$$\text{var(MD)} = \text{var}(\overline{R}_H) + \text{var}(\overline{R}_S) = \frac{1}{12}\overline{V} + \frac{1}{12}\overline{V} = \frac{1}{6}\overline{V} \tag{A3}$$

where $\overline{V}$ is the mean of the twelve V_m values. In this study var(MD) was estimated by replacing the V_m values by their jackknife estimates from the historical observations (to have the same value in each comparison with the simulated data).

To derive an expression for the mean and variance of MAD, we make use of the fact that for large sample sizes the estimated autocorrelation coefficients are approximately normally distributed. Therefore $|R_{H,m} - R_{S,m}|$ is approximately distributed as $\sigma_m\sqrt{2}|\chi_m|$,where χ_m is a standard normal variable. The absolute value of χ_m is known as the half-normal variable. Its mean and variance are given by (Leone *et al.*, 1961):

$$E|\chi_m| = \sqrt{\frac{2}{\pi}} \tag{A4}$$

$$\text{var}|\chi_m| = \frac{\pi - 2}{\pi} \tag{A5}$$

For the mean of MAD we then obtain

$$\text{E(MAD)} = \frac{2}{\sqrt{\pi}}\overline{\sigma} \tag{A6}$$

where $\overline{\sigma}$ is the mean of the twelve σ_m values, and the variance is given by:

$$\text{var(MAD)} = \frac{\pi - 2}{6\pi}\overline{V} \tag{A7}$$

Analogous to the estimation of var(MD), the jackknife estimates of the V_m values from the historical data were used to estimate E(MAD) and var(MAD).

The assumption that the two series are independent is somewhat questionable in case of conditional simulation, because both time series are related to the observed atmospheric circulation. The jackknife estimate of the correlation coefficient between $R_{H,M}$ and $R_{S,M}$ is, however, not more than 0.13 (annual average, case 4.1).

B. Bootstrap standard errors

The bootstrap is a computer-based method to obtain distribution-free confidence intervals and estimates of bias, standard errors and prediction errors. The method was introduced by Bradley Efron in 1979. A standard reference is Efron (1982). Better accessible and more up-to-date is the monograph of Efron and Tibshirani (1993). In this appendix we restrict ourselves to the use of the bootstrap for estimating standard errors like those given in Sections 3.4 and 3.6.2.

The bootstrap estimate of the standard error of a statistic $\hat{\theta}$ is obtained by recomputing $\hat{\theta}$ for a large number of independent bootstrap samples. Each bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$ is a random sample of size n drawn with replacement from the original sample $X_1, X_2, \ldots, X_n$. A consequence of resampling with replacement is here that some X_i values will not appear in the bootstrap sample, while others will be repeated. Let $\hat{\theta}^*(b)$ denote the value of the statistic for the bth bootstrap sample, $b = 1, \ldots, B$. Then the bootstrap estimate of the standard error is the sample standard deviation of the $\hat{\theta}^*(b)$ values:

$$\hat{\mathrm{se}}^* = \left\{ \sum_{b=1}^{B} \left[\hat{\theta}^*(b) - \hat{\theta}^*(\circ) \right]^2 / (B-1) \right\}^{1/2} \qquad \text{(B1)}$$

where $\hat{\theta}^*(\circ)$ is the mean of the $\hat{\theta}^*(b)$ values.

In our assessments of the performance of nearest neighbour methods, each element of the sample is a vector of the twelve monthly precipitation sums (Section 3.4) or the seven N-day winter maxima (Section 3.6.2) of a given year. A bootstrap sample of size n is then obtained by choosing years randomly with replacement. Every time when a particular year is selected the twelve monthly precipitation sums or the seven N-day winter maxima of that year are added to the bootstrap sample. In this study $B = 500$ bootstrap samples were generated.

B1. Annual averages of monthly standard deviations

In Section 3.4 the annual average $\hat{\theta} = \bar{s}$ of the monthly standard deviations s_m, $m = 1,\ldots,12$ for Stuttgart were considered. For the bth bootstrap sample we can calculate in the same way the standard deviations $s_m^*(b)$, $m = 1,\ldots,12$ of the monthly precipitation sums and their annual average $\hat{\theta}^*(b)$. The latter is used in equation (B1) to obtain the bootstrap standard error of $\bar{s}$.

B2. Spatial averages of relative deviations from the median

In Section 3.6.2 the percentage differences between the sample median of the historical winter maxima and the average sample median of four simulated records were averaged over all stations. This spatial average is affected by the sampling variability of the medians of the historical winter maxima. The

	N = 1	N = 4	N = 10	N = 20
Average of local standard deviations (%)				
Analytical	6.3	5.6	5.0	4.9
Bootstrap	8.1	5.8	4.8	4.6
Area-average percentage difference (%)				
Bootstrap	4.5	3.4	2.8	2.8

Table B1: Standard deviations of the percentage differences between the sample median of the winter maxima and the true median for the seven stations.

variability of the simulated data has much less influence because averaging over four independent estimates reduces the standard deviation of the sample median by a factor of two, and, in contrast to the historical data, the simulated records for the various stations are independent. In order to investigate how far the average percentage differences in the bottom lines of Tables 7 and 8 can be explained by sampling variability, we therefore define the statistics $\hat{\theta}_i = (M_i - \text{med}_i)/\text{med}_i$, where M_i is the sample median of the winter maxima at the ith station, $i = 1,\ldots,7$, and med_i is the true median. The average of the $\hat{\theta}_i$ values is denoted as $\hat{\theta}_\circ$. From the medians $M_i^*(b)$ in the bootstrap samples we calculate $\hat{\theta}_i^*(b) = [M_i^*(b) - M_i]/M_i$ and $\hat{\theta}_\circ^*(b) = \sum_{i=1}^{7} \theta_i^*(b)/7$. The use of the $\theta_i^*(b)$ values in equation (B1) results in a percentage standard deviation of the sample median that is comparable with the analytical results for the Gumbel distribution in Appendix D as shown in Table B1. The bootstrap standard deviation from the $\hat{\theta}_\circ^*(b)$ values in that table is considerably lower due to the spatial averaging of sample medians.

C. The change of the CV-score with the number of nearest neighbours

From equation (11) it is obvious that the expected *CV*-score is determined by the variance of the deletion residuals, $e_t^* = x_t - \hat{x}_t^*$. For the uniform kernel $p_j^* = 1/(k-1)$ and $\hat{x}_t^*$ is given by:

$$\hat{x}_t^* = \frac{1}{k-1}\sum_{j=2}^{k} x_{t(j)} \tag{C1}$$

Because $x_t = x_{t(1)}$, the variance of e_t^* can be written as:

$$\text{var}(e_t^*) = \text{var}(x_{t(1)}) - 2\text{cov}(x_{t(1)}, \hat{x}_t^*) + \text{var}(\hat{x}_t^*) \tag{C2}$$

For ease of exposition we assume that the $x_{t(j)}$ values are equi-correlated random variables with variance σ^2, i.e.:

$$\text{var}(x_{t(j)}) = \sigma^2, \quad j = 1,\ldots,k \tag{C3}$$

$$\text{cov}(x_{t(i)}, x_{t(j)}) = \rho\sigma^2, \quad i \neq j \tag{C4}$$

For days with normal or nearly normal weather conditions, this assumption is quite reasonable for small k. For $\text{cov}(x_{t(1)}, \hat{x}_t^*)$ and $\text{var}(\hat{x}_t^*)$ in equation (C2) we then obtain:

$$\mathrm{cov}(x_{t(1)}, \hat{x}_t^*) = \rho\sigma^2 \tag{C5}$$

$$\mathrm{var}\left(\hat{x}_t^*\right) = \frac{1}{k-1}\sigma^2 + \frac{k-2}{k-1}\rho\sigma^2 \tag{C6}$$

giving

$$\mathrm{var}\left(e_t^*\right) = \sigma^2(1-\rho)\left(1+\frac{1}{k-1}\right) \tag{C7}$$

According to equation (C7), $\mathrm{var}\left(e_t^*\right)$ decreases with increasing k. This must be ascribed to a decrease of $\mathrm{var}\left(\hat{x}_t^*\right)$, which is the only term in equation (C2) that depends on k. Changing k from 5 to 20 leads to a decrease of $\mathrm{var}\left(e_t^*\right)$ of about 20%, which is in reasonable agreement with the observed drop in *CV*-score in Figure 10.

In reality, $\mathrm{cov}\left(x_{t(1)}, \hat{x}_t^*\right)$ decreases with increasing k. For large k this decrease becomes more important than the decrease of $\mathrm{var}\left(\hat{x}_t^*\right)$, resulting in higher values of $\mathrm{var}\left(e_t^*\right)$.

D. Properties of statistics used to compare extreme-value distributions

The largest value $X_{\max}$, the upper quintile mean QM5, and the sample median M are examples of order statistics or simple linear functions of order statistics. The order statistics of a random sample $X_1, X_2, \ldots, X_n$ are obtained by arranging the n variables in ascending order of magnitude and then written as:

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)} = X_{\max} \tag{D1}$$

For the winter maxima in this study $n = 30$. These maxima are assumed to be statistically independent. The $X_{(i)}$ values are then necessarily dependent because of the inequality relations among them. For instance, the second-largest value $X_{(n-1)}$ cannot be extraordinarily large if $X_{\max}$ is not large.

In this appendix we assume that the random variables X_i have a Gumbel distribution with location parameter μ and scale parameter σ ($\sigma > 0$). The distribution function is given by:

$$F(x) = \Pr(X_i \le x) = \exp\left\{-e^{-(x-\mu)/\sigma}\right\}, \qquad -\infty < x < \infty \tag{D2}$$

The mean of X_i depends both on μ and σ:

$$\mathrm{E}(X_i) = \mu + \sigma\gamma \tag{D3}$$

where $\gamma = 0.5772...$ is Euler's constant. The variance of X_i is given by:

$$\mathrm{var}\, X_i = \pi^2\sigma^2 / 6 \tag{D4}$$

The special case $\mu = 0$ and $\sigma = 1$ is known as the standard or reduced variable Y_i. Every Gumbel variable X_i can be written as $X_i = \mu + \sigma Y_i$, and because this transformation is monotonic increasing, it also applies to the order statistics:

$$X_{(i)} = \mu + \sigma Y_{(i)} \tag{D5}$$

where $Y_{(i)}$ is the ith order statistic in a random sample of size n from the reduced variable.

A well-known result for the Gumbel variable is that the maximum X_{max} has also a Gumbel distribution with the same scale parameter σ but with location parameter $\mu + \sigma \ln n$. Equation (D4) therefore also applies to the variance of $X_{max.}$. This result, however, strongly relies on the validity of the Gumbel distribution. Regional analyses of long-term records of N-day maxima in the United Kingdom and the Low Countries (Dupriez and Demarée, 1988; Dales and Reed, 1989; Buishand, 1991) show that for $N = 1$ the upper tail of the distribution tends to be longer than that of the Gumbel distribution, whereas for large N ($N = 10$ or $N = 20$) the upper tail tends to be shorter. The former implies that the variance of X_{max} increases with increasing n and will be underestimated if a Gumbel distribution is assumed. The opposite holds if the distribution has a shorter upper tail than the Gumbel distribution. For the German stations used in this study there are, however, no indications of systematic departures from the Gumbel distribution in the upper tail.

To find the means and variances of QM5 and M, we consider the following linear combination of the order statistics:

$$L = \sum_{i=1}^{n} a_i X_{(i)} \tag{D6}$$

For QM5 we have $a_1 = \cdots = a_{24} = 0$ and $a_{25} = \cdots = a_{30} = 1/6$, and for the sample median $a_{15} = a_{16} = 1/2$, whereas the other a_i values are 0. Substituting equation (D5) in the expression for L, and using the fact that for the two statistics considered here the sum of the coefficients a_i equals 1, we get for the mean of L:

$$\mathrm{E}(L) = \mu + \sigma \sum_{i=1}^{n} a_i \mathrm{E}\left(Y_{(i)}\right), \tag{D7}$$

For the variance of L we have:

$$\mathrm{var}\, L = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{cov}(X_{(i)}, X_{(j)}) = \sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{cov}\left(Y_{(i)}, Y_{(j)}\right) \tag{D8}$$

The means and covariances of the order statistics of the reduced variable can be obtained from tables in Balakrishnan and Chan (1992). For the estimation of var (QM5) and var M, the scale parameter σ in equation (D8) was replaced by its probability-weighted moment estimate.

For the upper quintile mean QM5, equations (D7) and (D8) reduce to:

$$\mathrm{E}(\mathrm{QM5}) = \mu + 2.4840\sigma \tag{D9}$$

$$\mathrm{var}(\mathrm{QM5}) = 0.3130\sigma^2 \tag{D10}$$

The return period T associated with E(QM5) follows from $T = 1 / \Pr\{X_i > \mathrm{E(QM5)}\}$. Using

$$\Pr\{X_i > \mathrm{E(QM5)}\} = 1 - \Pr\{X_i \leq \mathrm{E(QM5)}\} = 1 - \exp\left(-e^{-2.4840}\right) = 0.0800 \qquad \text{(D11)}$$

we obtain $T = 12.5$ years.

For the sample median M, equations (D7) and (D8) reduce to:

$$\mathrm{E}(M) = \mu + 0.3774\sigma \qquad \text{(D12)}$$

$$\mathrm{var}(M) = 0.0671\sigma^2 \qquad \text{(D13)}$$

For the return period associated with E(M) it follows $T = 2.02$ years. Due to the positive bias of the sample median, this return period slightly differs from the value $T = 2$ years for the true median.