Rainfall Generator for the Rhine Basin

Nearest-neighbour resampling of daily circulation indices and conditional generation of weather variables

Jules J. Beersma

T. Adri Buishand

KNMI publication 186–III

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, P.O. Box 17, 8200 AA Lelystad (The Netherlands) Telephone: +31.320.298411; Telefax: +31.320.249218

Contents

Summary

This is the third progress report of a project on the development of a rainfall generator for the Rhine basin. The need for such a rainfall generator arose from the wish to study the likelihood of extreme river discharges in the Netherlands, using a hydrological/hydraulic model. The first progress report dealt with the single-site generation of weather variables by nearest-neighbour resampling for seven stations in the German part of the Rhine basin. In the second progress report a multi-site extension was presented, using daily precipitation and temperature data for 25 stations (1961-1995) in the German part of the basin. The present report deals with a number of relevant issues for long-duration ($\sim$ 1000-years) simulations of precipitation and temperature. Such long-duration simulations are needed as input for the hydrological/hydraulic model.

The nearest-neighbour resampling technique is a simulation method that can easily generate multi-site daily precipitation and temperature data without making restrictive assumptions about the underlying joint distribution of those data. The essence of this 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). In order to generate weather variables for day t, the method needs a feature vector $\mathbf{D}_t$ to find the nearest neighbours in the historical data. In the popular first-order model $\mathbf{D}_t$ contains variables that characterise the weather on day $t-1$. A finite number k of nearest neighbours in terms of a weighted Euclidean distance is selected from the historical record. One of these k nearest neighbours is finally "resampled" using a discrete probability kernel.

The report addresses two major topics. The first topic covers the generation of "synthetic" daily circulation indices needed for long-duration conditional simulations. Time series of synthetic daily circulation indices are simulated using the nearest-neighbour resampling technique. In contrast to nearest-neighbour resampling models for the simulation of precipitation and temperature discussed in the previous progress reports, those for the simulation of circulation indices should be of higher order, i.e. $\mathbf{D}_t$ should not only contain the circulation indices for day $t-1$ but also those for day $t-2$. The best reproduction of the autocorrelation properties and the run lengths of circulation indices is achieved using a second-order model with a relatively small value of k ($k = 5$).

The time series of the circulation indices used for resampling (1961–1995) is not entirely homogeneous. During five of the 35 years the mean sea level pressure fields, from which the circulation indices are derived, show excessive artificial smoothing. The inclusion of these five "oversmoothed" years in the resampling procedure does, however, not affect the model choice.

The second progress report already revealed that unconditional simulations of precipitation and temperature perform slightly better than simulations conditional on historical (1961–1995) circulation indices. The results for the conditional simulation of precipitation and temperature get worse if simulated circulation indices are used instead of the historical ones. In these simulations it makes little difference whether the circulation indices are simulated with a first or a second-order model. At best the underestimation of the extreme value properties of precipitation is as large as 10%. Attempts to remedy this shortcoming remained without success.

The second topic in this report concerns the influence that "natural variations" of the precipitation, temperature and circulation climate may have on the simulated results. Daily precipitation and temperature observations were only available for the period 1961–1995, but daily circulation indices were available for 1881–1995. Conditional simulations are used

to reconstruct precipitation statistics for the period 1891–1995. These simulations explain on average slightly more than 50% of the trends in the mean winter precipitation at five stations for which monthly data during this century were available. Reconstructed extreme winter precipitation for the two 35-year periods 1891–1925 and 1926–1960 is compared with that simulated for 1961–1995. The median and upper quintile mean of the simulated N-day precipitation maxima are for both earlier 35-year periods about 6% smaller than those for 1961–1995. Since the winters in the period 1961–1995 are relatively wet compared to the winters before 1961, long-duration simulations based on precipitation and temperature data for this period should be interpreted with care.

The sensitivity of simulated precipitation to changes in circulation indices is studied by performing three simulations conditional on the 1961–1995 circulation indices, in which in each simulation only one of the three circulation indices is systematically changed. These simulations show that the simulated precipitation is most sensitive to changes in the westerly flow index W, followed by changes in the vorticity index Z. The mean precipitation is typically much more sensitive to systematic changes in W and Z than the precipitation extremes. This is because a large part of the change in the mean precipitation is due to a change in the number of wet days, which has less influence on the extremes.

1 Introduction

1.1 Background

The Rhine, which is the most important river in the Netherlands, flows through several countries (Fig. 1). Large parts of its drainage basin are situated in Switzerland, Germany France and the Netherlands. Protection against flooding is a point of continuous concern. According to safety standards, laid down in the Flood Protection Act, measures against flooding in the non-tidal part of the Rhine in the Netherlands have to withstand a discharge that is exceeded on average once in 1250 years. Traditionally this design discharge has been obtained from a statistical analysis of large river discharges (data from 1901 onwards) at Lobith, where the river enters the country. Several probability distributions have been fitted to the discharge maxima of that record. The long return period requires an extrapolation far 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.

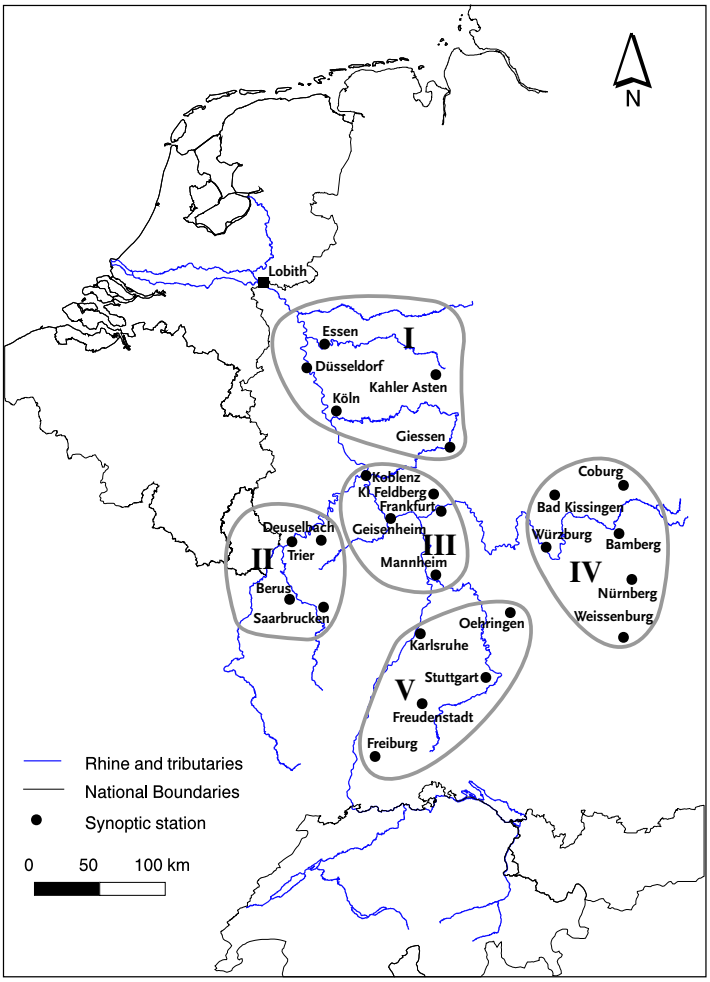

Figure 1: Location of Lobith in the Netherlands and the 25 German stations in the drainage basin of the river Rhine used in this study. A subdivision of the stations into five areas is also shown.

In the most recent re-evaluation of the design discharge at Lobith, there was a strong feeling that the uncertainties of extrapolation could be reduced by taking the physical behaviour of the river basin into account (Delft Hydraulics and EAC-RAND, 1993). For this purpose, it was suggested to develop a hydrological/hydraulic model for the whole basin. With such a model, it would also be possible to quantify the effects of changes in the catchment and the river bed and to predict the potential impacts of climate change. The Institute for Inland Water Management and Waste Water Treatment (RIZA) adopted this idea in a research plan for a new methodology to determine the design discharge (Bennekom and Parmet, 1998). Besides a hydrological/hydraulic model, the development of a stochastic rainfall generator was also planned in order to produce long-duration daily rainfall series over the basin. Unprecedented extreme rainfall situations are likely to occur if the simulation run is considerably longer (300–1000 years) than the observed record. Such unprecedented extreme rainfall events in turn, may lead to more extreme discharges at Lobith than those experienced in the past century. The use of synthetic rainfall data in combination with a hydrological/hydraulic model does not only provide the peak discharges but also the duration of extreme river discharges, which may lead to a better insight into the shape of the design flood.

1.2 Previous research

At the request of RIZA, KNMI carried out a feasibility study (Buishand and Brandsma, 1996). The generation of daily rainfall sequences was reviewed with emphasis on multi-site applications in large catchments and methods were discussed to make use of the influence of the atmospheric circulation on precipitation. The prospects were promising enough to start the development of a rainfall generator for the Rhine basin. In the first instance, the joint simulation of daily precipitation and temperature at single sites was considered, using a non-parametric nearest-neighbour method (Brandsma and Buishand, 1997, 1998; hereafter BB97 and BB98, respectively). The simulation of daily temperature is necessary to account for the effect of snow and frozen soils on large river discharges. The reproduction of the precipitation and temperature autocorrelation coefficients and the N-day maximum rainfall and snowmelt distributions was sufficiently encouraging to proceed with a multi-site extension (Brandsma and Buishand, 1999; further denoted as BB99).

In the hydrological literature the method of nearest-neighbour resampling has been introduced by Rajagopalan and Lall (1995) and Lall and Sharma (1996). Resampling is restricted to nearest neighbours or analogues in the historical record. In the present project the search for nearest neighbours has been based on precipitation and temperature of the last generated day and atmospheric circulation indices. Both unconditional simulation and conditional simulation on atmospheric circulation indices have been considered. In the multi-site study (BB99), the unconditional simulations performed slightly better than the conditional simulations with respect to the reproduction of autocorrelation properties and the distributions of N-day maximum precipitation amounts during the winter season.

An advantage of unconditional simulation is that the length of a simulation run is not restricted by that of the observed record. Unconditional 300-year single-site, 1000-year single-site and 1000-year multi-site simulations are discussed in BB97, BB98 and BB99 respectively. These long-duration simulations contained multi-day precipitation amounts that were much larger than the largest recorded values. In a single-site 1000-year simulation 10-day annual maxima up to 40% larger than the largest historical values were found (BB98). For one particular area, in a multi-site 1000-year simulation, a 10-day precipitation

maximum almost 70% larger than the historical maximum was simulated (BB99).

The conditional simulation of daily preciptation and temperature in earlier phases of the project considered the observed circulation indices for the period 1961–1990 (BB97, BB98) or 1961–1995 (BB99). For long-duration conditional simulations, a separate model which extends the historical record of these circulation indices is required. An important advantage of conditional simulation is its ability to study the effect of past or future changes in the circulation on extreme precipitation. Unconditional simulation by means of nearest-neighbour resampling is, in principle, unable to deal with changes in atmospheric circulation and can therefore only be applied under present climate conditions.

1.3 Scope and objectives

The present report is the third in a series of reports on the development of the rainfall generator. The first objective of this report is the simulation of (long-duration) series of daily circulation indices. Such series of daily circulation indices are required for long-duration simulations of precipitation and temperature conditional on circulation indices. In selecting the nearest neighbours, particular attention will be given to the reproduction of autocorrelation coefficients of the circulation indices and to the reproduction of average run lengths[1].

The second objective is to compare the simulations of precipitation and temperature conditional on the observed circulation indices from the period 1961–1995, which were previously also presented in BB99, with simulated precipitation and temperature conditional on: 1) observed circulation indices from 35-year periods before 1961; 2) systematically changed circulation indices from 1961–1995 (sensitivity experiment); and 3) *simulated* circulation indices for the period 1961–1995. Finally, a few long-duration conditional simulations and a long-duration unconditional simulation are analysed and compared.

Although conditional simulations of precipitation and temperature can in principle deal with anthropogenic climate change this topic is postponed to a future project.

[1] A run is a group of consecutive exceedances of a specified threshold, preceeded and followed by values below that threshold. The corresponding run length is defined as the number of exceedances (days) within the run. For a low threshold the definition of a run can be extended to a group of observations below the threshold.

2 Simulation of daily circulation indices

2.1 Model choice

A typical feature of air pressure is that the day-to-day variability is relatively small during periods of high pressure and generally large during periods of low pressure. Such state dependent behaviour can not be reproduced by classical autoregressive (AR) processes. Al-Awadhi and Jolliffe (1998) therefore studied the use of "threshold autoregressive (TAR) models" to describe time series of surface pressure in the UK. Zwiers and von Storch (1990) have applied this class of models, which they call "regime dependent autoregressive models", to time series of the Southern Oscillation index. It seems reasonable to suspect that the statistical properties of the circulation indices (which are based on air pressure maps) are also state dependent. In contrast to the univariate applications of the TAR models mentioned above we typically have to deal with three indices in case of atmospheric circulation.

Lall and Sharma (1996) showed that nearest-neighbour resampling is able to reproduce the nonlinear behaviour of a TAR model. In a later paper Sharma et al. (1997) demonstrated that this is also true for time series simulation based on non-parametric kernel estimates of the underlying probability densities. In contrast to nearest-neighbour resampling, the latter method is capable to generate values that differ from those in the historical series.

It is, a priori, not clear that threshold autoregressive modelling or kernel density estimations will produce better results than nearest-neighbour resampling. Considerable experience has been obtained with nearest-neighbour resampling in earlier phases of the project which is the main reason to study the suitablilty of this technique for (long-duration) simulations of circulation indices.

2.2 Nearest-neighbour resampling

The application of the method closely follows BB97, BB98 and BB99. For the simulation of circulation indices on day t a feature or state vector $\mathbf{D}_t$ is defined to find analogues in the historical record. This vector contains the values of the circulation indices simulated for day $t-1$ and optionally those for days $t-2$ and $t-3$. The k nearest neighbours (k-NN) are selected from the historical record, using the weighted Euclidean distance, which for two q-dimensional vectors $\mathbf{D}_t$ and $\mathbf{D}_u$ is defined as

$$\delta_{tu} = \left[\sum_{i=1}^{q} w_i (v_{ti} - v_{ui})^2 \right]^{1/2} \tag{1}$$

where v_{ti} and v_{ui} are the ith components of $\mathbf{D}_t$ and $\mathbf{D}_u$ respectively, and the w_i are scaling weights. The vector containing the circulation indices for day t is denoted as $\mathbf{C}_t$.

Let $t'(j),\ j = 1, \ldots, k$ be the times (days) associated with the k-NN, such that the distance between $\mathbf{D}_{t'(j)}$ and $\mathbf{D}_t$ increases with increasing j. One of these k-NN is sampled using a discrete probability distribution or kernel $\{p_j\}$. If the jth nearest neighbour is sampled, then the vector of simulated values $\mathbf{C}_t^*$ is taken as:

$$\mathbf{C}_t^* = \mathbf{C}_{t'(j)}. \tag{2}$$

Lall and Sharma (1996) recommended a decreasing kernel:

$$p_j = \frac{1/j}{\sum_{i=1}^{k} 1/i}, \quad j = 1, \ldots, k. \tag{3}$$

Brandsma and Buishand obtained satisfactory results with decreasing kernels using $k = 20$ in the single-site simulations (BB98). A further study for the multi-site simulations showed that the autocorrelation coefficients and the distributions of N-day maximum precipitation amounts were better reproduced by taking k as small as 5 or even 2 (BB99). Besides decreasing kernels with different k, also kernels in which p_j is inversely proportional to the Euclidean distance $\delta_{tt'(j)}$ are studied in the present report.

Brandsma and Buishand (1997) noted that there is too little autocorrelation in the simulated vorticity index if the circulation indices are simultaneously resampled with precipitation and temperature. In these unconditional simulations the selection of nearest neighbours was based on the simulated vector for day $t-1$, disregarding earlier days. Time series analyses of daily surface pressure (Beersma 1992, Al-Awadhi and Jolliffe 1998) show that first-order AR processes are insufficient to describe the temporal correlation. In order to improve the autocorrelation of simulated circulation indices the effect of including the circulation indices up to day $t-3$ in the feature vector $\mathbf{D}_t$ is studied.

To account for the seasonal variation in the vector of circulation indices the search for the nearest neighbours (analogues) is restricted to days within a specified moving window of width W_{mw} days, centred at the day of interest. The use of a moving window, instead of fixed seasons, prevents sharp transitions between seasons. Brandsma and Buishand (1997) found comparable results for $W_{\mathrm{mw}} = 61$ and $W_{\mathrm{mw}} = 121$. The results in this section were based on $W_{\mathrm{mw}} = 121$ days. In correspondence with BB99 $W_{\mathrm{mw}} = 61$ was used for the precipitation and temperature simulations in section 3.

2.3 Data description

The atmospheric circulation indices used are derived from daily mean sea level pressure (MSLP) data from the UK Meteorological Office on a 5° latitude by 10° longitude grid. For a grid centred at the Rhine basin three circulation indices were calculated: 1) total shear vorticity Z; 2) strength of the westerly flow W; and 3) strength of the southerly flow S (for details see Jones et al. 1993). These three indices form the elements of the circulation vector $\mathbf{C}$. Complete calendar years were obtained for the period 1881–1995. For the 35-year period 1961–1995 daily precipitation and temperature data were made available by the Deutscher Wetterdienst (DWD) for 25 stations in the German part of the Rhine basin.

It is known that there are discontinuities in the daily MSLP data set, partly because of changes in the procedure used for gridding MSLP from surface analyses. As a result the data are smoother (both in the spatial and temporal domain) in some periods than in other periods, in particular over the North Atlantic Ocean. In a study of the number of gales over the British Isles, Hulme and Jones (1991) found that the data are particularly smooth in the period 1960–1965. The grid used here is shifted 10° east and 5° south compared to theirs and therefore contains less data over the ocean. Nevertheless the pressure fields are too smooth over the study area during the period 1961–1965. The effect of smoothing on the estimated number of gales is, however, somewhat smaller than in Hulme and Jones (1991). The effects of the use of these "oversmoothed" data in the resampling procedure will be discussed in section 2.5.

Before resampling, the data were standardised by subtracting the calendar day's mean m_d and dividing by the calendar day's sample standard deviation s_d:

$$\tilde{x}_t = (x_t - m_d)/s_d, \quad t = 1, \ldots, 365J;\ d = (t-1) \bmod 365 + 1 \tag{4}$$

where x_t and $\tilde{x}_t$ are the original and the standardised variable, respectively, for day t, and

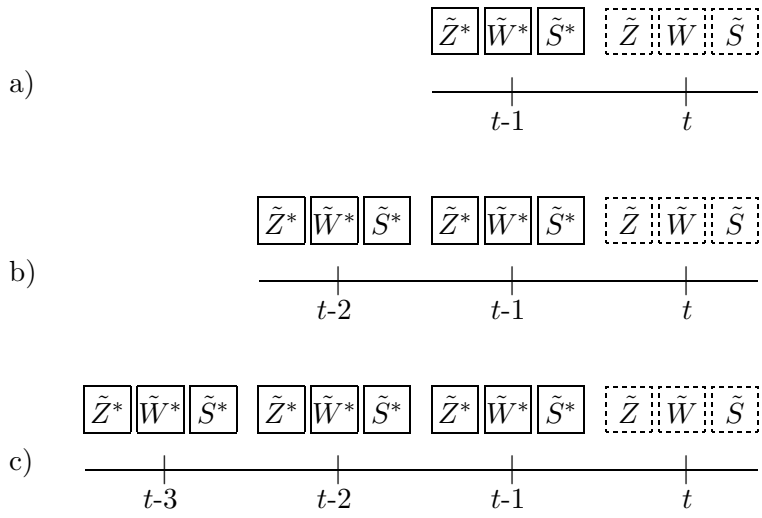

Figure 2: Elements of the feature vector (solid boxes) for unconditional simulation of circulation indices Z, W, S (dashed boxes). An asterisk indicates that a value was resampled in a previous time step and a tilde denotes standardised values. First-order a); second-order b); and third-order models c).

J is the total number of years in the record. Here smoothed values of m_d and s_d were used based on Friedman's supersmoother (Härdle 1990). Before calculating the smooths the values for $d = 336, \ldots, 365$ were inserted for $d < 1$ and the values for $d = 1, \ldots, 30$ were inserted for $d > 365$ to harmonize the smoothed values at changes of years.

2.4 Model construction

Various resampling models are presented in this section. Differences between models are the result of variations in the composition of the feature vector, the scaling weights or the resampling kernel. The feature vectors are schematically presented in Fig. 2. An overview of the different models is given in Table 1.

In the first model (circ1.k) the feature vector contains the three circulation indices on day $t-1$ with equal weights w_i. This model is therefore called a first-order model. Resampling with this model is based on a decreasing kernel with $k = 20$ (circ1.20) or $k = 2$ (circ1.2).

Five second-order models are presented. In these models the feature vector does not only contain the standardised circulation indices for day $t-1$ but also those for day $t-2$. The weights for $\tilde{Z}_{t-1}$, $\tilde{W}_{t-1}$, $\tilde{S}_{t-1}$, $\tilde{Z}_{t-2}$, $\tilde{W}_{t-2}$ and $\tilde{S}_{t-2}$ are 1,2,1,1,0 and 1 respectively. The weight for $\tilde{W}_{t-2}$ is taken to be zero here because the autocorrelation structure of the W index closely resembles that of a first-order AR process. Three of the second-order models (circ2.20, circ2.5 and circ2.2) have a decreasing kernel with $k = 20, 5$ and 2 respectively. In the fourth second-order model (circ2.var) the distances between the feature vector and its nearest neighbours are included in the discrete kernel for generating a new value. The probability p_j is inversely proportional to the volume V_j, of a q-dimensional ball with radius $\delta_{tt'(j)}$ as suggested by Lall and Sharma (1996), where $\delta_{tt'(j)}$ is the weighted Euclidean distance between the feature vector for the day of interest and its jth nearest neighbour, and q is the dimension of these vectors:

$$p_j = \frac{1/V_j}{\sum_{i=1}^{k} 1/V_i}, \quad j = 1, \ldots, k. \tag{5}$$

In contrast to the fixed decreasing kernel, a separate kernel is calculated for each day. There is often a singularity at $j = 1$ due to the possibility that $\mathbf{D}_{t'(1)} = \mathbf{D}_t$ (in a first-order model this occurs every day), giving $\delta_{tt'(1)} = 0$ and $V_1 = 0$. This singularity is circumvented by setting the volume V_1 equal to $V_1 = V_2(V_3 - V_2)/V_3$. Note, that V_1 can take values between 0 and V_2. Alternative variable kernels did not produce better results.

The last of the second-order models (circ2.5 fp) resembles the model with $k = 5$ but for finding the nearest neighbours we also compare the Euclidean distance between the previously simulated vector $\mathbf{C}^*_{t-1}$ and the new vector $\mathbf{C}^*_t = \mathbf{C}_{t'(j)}$ with the Euclidean distance between this new vector and its historical predecessor $\mathbf{C}_{t'(j)-1}$. The absolute value of the difference of these two distances:

$$| \, \|\mathbf{C}^*_{t-1} - \mathbf{C}^*_t\| - \|\mathbf{C}_{t'(j)-1} - \mathbf{C}_{t'(j)}\| \, | \tag{6}$$

is added to $\delta_{tt'(j)}$ with a weight of 0.02. Compared to the weights of the feature vector elements this weight is relatively low because the average Euclidean distance between two consecutive days, in the historical and in the simulated series, is generally much larger than that between two analogues. Equation 6 compares the forward persistence (fp) in the simulated series with that in the historical series and therefore this model is referred to as second-order fp model (circ2.5 fp).

Finally, there are two third-order models (circ3.5 and circ3.2 with $k = 5$ and $k = 2$ respectively) for which $\tilde{Z}_{t-3}$ and $\tilde{S}_{t-3}$ are included in the feature vector both with unit weight.

2.5 Model results

To make an objective comparison between different choices of $\mathbf{D}_t$ and p_j some relevant characteristics of the simulated circulation indices are compared with their historical values. First, the reproduction of their autocorrelation coefficients is studied. Then the duration or persistence of particular circulation types is examined. The latter is usually expressed in terms of run lengths, i.e. the number of consecutive days that a particular circulation index exceeds a prescribed threshold without interruption. For high thresholds the average run length is a measure of the mean duration of extreme events. A consequence of resampling

Table 1: Definition of models for unconditional simulation of circulation indices (Z, W, S). An asterisk indicates that a value was resampled in a previous time step; the tilde refers to standardised values.

Model	$\mathbf{D}_t$	w_i
First-order		
circ1.k	$(\tilde{Z}^*_{t-1},\tilde{W}^*_{t-1},\tilde{S}^*_{t-1})$	(1,1,1)
Second-order		
circ2.k		
circ2.var	$(\tilde{Z}^*_{t-1},\tilde{W}^*_{t-1},\tilde{S}^*_{t-1},\ \tilde{Z}^*_{t-2},\tilde{W}^*_{t-2},\tilde{S}^*_{t-2})$	(1,2,1,1,0,1)
circ2.5 fp		
Third-order		
circ3.k	$(\tilde{Z}^*_{t-1},\tilde{W}^*_{t-1},\tilde{S}^*_{t-1},\ \tilde{Z}^*_{t-2},\tilde{W}^*_{t-2},\tilde{S}^*_{t-2},\ \tilde{Z}^*_{t-3},\tilde{W}^*_{t-3},\tilde{S}^*_{t-3})$	(1,3,1,1,0,1,1,0,1)

Table 2: Differences between the lag 1 and lag 3 autocorrelation coefficients of daily circulation indices in 350-year simulations and the historical record (1961–1995). The bottom lines give the values of $r(1)$ and $r(3)$ in the historical record with their standard errors. Estimates in italics differ more than 2 × std. err. from the historical values. Z, W and S denote the vorticity, the west component and the south component of the flow respectively.

	$r(1)$			$r(3)$		
Model	Z	W	S	Z	W	S
circ1.20	-0.015	*-0.016*	-0.013	*-0.111*	0.009	*-0.107*
circ1.2	-0.010	-0.007	-0.009	*-0.062*	0.005	*-0.060*
circ2.20	*-0.036*	*-0.021*	*-0.046*	*-0.028*	0.012	*-0.027*
circ2.5	-0.014	-0.010	-0.022	-0.015	0.012	-0.014
circ2.2	-0.013	-0.003	-0.007	-0.012	0.012	-0.012
circ2.var	-0.018	-0.006	-0.014	-0.019	0.011	-0.014
circ2.5 fp	-0.007	-0.004	-0.011	-0.015	*0.024*	-0.017
circ3.5	*-0.031*	-0.011	*-0.031*	*-0.022*	0.013	-0.022
circ3.2	-0.010	-0.002	-0.013	*-0.022*	0.005	-0.017
Historical record	0.498	0.755	0.521	0.182	0.393	0.190
std. err.	0.012	0.006	0.011	0.010	0.012	0.011

is that some historical days will not be resampled at all while others will be resampled quite frequently. The influence of the model choice on the strength of this selection effect is also investigated. The results presented in this section are based on 350-year simulations in which the circulation indices were resampled from the period 1961–1995.

2.5.1 Autocorrelation of circulation indices

Table 2 presents the differences between the lag 1 and lag 3 autocorrelation coefficients of the simulated circulation indices and those of the historical record (1961–1995). The historical estimates and their standard errors are also given. The standard errors were obtained with the jackknife method in Buishand and Beersma (1993). When the autocorrelation coefficients of the simulated data deviate more than twice the standard error from the historical data they are referred to as being significantly different from the historical estimates.

The lag 1 autocorrelation coefficient of the three indices is underestimated in both first-order models. This underestimation is significant only for the W index of the model with $k = 20$. The lag 3 autocorrelation coefficients of the Z and S indices are significantly underestimated. The largest underestimation is found for $k = 20$. The lag 3 autocorrelation coefficient of the W index, finally, is slightly overestimated in both models. Note that for small k quite large parts of the historical record are resampled. From Eq. (A12) in BB99 it follows that the expected largest historical part in a 350-year simulation is 28.2 days for $k = 2$ and 9.9 days for $k = 20$.

Three second-order models using the standard decreasing kernel are presented in Table 2. The second-order model with $k = 20$, significantly underestimates all three lag 1 autocorrelation coefficients. Again the lag 3 autocorrelation coefficients of the Z and S

Table 3: Historical lag 1 and lag 3 autocorrelation coefficients of daily circulation indices and their standard errors for different historical periods. Estimates in italics differ more than 2 × std. err. from the 1961–1965 values. Z, W and S denote the vorticity, the west component and the south component of the flow respectively.

	$r(1)$			$r(3)$		
Period	Z	W	S	Z	W	S
1891–1925	*0.507*	0.761	0.563	0.174	0.391	*0.234*
std. err.	0.009	0.005	0.010	0.010	0.013	0.014
1926–1960	*0.493*	0.758	0.564	0.187	0.378	*0.233*
std. err.	0.012	0.007	0.010	0.016	0.013	0.012
1961–1995	*0.498*	0.755	*0.521*	0.182	0.393	0.190
std. err.	0.012	0.006	0.011	0.010	0.012	0.011
1966–1995	*0.484*	0.749	*0.510*	0.175	0.387	0.188
std. err.	0.012	0.005	0.010	0.011	0.012	0.012
1961–1965	0.612	0.778	0.606	0.227	0.398	0.180
std. err.	0.033	0.017	0.030	0.032	0.033	0.024

indices are significantly underestimated but the differences are much smaller than for the first-order model. For the two models with $k = 5$ and $k = 2$ none of the autocorrelation coefficients differs significantly from the historical ones. The model with $k = 2$ seems to perform slightly better. The performance of the second-order model with variable kernel (circ2.var) is comparable to that of the second-order model with a decreasing kernel with $k = 5$ (circ2.5). In the second-order forward persistence model (circ2.5 fp) the systematic errors in the lag 1 autocorrelation coefficients are somewhat further reduced compared to those for the other second-order models. On the other hand, the lag 3 autocorrelation coefficient of the W index is significantly overestimated in that model.

The third-order models are no improvement compared to the second-order models. In fact, some of the autocorrelation coefficients, in particular those for the Z index (vorticity), are significantly different from the historical ones.

In Table 3 the autocorrelation coefficients for the period 1961–1965 are compared with the three 35-year periods used in this study and those for the 30-year period 1966–1995. The only case in which the 1961–1965 period significantly differs (difference larger than twice the 1961–1965 standard error) from the other periods is for the lag 1 autocorrelation coefficient of the Z index. The larger autocorrelation is in agreement with the oversmoothed data in this period.

With two resampling models (circ1.20 and circ2.5) 350-year simulations were performed based on the period 1966–1995. This period does not contain any oversmoothed data. Table 4 compares the autocorrelation coefficients of the simulations with those for the period from which is resampled (1966–1995). The results are very similar to those of the same models for the period 1961–1995 presented in Table 2. From this it may be concluded that the inclusion of the oversmoothed data (1961–1965) in the resampling procedure does not influence the conclusions concerning the model choice.

Table 4: Differences between the lag 1 and lag 3 autocorrelation coefficients of daily circulation indices in 350-year simulations and the historical record (1966–1995). The bottom lines give the values of $r(1)$ and $r(3)$ in the historical record with their standard errors. Estimates in italics differ more than 2 × std. err. from the historical values. Z, W and S denote the vorticity, the west component and the south component of the flow respectively.

	$r(1)$			$r(3)$		
Model	Z	W	S	Z	W	S
circ1.20	-0.019	*-0.018*	-0.019	*-0.118*	0.002	*-0.117*
circ2.5	-0.016	*-0.010*	-0.017	-0.017	0.017	-0.010
1966–1995	0.484	0.749	0.510	0.175	0.387	0.188
std. err.	0.012	0.005	0.010	0.011	0.012	0.012

2.5.2 Run lengths of circulation indices

The run length refers to the number of consecutive days that a particular circulation index is continuously above or below a certain threshold. The thresholds that are considered in the first instance are the average of a circulation index plus one standard deviation and the average of that circulation index minus one standard deviation. Since the data are standardised the thresholds correspond to 1.0 and -1.0 respectively. For the standardised indices of Z, W and S, values larger than 1.0 correspond to cyclonic, strong westerly and southerly flow respectively while values smaller than -1.0 correspond to anticyclonic, easterly and northerly flow. Table 5 presents the average run lengths of the historical circulation indices for 1961–1995 as well as the relative differences between the average run lengths of the simulated circulation indices and those of the historical indices.

The first-order models underestimate the average run lengths, in particular those for strong westerly flows ($\tilde{W} > 1.0$) and easterly flows ($\tilde{W} < -1.0$). In the second-order model with $k = 20$ all average run lengths are significantly underestimated (i.e. the differences are larger than twice the standard error of the average run length in the historical record). In general, the reproduction of the mean run lengths improves with decreasing k. The reproduction of run lengths of the second-order forward persistence model (circ2.5 fp) is comparable to that of the second-order model with variable kernel (circ2.var). Their performance lies between that of the second-order models with $k = 5$ and $k = 2$. The third-order model with $k = 5$ significantly underestimates the average run lengths for the cyclonic ($\tilde{Z} > 1.0$) and anticyclonic ($\tilde{Z} < -1.0$) flows as well as the southerly ($\tilde{S} > 1.0$) and northerly ($\tilde{S} < -1.0$) flows.

Note that in the historical record the average run length for easterly flows ($\tilde{W} < -1.0$) is about 25% larger than for strong westerly flows ($\tilde{W} > 1.0$). All models are able to reproduce this asymmetry between the average duration of strong westerly and easterly flows.

For each circulation component also the average durations corresponding to the 2.0 and -2.0 thresholds are considered. In only 2.5% of the cases the values are > 2.0, respectively < -2.0. Table 6 presents the relative differences between the average run lengths of the simulated circulation indices and those of the 1961–1995 indices. The results for these larger thresholds are comparable to those in Table 5. Although the same models show the largest differences compared to the observations, the number of significant differences is smaller.

Table 5: Relative differences (%) between average run lengths (see text for details) of circulation indices in 350-year simulations and the historical record (1961–1995) for the 1.0 and -1.0 thresholds. The bottom lines give the average run lengths (days) in the historical record with their relative standard errors (%). Estimates in italics differ more than 2 × std. err. from the historical values. Z, W and S denote the vorticity, the west component and the south component of the flow respectively.

	Standardised index > 1.0			Standardised index < -1.0		
Model	Z	W	S	Z	W	S
circ1.20	-1.78	-3.96	-0.61	-1.35	-4.70	-3.12
circ1.2	-0.32	-0.89	-0.26	-1.27	-2.00	-1.76
circ2.20	*-4.52*	*-5.26*	*-4.84*	*-4.68*	*-5.82*	*-7.80*
circ2.5	-2.67	-3.63	-3.22	-2.26	-3.60	-4.00
circ2.2	2.72	-1.68	-1.73	-2.53	-0.63	-0.94
circ2.var	-3.53	-2.68	-1.65	-2.00	-1.32	-3.13
circ2.5 fp	-1.83	-3.15	-2.24	-2.34	-1.80	-3.01
circ3.5	*-4.45*	-4.28	*-3.85*	*-4.46*	-2.40	*-5.89*
circ3.2	-1.23	-1.47	-1.25	-2.51	-0.83	-2.92
Historical record	1.74	2.14	1.69	1.62	2.70	1.81
std. err. (%)	1.97	2.32	1.83	1.74	2.86	2.10

This is mainly due to the larger relative standard error of the average run length.

2.5.3 Selection effects

As a result of the resampling procedure some days from the historical record may not be present in the resampled series while other days will be resampled more frequently than expected. Here we will have a closer look into this so called selection effect.

Assume we have a random series $X_1, X_2, \ldots, X_n$. The probability that X_i does not occur in a sample of n independent drawings with replacement (the standard bootstrap) is

$$P_n = (1 - 1/n)^n = e^{n\ln(1-1/n)} \approx e^{-1} \approx 0.368. \tag{7}$$

Similarly we find for Ln independent drawings that

$$P_{Ln} \approx e^{-L}. \tag{8}$$

Let us now consider the number K_i of times that X_i occurs in Ln drawings. The probability that $K_i = r$ is given by:

$$\Pr(K_i = r) = \left(\begin{array}{c} Ln \\ r \end{array} \right) (1/n)^r (1 - 1/n)^{Ln-r}. \tag{9}$$

This binomial distribution can be approximated by a Poisson distribution with parameter L:

$$\Pr(K_i = r) = \frac{L^r e^{-L}}{r!}. \tag{10}$$

Table 6: Relative differences (%) between average run lengths (see text for details) of circulation indices in 350-year simulations and the historical record (1961–1995) for the 2.0 and -2.0 thresholds. The bottom lines give the average run lengths (days) in the historical record with their relative standard errors (%). Estimates in italics differ more than 2 × std. err. from the historical values. Z, W and S denote the vorticity, the west component and the south component of the flow respectively.

	Standardised index > 2.0			Standardised index < -2.0		
Model	Z	W	S	Z	W	S
circ1.20	-1.44	-3.77	-2.27	-0.63	-7.01	-3.08
circ1.2	0.02	-2.44	-0.68	1.28	2.50	0.53
circ2.20	*-7.12*	-4.98	-4.03	-1.91	-8.02	*-6.37*
circ2.5	-2.77	-1.93	-2.39	-0.65	-7.00	-3.35
circ2.2	-1.86	-2.10	-0.49	0.82	-1.36	-2.21
circ2.var	-3.91	-3.87	-1.14	-1.59	-2.91	-4.05
circ2.5 fp	-4.18	-2.89	-1.19	-1.46	-4.85	-3.53
circ3.5	-4.94	-3.39	-3.86	-1.67	-3.91	-3.05
circ3.2	-2.64	-2.63	-1.88	-1.83	-1.77	-0.81
Historical record	1.34	1.23	1.17	1.05	1.44	1.24
std. err.(%)	2.68	2.90	2.45	2.18	4.12	2.69

For $r = 0$ this equation reduces to Eq. (8). Equations (8) and (10) remain valid if resampling is restricted to the data within a moving window of width W_{mw}. The expected number of X_i's drawn r times in Ln drawings equals $n\Pr(K_i = r)$. For nearest-neighbour resampling the number of historical days that is drawn r times can be compared with the number expected from the Poisson distribution with parameter L.

When the frequency distribution of the number of historical days in a simulation run becomes wider compared to the one for independent resampling (typically when the number of days that is never or seldomly resampled increases), it becomes more likely that certain important characteristics of the historical data are not well reproduced. In Table 7 the frequency distributions of the 350-year simulations are compared with the values for the Poisson distribution with $L = 10$. The frequency distributions from the simulations are wider than the theoretical frequency distribution for independent resampling with replacement, because resampling is restricted to nearest-neighbours. The table shows that the selection effect becomes somewhat stronger (wider frequency distribution) with increasing order of the resampling model. The kernel size k has only little influence on the width of the frequency distribution. Remember, however, that k has a considerable influence on the expected largest historical part in a 350-year simulation (28.2, 14.7 and 9.9 days for $k = 2$, $k = 5$ and $k = 20$ respectively).

Table 7: Number of historical days drawn r times in 350-year simulations compared with the number expected for the standard bootstrap. The largest number of times that a historical day is drawn is given in the last column.

Model	r							$r_{\max}$
	0	1-5	6-10	11-15	16-20	21-25	>25	
bootstrap	1	856	6591	4704	602	20	<1	
circ1.20	9	1201	6200	4493	813	57	2	27
circ1.2	11	1741	5703	4039	1066	196	19	33
circ2.20	49	1954	5120	4284	1240	120	8	29
circ2.5	15	1714	5620	4165	1114	139	8	28
circ2.2	21	1753	5665	3991	1137	182	26	33
circ2.var	22	1728	5571	4203	1093	142	16	37
circ2.5 fp	26	1745	5590	4079	1148	169	18	30
circ3.5	65	2415	4810	3727	1358	337	63	36
circ3.2	40	2080	5358	3733	1252	266	46	38

3 Conditional resampling of precipitation and temperature

The previous section was devoted to the simulation of circulation indices. In this section simulations of precipitation P and temperature T conditional on the atmospheric circulation are discussed. The resampling models from BB99 as well as some alternatives are used. Much attention is given to the influence of the atmospheric circulation on the simulated precipitation and temperature characteristics. This influence is investigated by: 1) comparing the precipitation statistics from simulations that are based on different 35-year periods of historical circulation indices; and 2) performing sensitivity simulations with systematic changes in the 1961–1995 circulation indices. In addition, 35-year as well as long-duration simulations are presented conditional on the *simulated* circulation indices obtained with the resampling models described in the previous section.

3.1 Simulations conditional on the historical circulation

3.1.1 Summary statistics for P and T

Daily P and T observations were available for 25 German stations for the period 1961–1995 (Fig. 1). To keep the dimension of the feature vector low, a small number of summary statistics of P and T were selected as elements of this vector. Both for P and T the arithmetic mean of the standardised daily values of the 25 stations was considered. The T data were standardised, similar to the circulation indices, by subtracting the calendar day's mean and dividing by the calendar day's standard deviation, according to Eq. (4). To avoid negative standardised P values, the precipitation data were standardised by dividing by the calendar day's mean precipitation $m_{d,\mathrm{wet}}$ for wet days:

$$\tilde{x}_t = x_t / m_{d,\mathrm{wet}}. \tag{11}$$

The values of $m_{d,\mathrm{wet}}$ were smoothed using Friedman's supersmoother (Härdle 1990). After resampling, the final simulated values of T and P were obtained from the standardised resampled values by inverting, respectively Eq. (4) and (11).

The fraction F of stations with precipitation ($P > 0.2$ mm) was used as an additional summary statistic. F helps to distinguish between large-scale and convective precipitation. The feature vector further contained standardised circulation indices.

3.1.2 Simulated precipitation and temperature for 1961–1995

Initially, a large number of test cases was compared in BB99 for the simulation of P and T using a decreasing kernel with $k = 20$. Regarding the choice of k they explored two cases further: their case 4.4 for unconditional simulation and their case 4.1 for conditional simulation. Their simulations with $k = 5$ (and for conditional simulation also $k = 2$) are repeated here. In an attempt to improve the reproduction of the autocorrelation structure of daily precipitation, several new cases for conditional simulation are also investigated. The model details are given in Table 8. The composition of some of the feature vectors is presented schematically in Fig. 3.

Table 9 shows for the winter (October–March) the deviations of some relevant statistics from the historical (1961–1995) values averaged over the 25 stations for the unconditional model (case 4.1 with $k = 5$) and several conditional models. Besides the reproduction of mean values, Table 9 also presents that of the standard deviation s of daily values, the lag 1 autocorrelation coefficient $r(1)$, the standard deviation s_m of monthly values, and three

Table 8: Definition of cases for unconditional and conditional simulation. The w_i values for the circulation apply to all three components of $\tilde{\mathbf{C}} = (\tilde{Z}, \tilde{W}, \tilde{S})$. $\tilde{P}$ and $\tilde{T}$ denote respectively the standardised precipitation and standardised temperature averaged over the 25 German stations, and F denotes the fraction of these stations with $P > 0.2$ mm. An asterisk indicates that a value was resampled in a previous time step.

Case	$\mathbf{D}_t$	w_i
unconditional		
4.4	$(\tilde{\mathbf{C}}^*_{t-1},\, \tilde{P}^*_{t-1},\, F^*_{t-1},\, \tilde{T}^*_{t-1})$	(1,2,2,2)
conditional		
4.1	$(\tilde{\mathbf{C}}_t,\, \tilde{P}^*_{t-1},\, F^*_{t-1},\, \tilde{T}^*_{t-1})$	(1,1,1,1)
6.1	$(\tilde{\mathbf{C}}_t,\, \tilde{\mathbf{C}}_{t-1},\, \tilde{P}^*_{t-1},\, F^*_{t-1},\, \tilde{T}^*_{t-1})$	(1,1,1,1,1)
6.2	$(\tilde{\mathbf{C}}_t,\, \tilde{\mathbf{C}}_{t-1},\, \tilde{P}^*_{t-1},\, F^*_{t-1},\, \tilde{T}^*_{t-1})$	(1,1,2,2,2)
7.1	$(\tilde{\mathbf{C}}_t,\, \tilde{P}^*_{t-1},\, F^*_{t-1},\, \tilde{T}^*_{t-1},\, \tilde{P}^*_{t-2})$	(1,1,1,1,1)
7.2	$(\tilde{\mathbf{C}}_t,\, \tilde{P}^*_{t-1},\, F^*_{t-1},\, \tilde{T}^*_{t-1},\, \tilde{P}^*_{t-2},\, F^*_{t-2},\, \tilde{T}^*_{t-2})$	(1,1,1,1,1,1,1)
8.1	as case 4.1 with fp P and F terms and weights (1,1)	
8.2	as case 4.1 with fp P and T terms and weights (1,1)	
8.3	as case 4.1 with fp P and T terms and weights (0.1,0.1)	

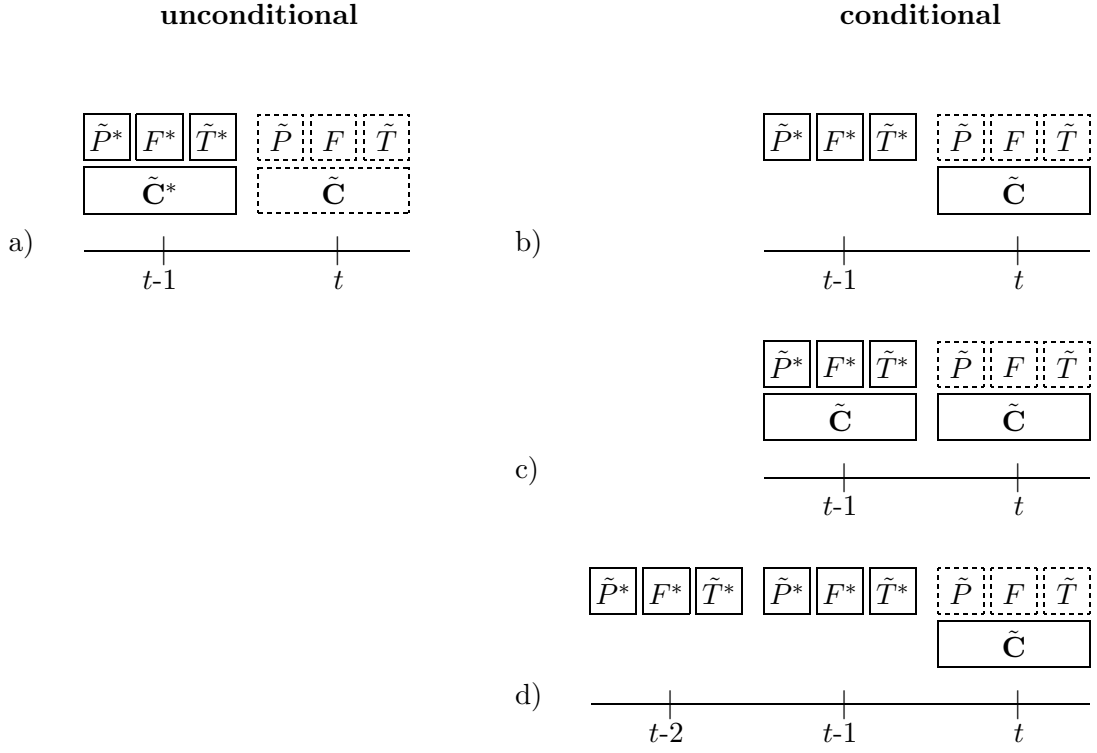

Figure 3: Elements of the feature vector (solid boxes) for unconditional simulation a (case 4.4); and conditional simulations b (case 4.1); c (cases 6.1 and 6.2); and d (case 7.2) of new variables (dashed boxes). The asterisks indicate that the corresponding variables are resampled values of the previous time step, and the tilde refers to standardised values.

Table 9: Performance of the 1961–1995 simulations (ten runs of 35 years for each case) for the winter (October–March). For each characteristic the differences (mean temperature in °C and lag 1 autocorrelation coefficient) or percentage differences (other characteristics) are given between the simulated and historical data (1961–1995), averaged over 25 stations. The same units apply to the standard errors from the historical data in the bottom line. Estimates in italics differ more than 2 × std. err. from the historical values.

Case	mean		s		$r(1)$		s_m		10-day P		
	P	T	P	T	P	T	P	T	Max	QM5	Med
unconditional											
4.4 ($k=5$)[1]	-3.6	0.12	-2.5	-2.0	*-0.021*	*-0.032*	-3.5	-9.2	-4.8	-4.1	-2.4
4.4 ($k=5$)[1]	-4.4	0.08	-1.9	-2.4	-0.016	*-0.035*	-2.1	-10.4	1.1	-1.2	-2.9
4.4 ($k=5$)[2]	-2.1	0.12	-0.5	-3.3	-0.013	*-0.041*	-3.8	*-13.2*	-1.7	-1.4	-0.3
conditional											
4.1 ($k=5$)[1]	-0.3	0.19	-1.1	*-6.9*	*-0.046*	*-0.090*	-3.8	*-19.2*	1.6	-1.8	-3.3
4.1 ($k=5$)[1]	-2.3	0.14	-2.9	*-6.0*	*-0.057*	*-0.084*	-4.8	*-17.0*	-3.8	-4.9	-4.9
4.1 ($k=5$)[2]	-1.6	0.12	-1.5	*-5.6*	*-0.053*	*-0.082*	-3.9	*-17.2*	-0.6	-2.9	-3.9
6.1 ($k=5$)	-5.0	0.16	-4.2	*-5.6*	*-0.051*	*-0.101*	-7.2	*-17.5*	-6.2	-7.0	-6.7
6.2 ($k=5$)	-4.1	0.22	-3.1	*-5.6*	*-0.041*	*-0.082*	-6.4	*-15.4*	-6.1	-6.5	-4.7
7.1 ($k=5$)	-2.3	0.21	-3.5	*-7.1*	*-0.058*	*-0.094*	-4.8	*-20.2*	-4.6	-5.6	-5.4
7.2 ($k=5$)	-3.7	0.20	-3.2	*-7.4*	*-0.062*	*-0.102*	-5.3	*-18.0*	-4.5	-5.5	-5.3
8.1 ($k=5$)	*-22.7*	0.19	*-20.8*	*-7.1*	*0.039*	*-0.083*	*-15.8*	*-17.9*	-8.4	-13.4	-19.3
8.2 ($k=5$)	*-24.1*	0.23	*-22.2*	*-13.9*	*0.026*	*-0.043*	*-19.6*	*-15.4*	-14.5	-18.0	-20.7
8.3 ($k=5$)	-7.5	0.18	*-7.9*	*-8.0*	*-0.035*	*-0.075*	*-10.2*	*-18.1*	-2.4	-6.6	-8.8
4.1 ($k=2$)	-0.6	0.11	-1.6	-4.6	*-0.044*	*-0.063*	-3.5	*-13.8*	-1.8	-3.3	-4.2
4.1 {0.8,0.2}	-1.9	0.07	-2.3	-3.4	*-0.036*	*-0.048*	-3.5	-9.9	-1.0	-3.5	-3.2
std. err.	3.8	0.17	2.6	2.5	0.009	0.007	4.8	6.2	–	–	–

[1] simulations with different random number seeds
[2] simulations from BB99

properties of 10-day maximum precipitation, namely the largest value (Max), the upper quintile mean (QM5), and the median value (Med). It should be noted that the success of reproducing s_m and the extreme-value properties depends on the reproduction of s and $r(1)$. As for the standard errors of $r(1)$, the standard errors of s and s_m in Table 9 were based on a jackknife procedure (see Appendix).

First, the unconditional case 4.4 and conditional case 4.1 both with $k = 5$ from BB99 are compared. The unconditional model shows a smaller bias of the lag 1 autocorrelation coefficients of both daily precipitation and temperature, and a smaller underestimation of the daily standard deviations of temperature. As a result the unconditional model slightly better reproduces the monthly standard deviations of P and T and the extreme-value properties. Note that the results for the unconditional case 4.4 and conditional case 4.1 in Table 9 differ somewhat from those in BB99 due to a different random number seed. The simulations for case 4.1 were repeated using the data for the period 1966–1995 only. Like in section 2.5.1 removal of the years with oversmoothed data had little effect on the differences between the

observed and simulated values.

The results of the new conditional simulations, are discussed next. With the inclusion of $\mathbf{C}_{t-1}$ in the feature vector $\mathbf{D}_t$ (cases 6.1 and 6.2) the autocorrelation of P and T in the simulated data remains more or less the same as in case 4.1 but the daily standard deviation of precipitation is smaller which has a negative effect on the reproduction of the monthly standard deviation and the 10-day precipitation maxima. Stronger persistence of precipitation is expected with the inclusion of values at $t-2$ (cases 7.1 and 7.2) in the feature vector. The results show, however, that second order P, F and T terms in the model slightly deteriorate the reproduction of the lag 1 autocorrelation coefficient and the standard deviations of the daily and monthly values. The lag 1 autocorrelation can be enhanced by considering the next value that will be resampled in the search for nearest neighbours. For unconditional simulation of circulation indices such a model was already introduced in section 2.4 and referred to as forward persistence (fp) model. Case 8.1 represents a conditional simulation based on case 4.1 but with fp terms in both P and F. This simulation shows indeed an enhancement of the precipitation lag 1 autocorrelation. However, it is overestimated with about the same amount as it is underestimated in the original model (conditional case 4.1). The model also underestimates the average winter precipitation by more than 20% and the underestimation of the standard deviation of the daily and monthly precipitation values is of the same order. This turns out to be caused by a strong selection effect (see section 3.3). A second forward persistence model with fp terms in P and T (case 8.2) shows similar behaviour. When the weights for the fp terms in P and T are considerably reduced (case 8.3), the systematic underestimations reduce as well, but the lag 1 precipitation autocorrelation is again significantly underestimated.

A slight improvement of conditional case 4.1 can be obtained by reducing the size k of the kernel $\{p_j\}$. For $k = 2$ there is some improvement of the reproduction of the temperature statistics but not for precipitation. A final simulation using a kernel with probabilities $p_1 = 0.8$ and $p_2 = 0.2$ gives sligthly better results, comparable to those of the unconditional model. It should be noted, however, that for kernels with a high probability p_1 (small k) the number of times that the resampled values of P and T for the day of interest equal the observed values for that day becomes large as well. The resampled time series becomes thus more similar to the historical time series for larger values of p_1. For the simulation with $k = 2$ ($p_1 = 0.67$) and the one with $p_1 = 0.8$, the percentage of days in the resampled series that is unchanged compared to the historical series is respectively 32% and 47%.

3.1.3 Reconstructed and observed trends in winter precipitation

Several authors have reported on precipitation trends during the past 100 years in Europe. Rapp and Schönwiese (1996) fitted a linear trend to precipitation data from Germany for the period 1891–1990. For the winter (November–April) the slope turned out to be statistically significant over large parts of the country, corresponding to an increase in the mean of 10–15% in the second half of the period compared to the first half. A strong increase has been observed in the number of days with a zonal circulation type during December, January (Bárdossy and Caspary, 1990) and in the number of days with a west-cyclonic circulation during December–February (Caspary and Bárdossy, 1995) since the early 1970s, which may (partly) be responsible for the observed trend in winter precipitation. On the other hand, Widmann and Schär (1997) demonstrated that the observed trends in winter (December–February) precipitation in Switzerland during the period 1961–1990 were primarily due to an increase in the mean precipitation amounts of the most rain-producing weather types

Table 10: Average precipitation amounts (mm) in winter (October–March). The observed values are obtained from monthly precipitation data. The simulated values are the average of ten simulation runs with conditional case 4.1 ($k = 5$). Note that in the simulations the precipitation is resampled from the period 1961-1995.

	1901–1960		1961–1990		rel. diff. (%)	
Station	obs.	sim.	obs.	sim.	obs.	sim.
Stuttgart	245.9	248.7	261.2	263.4	6.2	5.9
Frankfurt	297.9	284.8	315.6	295.0	5.9	3.6
Trier	347.6[†]	369.6	393.4	384.0	13.2	3.9
Bamberg	264.9	267.8	277.1	278.9	4.6	4.1
Karlsruhe	312.6	331.3	354.0	346.6	13.2	4.6

[†] 1908–1960

rather than to changes in the frequency of weather types.

For five stations, for which monthly data were available at KNMI, Table 10 compares the winter precipitation in the period 1901–1960 with that in the period 1961–1990 for both the observations and the simulations with a conditional model (case 4.1; $k = 5$). The simulated average winter precipitation amounts are in general close to the observed values. For two stations, Stuttgart and Bamberg, the simulated increase in precipitation is in good agreement with the observations. For the other stations, in particular Trier and Karlsruhe, the increase is underestimated. On average the simulations explain slightly more than 50% of the observed trends.

3.1.4 Reconstructed extreme winter precipitation for 1891–1925 and 1926–1960

In this section properties of the extreme N-day winter amounts in simulations conditional on the atmospheric circulation indices for the periods 1891–1925 and 1926–1960 are compared with those simulated and observed for the period 1961–1995. For seven individual stations Table 11 shows that the simulated largest values, upper quintile means and medians for the periods 1891–1925 and 1926–1960 are smaller than those for the period 1961–1995 (respectively 79 and 78 out of the 84 values are smaller). This also holds for the average relative differences of all 25 stations. On average the median and the upper quintile mean of the simulated N-day precipitation maxima in these two periods are about 6% smaller than those simulated for 1961–1995.

Figure 4 presents, for three stations and the average of the 25 stations, Gumbel plots of the 10-day winter precipitation maxima for conditional simulations on the observed circulation of the three historical 35-year periods and for the historical 1961-1995 data. The graphs of the simulations for the periods 1891–1925 and 1926–1960 are systematically below those of the simulations and observations for 1961–1995. The panel for the 25 station average suggests that these differences are "significant" since the largest values in the ten simulation runs for the periods 1891–1925 and 1926–1960 (plusses) are almost always below the historical values.

Table 11: Relative differences (%) between the largest value (Max), the upper quintile mean (QM5) and the median value of the N-day winter (October-March) precipitation maxima for the simulated data (ten runs of 35 years with conditional case 4.1; $k = 5$) and the historical data (1961–1995) for seven stations in the three 35-year periods. Avg denotes the average relative difference of all 25 stations.

	Max				QM5				Median			
Station	N=1	N=4	N=10	N=20	N=1	N=4	N=10	N=20	N=1	N=4	N=10	N=20
1961-1995												
Essen	10.1	-8.2	5.4	0.0	3.2	-3.7	4.8	3.5	-1.6	1.4	-1.0	1.7
Kahler Asten	2.8	-5.8	-5.8	-3.9	0.3	-2.0	-6.5	-3.5	-4.8	-5.1	0.3	-0.5
Trier	-10.4	7.3	-12.4	1.6	-8.4	-4.5	-7.8	0.8	-10.8	-8.1	-6.3	-3.1
Frankfurt	-3.0	-11.3	7.4	10.9	-4.4	-4.0	-3.5	0.5	-8.5	-6.9	-6.8	-2.8
Bamberg	-5.3	6.8	18.7	-9.6	-1.1	-5.3	4.9	0.1	-6.8	2.4	-3.1	-0.2
Freudenstadt	-0.6	-4.4	0.3	-9.7	-3.1	-7.9	0.8	-5.8	-13.8	-16.9	-6.2	-7.7
Stuttgart	-0.3	1.6	2.1	0.2	0.1	0.2	3.4	-0.8	-4.9	-6.5	-7.9	-7.2
Avg	-6.0	-2.5	1.6	-1.0	-3.6	-4.2	-1.8	-0.4	-5.2	-5.1	-3.3	-2.2
1926-1960												
Essen	7.7	-2.9	-0.3	-5.4	2.7	-5.9	-1.0	-4.4	-3.4	-3.2	-9.6	-7.0
Kahler Asten	6.3	-9.8	-11.9	-13.6	1.5	-8.3	-15.0	-14.1	-3.7	-10.8	-8.4	-13.7
Trier	-20.1	-7.7	-24.6	-12.6	-10.5	-11.4	-14.7	-6.7	-15.4	-12.1	-10.9	-6.6
Frankfurt	-7.5	-23.8	-11.6	-2.1	-5.5	-10.0	-11.3	-5.6	-8.0	-10.5	-9.7	-6.9
Bamberg	-17.1	2.5	6.3	-10.8	-7.0	-8.6	-3.6	-7.0	-9.4	-5.8	-8.8	-11.4
Freudenstadt	-0.3	-8.3	-8.5	-22.2	-8.7	-15.2	-10.5	-17.0	-17.7	-23.4	-20.1	-21.9
Stuttgart	-12.2	-0.2	-8.5	-10.0	-14.5	-10.4	-7.1	-9.2	-10.7	-13.4	-15.3	-15.1
Avg	-10.5	-9.5	-9.4	-10.1	-8.1	-9.8	-10.3	-8.1	-7.7	-9.2	-10.1	-9.7
1891-1925												
Essen	3.1	-7.7	4.3	-8.3	-0.6	-6.4	-0.9	-4.8	-3.6	-5.6	-9.2	-5.5
Kahler Asten	2.6	-13.5	-14.6	-19.0	-2.7	-12.2	-17.1	-17.0	-12.3	-14.6	-9.2	-12.0
Trier	-24.2	-11.6	-22.2	-11.6	-12.6	-9.3	-13.3	-6.6	-12.1	-10.4	-10.8	-6.1
Frankfurt	-7.5	-24.0	-5.5	-5.4	-7.0	-8.6	-9.6	-8.7	-8.7	-10.5	-10.1	-5.4
Bamberg	-4.3	7.4	6.9	-15.8	4.5	-5.1	-3.3	-7.8	-10.4	-4.5	-8.8	-8.8
Freudenstadt	-3.7	-9.2	-10.7	-23.3	-10.8	-16.7	-11.9	-20.3	-19.0	-25.5	-19.0	-21.3
Stuttgart	-13.1	-2.2	-8.2	-8.4	-9.0	-7.2	-3.9	-6.3	-10.1	-12.9	-11.4	-12.6
Avg	-10.6	-8.6	-7.5	-12.3	-8.1	-8.9	-9.8	-9.5	-8.5	-10.4	-10.4	-9.5

Table 12 (Part I) finally, compares some more characteristics of the simulated precipitation and temperature for the periods 1926–1960 and 1891–1925 with those for the 1961–1995 simulations. It can be seen that not only the simulated precipitation amounts are relatively low for the two earlier periods but also their standard errors, in particular those for the monthly precipitation amounts.

3.1.5 Sensitivity of simulated precipitation to systematic changes in the circulation

Besides the reconstruction of precipitation trends and properties of N-day extremes for the years before 1961, it is interesting to study for conditional resampling the sensitivity of the simulated precipitation to systematic changes in the individual circulation indices. A

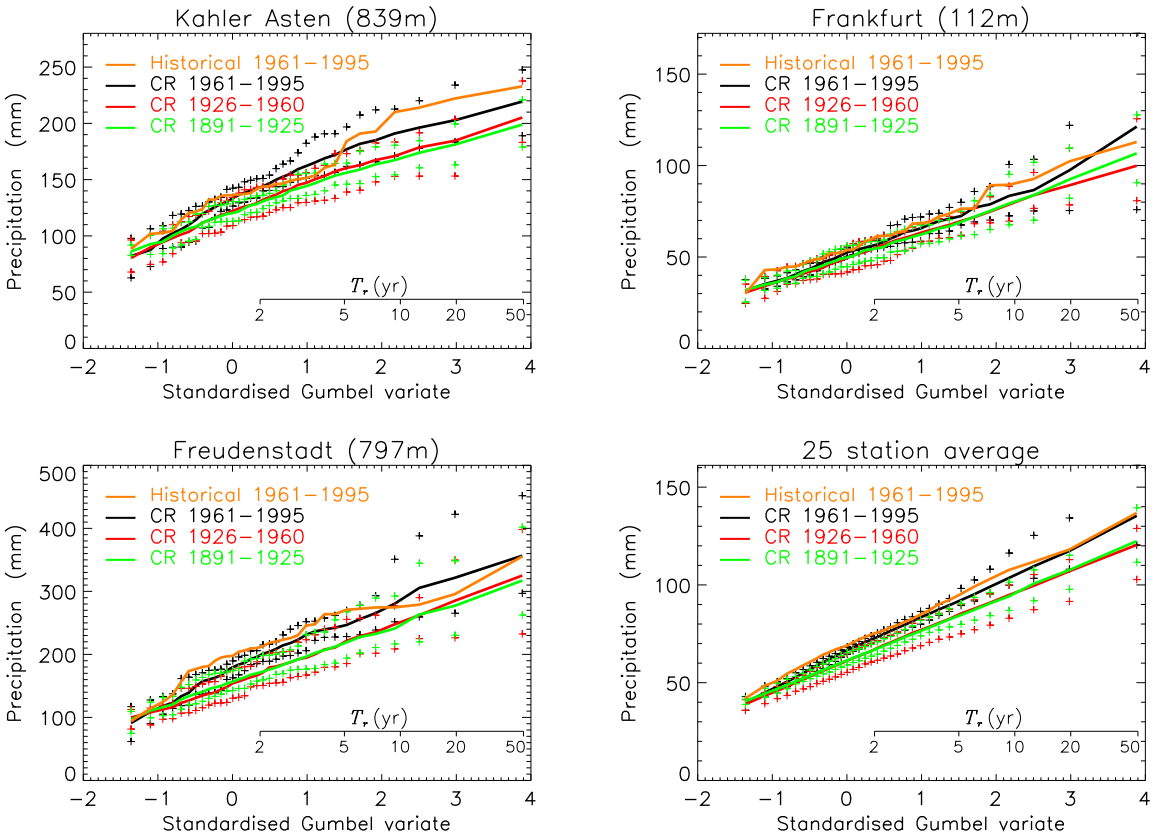

Figure 4: Gumbel plots of the 10-day winter precipitation maxima for conditional simulations on the observed circulation of three historical 35-year periods (average of ten 35-year simulation runs; case 4.1; $k = 5$) and for the historical 1961-1995 data. The plusses denote the largest and smallest values in the ten simulation runs.

sensitivity study was performed consisting of three simulations conditional on the 1961–1995 circulation indices in which in each simulation one of the three indices was systematically increased by half of its standard deviation. For the W index this means more and stronger west circulations, for the S index more and stronger southerly flow and for the Z index more and stronger cyclonic activity.

The results of this sensitivity study are presented in Part II of Table 12. The simulation with a systematic increase in the W index gives, averaged over all 25 stations, a 27% increase in mean winter precipitation and an increase in mean winter temperature of about 1 °C. The average increase in the 10-day winter precipitation maxima (almost 10%) is much smaller than that in the mean. This is because a large part of the increase in mean winter precipitation is due to an increase of 19% in the number of wet days, which has less influence on the extremes. The temperature increase is consistent with an enhanced advection of relatively warm maritime air during winter. A systematic increase in the S index results on average in a 6% decrease in mean winter precipitation and a small decrease in the extreme value properties. Systematically increasing the Z index gives an average increase of 15% in mean winter precipitation without an accompanying temperature change. Because the increase in mean precipitation is mainly due to an increase in the number of wet days, the change in the average 10-day precipitation maxima is relatively small. For the individual stations the precipitation change due to the systematic increases in the W and S indices

Table 12: Sensitivity of simulated precipitation and temperature characteristics to the circulation indices used in the resampling procedure (ten runs of 35 years with case 4.1) for the winter (October–March). For each characteristic the differences (mean temperature in °C and lag 1 autocorrelation coefficients) or percentage differences (other characteristics) are given between the simulated and the historical data (1961–1995). Apart from Kahler Asten, Freudenstadt and Köln the values are averaged over 25 stations.

Case or station	Circ. indices	mean		s		$r(1)$		s_m		10-day P		
		P	T	P	T	P	T	P	T	Max	QM5	Med
Part I. Historical indices												
4.1 (k=5)	1961-1995	-0.3	0.19	-1.1	-6.9	-0.046	-0.090	-3.8	-19.2	1.6	-1.8	-3.3
4.1 (k=5)	1926-1960	-7.6	-0.03	-5.8	-4.5	-0.060	-0.074	-11.9	-15.8	-9.4	-10.3	-10.1
4.1 (k=5)	1891-1925	-4.2	0.15	-5.1	-7.3	-0.069	-0.089	-15.3	-23.5	-7.5	-9.8	-10.4
Part II. Changed 1961–1995 indices												
4.1 (k=5)	$\tilde{Z}+0.5$	14.6	0.05	3.3	-10.0	-0.082	-0.100	-5.5	-23.5	0.2	-1.2	0.9
4.1 (k=5)	$\tilde{W}+0.5$	26.8	1.27	10.7	-12.7	-0.075	-0.126	2.7	-27.2	9.3	7.1	9.3
4.1 (k=5)	$\tilde{S}+0.5$	-6.0	0.82	-2.6	-4.0	-0.045	-0.088	-7.7	-17.7	-2.1	-4.3	-5.8
Kahler A.	$\tilde{Z}+0.5$	7.2	-0.22	-2.5	-9.7	-0.135	-0.168	-10.4	-25.4	-10.1	-11.5	-0.4
Freudenst.	$\tilde{Z}+0.5$	12.1	-0.05	-1.6	-10.8	-0.106	-0.149	-13.7	-22.6	-6.5	-2.2	-6.1
Köln	$\tilde{Z}+0.5$	10.8	-0.04	0.2	-9.4	-0.079	-0.106	-3.6	-22.5	-1.8	-2.8	-10.2
Part III. Simulated indices												
4.1 (k=5)	circ1.20	-2.7	0.10	-2.8	-7.4	-0.065	-0.096	-11.5	-24.2	-7.9	-9.1	-9.0
4.1 (k=5)	circ2.5	-4.8	0.11	-3.8	-7.7	-0.060	-0.096	-10.9	-25.4	-6.9	-7.9	-7.2
4.1 {.8, .2}	circ2.5	-2.7	0.06	-3.1	-6.1	-0.059	-0.078	-8.3	-20.2	-4.7	-5.2	-6.4
4.1 (k=1)	circ2.5	-2.7	0.03	-3.0	-5.6	-0.059	-0.074	-9.5	-20.0	-9.0	-7.4	-6.4

has always the same sign. This is, however, not the case for a systematic increase in the Z index. Even though, for individual stations a systematic increase in the Z index leads to enhanced mean winter precipitation, for some stations (Kahler Asten, Freudenstadt and Köln), the median and the upper quintile mean of the 10-day winter precipitation maxima are reduced.

3.2 Simulations conditional on simulated circulation

The main objective of the project is to produce long-duration precipitation and temperature simulations. Long-duration conditional simulations can only be based on "synthetic" circulation time series. Synthetic time series of circulation indices can be obtained from an unconditional resampling model (as discussed in section 2) or from GCM experiments, although the latter usually have limited length. Here, precipitation and temperature were generated conditional on simulated circulation indices. These indices were simulated with the unconditional resampling models discussed in section 2. First, two simulations are compared that are based on the indices from respectively the first-order model with $k = 20$ (circ1.20) and the second-order model with $k = 5$ (circ2.5). With respect to the reproduction of the characteristics of the circulation indices circ2.5 is considered the "best" model

while circ1.20 is more or less the "worst" model. The results (Table 12, Part III) show that there is little difference between the two simulations. The simulation conditional on the circ2.5 indices slightly better reproduces the properties of the 10-day precipitation maxima. For both simulations the reproduction of these characteristics, however, has become worse compared to the simulation of the same model conditional on the historical circulation indices (Top of Table 12).

Table 12 (Part III) gives the results of two more simulations, one using a discrete kernel with probabilities 0.8 and 0.2, and another with $k = 1$. Note that for $k = 1$ only one conditional realisation exists, but since the simulated indices consist of 350 years there still are 10 independent 35-year simulations available. The slight improvement for these two simulations is mainly seen in the reproduction of the temperature characteristics.

3.3 Selection effects

Table 13 compares the frequency distributions of the number of times that the various historical observations occur in the simulation runs discussed in the previous sections. The simulations of P and T conditional on the circulation indices for the periods 1926–1960 and 1891–1925 show much stronger selection effects than those conditional on the circulation for the period 1961–1995. In the last case, the pool of days from which is resampled (1961–1995) contains the same circulation vectors as the series on which is conditioned (also 1961–1995). This correspondence makes that for each day in the period 1961–1995 the day with the same $\mathbf{C}$ in the pool from which is resampled becomes one of the neighbours with a small Euclidean distance (because the contribution of $\mathbf{C}$ to the Euclidean distance is zero), and therefore this day has a relatively large probability of being resampled. As a consequence, almost every day in the pool from which is resampled has a finite probability of being resampled. This situation does not hold for the other two periods since the circulation indices for the period 1961–1995 differ in some respects from the two earlier periods.

The strongest selection effect is found for the simulations in which the 1961–1995 indices are systematically increased. It is most pronounced in the simulations with a systematic increase in the Z index. For the simulations with a systematic increase in the W or S index it is weaker but still considerable. The simulations conditional on resampled circulation indices from the period 1961–1995 have a slightly wider frequency distribution than those conditional on the historical 1961–1995 indices.

The forward persistence models (cases 8.1 and 8.2) show a strong selection effect. This may explain the large systematic deviations in the mean and the standard deviations of the simulated data from these models.

3.4 Long-duration simulations

A few long-duration conditional simulations were performed. These are compared with a long-duration *un*conditional simulation and the historical data. Figure 5 shows Gumbel plots of the 10-day winter precipitation maxima for three stations and the average of the 25 stations. Three different simulation experiments are presented. The points for the simulated data refer to average ordered 10-day maxima in three runs of 1000 years each. As expected the points for the conditional 1000-year simulations are below those for the unconditional 1000-year simulation. It can be seen that the differences between the conditional and unconditional simulations do not increase at long return periods (T_r). Table 14 presents some properties of the N-day winter maxima. The relative differences between the conditional and unconditional 1000-year simulations are for the 200-year event somewhat smaller than

Table 13: Number of historical days drawn r times in ten 35-year simulation runs compared with the number expected for the standard bootstrap. The largest number of times that a historical day is drawn is given in the last column.

Case	Circulation indices	r							$r_{\max}$
		0	1-5	6-10	11-15	16-20	21-25	>25	
bootstrap		1	856	6591	4704	602	20	<1	
conditional									
4.1 ($k=5$)	1961–1995	30	2049	5280	3923	1217	229	47	29
4.1 ($k=5$)	1926–1960	271	2903	4393	2984	1455	526	243	44
4.1 ($k=5$)	1891–1925	261	3036	4194	3081	1402	561	240	55
4.1 ($k=5$)	$\tilde{Z}+0.5$	552	3776	3358	2386	1396	735	572	51
4.1 ($k=5$)	$\tilde{W}+0.5$	414	3489	3769	2598	1416	614	475	61
4.1 ($k=5$)	$\tilde{S}+0.5$	291	3679	4013	2433	1199	571	589	62
4.1 ($k=5$)	circ1.20	64	2294	5064	3684	1289	301	79	41
4.1 ($k=5$)	circ2.5	79	2262	5029	3735	1299	302	69	38
4.1 {0.8,0.2}	circ2.5	54	2456	4990	3487	1315	362	111	54
4.1 ($k=1$)	circ2.5	66	2542	4994	3329	1290	409	145	55
8.1 ($k=5$)	1961–1995	161	3278	4172	2909	1368	590	297	55
8.2 ($k=5$)	1961–1995	227	3515	3994	2588	1414	631	406	59
unconditional									
4.4 ($k=5$)		29	2028	5327	3884	1259	209	39	34

for the 50-year event[2]. The values for the conditional simulation based on the circulation indices simulated with circ2.5 (the second order model with $k = 5$) are closer to the unconditional simulation than those based on circ1.20 indices. For the simulation with circ2.5 indices the average difference with the unconditional model for 50-year and higher events of N-day winter precipitation is about -4%.

From the results in section 3.1.3 it follows that the average precipitation amounts in winter in the period 1961–1995 are 5–10% larger than those for the first 60 years of this century. The long-term average precipitation (representative of the period 1891–1995) is therefore 3–7% smaller than that for the period 1961–1995 from which is resampled. From the conditional simulations in section 3.1.4 it may be concluded that the differences in the T_r-year events of the N-day precipitation amounts are of the same size.

Let us now assume that the long-term average precipitation as well as the 50-year events are 5% smaller than those for 1961–1995. The unconditional simulation then overestimates the 50-year events on average by about 1% and the conditional simulations underestimate it by 3–4%.

[2]For the 200-year and 50-year events, respectively the 5th and 20th highest values out of 999 simulated winters were used.

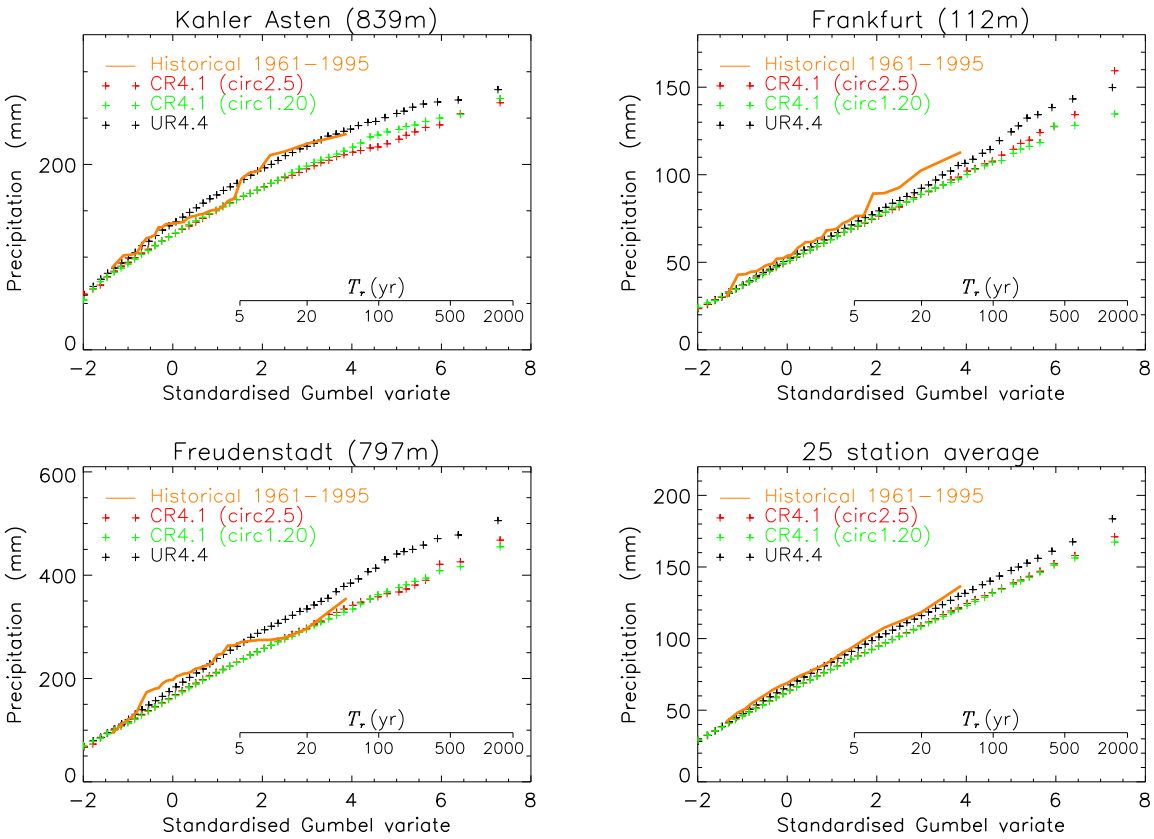

Figure 5: Gumbel plots of 10-day precipitation maxima in winter for observed and simulated data (average of three runs of 1000 years).

Table 14: The largest value (Max), the 200-year and 50-year events of N-day winter (October-March) precipitation (mm) in the 1000-year simulations (averages of three runs of 1000 years each), averaged over all 25 stations. The bottom part of the table gives the relative differences (%) between the conditional simulations and the unconditional simulation, averaged over the 25 stations.

Case	Max			200-year event			50-year event		
	N=4	N=10	N=20	N=4	N=10	N=20	N=4	N=10	N=20
unconditional									
4.4 (k=5)	121.0	183.6	239.5	102.6	152.7	205.9	89.1	130.8	180.0
conditional									
4.1 (k=5); circ1.20	117.1	167.0	225.6	99.5	142.5	191.9	85.5	121.9	167.4
4.1 (k=5); circ2.5	120.4	171.1	228.1	101.2	143.2	193.2	85.9	123.1	168.2
avg. relative differences (%)									
4.1 (k=5); circ1.20	-1.8	-7.8	-4.9	-2.0	-5.7	-6.2	-3.3	-5.9	-6.3
4.1 (k=5); circ2.5	1.1	-5.8	-3.9	-0.3	-4.9	-5.2	-2.6	-5.1	-5.8

4 Discussion and conclusions

It was demonstrated that nearest-neighbour resampling is a useful method for the simulation of daily circulation indices. To satisfactorily reproduce the autocorrelation, in particular that of higher order, and the average run lengths of the circulation indices the resampling model should be of second order. Best results are obtained with a decreasing kernel with k as small as 5 (or even 2). The use of the oversmoothed data from the period 1961–1965 in the resampling procedure is, as expected, reflected in the statistics of the simulated circulation indices but it does not have any influence on the model choice.

Compared to unconditional simulation of P and T (case 4.4) conditional simulation (case 4.1) lags behind. In the conditional simulations the daily standard deviation of T and the lag 1 autocorrelation coefficients of P and T are underestimated somewhat stronger than in the unconditional simulations. As a result the monthly standard deviations and the properties of the 10-day precipitation maxima are also underestimated stronger in the conditional simulations. In simulations conditional on simulated circulation indices instead of historical indices, the underestimation of the extreme value properties of precipitation increases further and becomes as large as 10%. To make long-duration conditional simulations we consider such an underestimation undesirable. Several attempts to improve the conditional simulation, in particular to improve the reproduction of the lag 1 autocorrelation coefficient of P were without success. Model changes leading to small improvements of the simulations conditional on the historical indices are without effect when applied to simulations conditional on simulated indices. A relatively quick (and dirty) way to improve conditional simulation could be the introduction of correlated noise to enhance the daily standard deviations and the lag 1 autocorrelation.

The choice of the variables on which is conditioned could be questioned. Work in the EC funded POPSICLE project suggests that the area-average MSLP could be a better predictor for precipitation than Z (Kilsby et al. 1998). Later work in the framework of the WRINCLE project demonstrates that atmospheric moisture is a crucial predictor for precipitation. Further, daily averages of four 6-hourly observations seem to perform better as predictor variables than an instantaneous observation each day as used here. A lot of these observations are, however, not available for the first half of the 20th century, which is essential for the present application.

In BB99 long-duration unconditional simulations were performed based on $\mathbf{C}$, P and T from the 35-year period 1961–1995. In this period the average precipitation in large parts of the Rhine basin is 5–10% larger in comparison with observations earlier this century. The conditional simulations in section 3.1.4 show that the median and the upper quintile mean of the N-day precipitation maxima for the same period can be about 6% larger. In addition, changes in the circulation have been observed (decreased vorticity; enhanced strength of westerly flow related to the North Atlantic Oscillation) in particular in the most recent 20 years. Structures protecting the country against flooding have a design life of 50–100 years. Because of the observed non-stationarity of the (precipitation) climate during a period comparable to the design life, long-duration simulations based on the short, relatively wet, 1961–1995 period, as those presented in section 3.4, should be interpreted with care. It would therefore be extremely useful if daily P and T values would be available for the 25 German stations before 1961 (e.g. 1931–1960). The generation of the area-average precipitation for all $\sim$ 150 subcatchments, required as input for the hydrological/hydraulic model, needs additional research if the time series of these area-averages are shorter than those of the point values at the individual stations.

Acknowledgements

The authors are grateful to T. Brandsma (KNMI) for comments on an earlier version of the report and to B.W.A.H. Parmet and H.C. van Twuiver (RIZA) for fruitful discussions during the progress meetings. 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 Deutscher Wetterdienst via the International Commission for the Hydrology of the Rhine Basin (CHR).

References

Al-Awadhi, S. and I. Jolliffe, 1998: Time series modelling of surface pressure data, *Int. J. Climatol.*, **18**, 443–455.

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

Beersma, J.J., 1992: *GCM control run of UK Meteorological Office compared with the real climate in the NW European winter*, Scientific report 92-02, KNMI, De Bilt, 32 pp.

Beersma, J.J. and T.A. Buishand, 1999: A simple test for equality of variances in monthly climate data, *J. Climate*, **12**, 1770–1779.

Bennekom, A.R. van, and B.W.A.H. Parmet, 1998: *Bemessungsabfluβ in den Niederlanden; menschliche Einflüsse und andere Unsicherheiten*, In: Zukunft der Hydrologie in Deutschland, BfG Mitteilung 16, 125–131, Bundesanstalt für Gewässerkunde, Koblenz (in German).

Brandsma, T., and T.A. Buishand, 1997: *Rainfall Generator for the Rhine basin: single-site generation of weather variables by nearest-neighbour resampling*, KNMI-publication 186-I, KNMI, De Bilt, 47 pp.

Brandsma, T., and T.A. Buishand, 1998: Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling, *Hydrol. Earth Syst. Sci.*, **2**, 195–209; Corrigendum, **3**, p.319.

Brandsma, T., and T.A. Buishand, 1999: *Rainfall Generator for the Rhine basin: Multi-site generation of weather variables by nearest-neighbour resampling*, KNMI-publication 186-II, KNMI, De Bilt, 58 pp.

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

Buishand, T.A. and J.J. Beersma, 1996: Statistical tests for comparison of daily variability in observed and simulated climates, *J. Climate*, **9**, 2538–2550; Corrigendum, **10** (1997), p.818.

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

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

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

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

Hulme, M. and P.D. Jones, 1991: Temperatures and windiness over the United Kingdom during the winters of 1988/89 and 1989/90 compared with previous years, *Weather*, **47**, 126–136.

Jones, P.D., M. Hulme and K.R. Briffa, 1993: A comparison of Lamb circulation types with an objective classification scheme, *Int. J. Climatol.*, **13**, 655–663.

Kilsby, C.G., P.S.P. Cowpertwait, P.E. O'Connell and P.D. Jones, 1998: Predicting rainfall statistics in England and Wales using atmospheric circulation variables, *Int. J. Climatol.*, **18**, 523–539.

Lall, U. and A. Sharma, 1996: A nearest neighbor bootstrap for resampling hydrologic time

series, *Water Resour. Res.*, **32**, 679–693.

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

Rapp, J. and C.-D. Schönwiese, 1996: Atlas der Niederschlags- and Temperaturtrends in Deutschland 1891–1990, Frankfurter Geowissenschaftliche Arbeiten, Serie B, Band 5, Frankfurt, 255 pp.

Sharma, A., D.G. Tarboton and U. Lall, 1997: Streamflow simulation: A nonparametric approach, *Water Resour. Res.*, **33**, 291–308.

Stuart, A. and J.K. Ord, 1987: *Kendall's Advanced Theory of Statistics, Vol. 1 Distribution Theory*, 5th edition, Charles Griffin & Co., 604 pp.

Widmann, M. and C. Schär, 1997: A principal component and long-term trend analysis of daily precipitation in Switzerland, *Int. J. Climatol.*, **17**, 1333–1356.

Zwiers, F. and H. von Storch, 1990: Regime-dependent autoregressive time series modeling of the Southern Oscillation, *J. Climate*, **3**, 1347–1363.

APPENDIX

Variability of estimated standard deviations

The standard deviations s and s_m in Table 9 refer to the average of six monthly values of these statistics for the winter half-year. Let τ_i be the average of s or s_m at the ith site. The average relative differences between the simulated and observed standard deviations in Table 9 can then be written as:

$$\Delta(\tau, \tau^*) = \frac{1}{25} \sum_{i=1}^{25} \left[\frac{\tau_i^* - \tau_i}{\tau_i} \right] \tag{A1}$$

where τ_i^* is the average simulated standard deviation for the ith site (averaged over ten simulation runs). The standard errors in the bottom line of that table refer to the standard deviation of the average relative differences,

$$\Delta[\tau, \mathrm{E}(\tau)] = \frac{1}{25} \sum_{i=1}^{25} \left[\frac{\tau_i - \mathrm{E}(\tau_i)}{\mathrm{E}(\tau_i)} \right]. \tag{A2}$$

In order to estimate the variance of $\Delta[\tau, \mathrm{E}(\tau)]$, we consider the logarithms $\theta_i = \ln \tau_i$ rather than the τ_i values themselves. The average over the θ_i values over the region is denoted as $\bar{\theta}$. From a first-order Taylor expansion of the τ_i values about their mean (Stuart and Ord 1987, p.324), it follows that:

$$\mathrm{var}\bar{\theta} \approx \mathrm{var}\Delta[\tau, \mathrm{E}(\tau)] \tag{A3}$$

The jackknife estimates $\mathrm{var}\,\bar{\theta}$ by recomputing the θ_i values for all subsamples wherein one year is deleted from the complete sample (Buishand and Beersma 1996; Beersma and Buishand 1999).