

Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation

Jules J. Beersma*, T. Adri Buishand

Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730 AE De Bilt, The Netherlands

ABSTRACT: Nearest-neighbour resampling was used to generate multi-site sequences of daily precipitation and temperature in the Rhine basin. The simulation is conditional on the values of 3 continuous indices of the atmospheric circulation. An advantage of nearest-neighbour resampling is that the spatial correlations of the daily precipitation and temperature data are automatically preserved in the simulated data. Comparison of different resampling models showed that the simulation of the precipitation and temperature for a new day should not only be conditioned on the circulation characteristics of that day but also on the simulated precipitation and temperature for the previous day, in order to achieve the appropriate level of persistence and variability in the generated data. With a hydrological application in mind, 980 yr multi-site simulations of daily precipitation and temperature were performed conditional on a simulated time series of circulation indices that was obtained with a second resampling model. The distribution of the extreme 10 d area-average precipitation amounts in these long-duration simulations was compared with the distribution of the historical 10 d area averages. Again, the models in which the precipitation and temperature of the previously simulated day were taken into account performed best, but even these models somewhat underestimate the quantiles of the distribution of the 10 d area-average precipitation. The long-duration simulations demonstrate that nearest-neighbour resampling is capable of producing much larger 10 d area-average precipitation amounts than the historical maximum.

KEY WORDS: Rainfall and temperature simulation · Nearest-neighbour resampling · Multi-site · Atmospheric circulation · Rhine basin

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

A wide range of stochastic models is in use to simulate synthetic daily time series of precipitation. Some models are also able to simulate daily precipitation simultaneously with other weather variables. A limited number of techniques is available to generate weather variables simultaneously at multiple locations, which is of particular interest for hydrological applications. For assessments of the effects of anthropogenic climate change, there has been considerable interest in conditioning stochastic daily precipitation models on the large-scale atmospheric circulation. Quite often a classification of observed pressure fields into weather classes has been used for this purpose. The parameters

of the precipitation model are then determined for each weather class separately (e.g. Bárdossy & Plate 1992, Wilson et al. 1992, Schubert 1994, Corte-Real et al. 1999, Fowler et al. 2000, Qian et al. 2002, Stehlik & Bárdossy 2002). An alternative is to resample from the observed precipitation in the appropriate weather class (e.g. Hughes et al. 1993, Conway et al. 1996, Palutikof et al. 2002). A somewhat different approach is to describe daily precipitation by non-homogeneous hidden Markov models (e.g. Charles et al. 1999, Bellone et al. 2000). The parameters of such a stochastic precipitation model also depend on a discrete set of weather states, but these states are unobserved (hidden). The sequence of weather states is modelled as a first-order Markov chain of which the transition proba-

*Email: beersma@knmi.nl

bilities are determined by atmospheric predictor variables. Wilby et al. (1998) used regression techniques to link wet-dry transition probabilities and the means of a suite of weather variables to atmospheric circulation characteristics.

Zorita et al. (1995) and Zorita & von Storch (1999) used the analog method for the conditional simulation of multi-site daily precipitation. The word analog refers to the historical day that is closest to the target day in terms of atmospheric circulation characteristics. In this method the analog is sampled rather than days in a specific weather class. An advantage of resampling methods is that no conceptual extensions are required to generate multivariate and/or multi-site daily sequences. Neither do they require assumptions about the underlying distributions and the spatial correlations. The analog method can be seen as a special case of nearest-neighbour resampling. The earliest applications of nearest-neighbour resampling to weather data were on multivariate single-site simulation, without conditioning on the atmospheric circulation (Young 1994, Rajagopalan & Lall 1999). Using nearest-neighbour resampling, Brandsma & Buishand (1998) compared unconditional simulations of daily temperature and precipitation with several simulations conditional on the atmospheric circulation. For the conditional simulations, it was found that the autocorrelation coefficients and extreme-value distributions of precipitation were better reproduced if, apart from circulation characteristics of the target day, the simulated precipitation and temperature of the previous day were also taken into account. A multi-site extension of unconditional nearest-neighbour simulation of daily precipitation and temperature was presented in Buishand & Brandsma (2001).

The present paper compares a stochastic version of the analog method with more general nearest-neighbour resampling techniques for conditional multi-site simulation of daily precipitation and temperature in the German part of the Rhine basin. This area was chosen with a specific application to rainfall-runoff modelling in mind. Since in the downstream area of the river Rhine the largest discharges occur in winter, the reproduction of precipitation statistics is studied for the winter half-year (October–March). Temperature is generated because rainfall-runoff models often use temperature to determine evapotranspiration, snow accumulation and snow melt.

The methodology and the data are presented in Section 2. Section 3 gives a description of the resampling models used and compares statistical properties of simulated data with those of observed data. In Section 4, finally, the results are summarized and conclusions are drawn.

2. METHODOLOGY

2.1. Nearest-neighbour resampling

Nearest-neighbour resampling was originally proposed by Young (1994) to simulate daily minimum and maximum temperatures and precipitation. Independently, Lall & Sharma (1996) discussed a nearest-neighbour bootstrap to generate hydrological time series. Buishand & Brandsma (2001) presented an application to daily precipitation and 5 other weather variables. Basically the same method was used by Buishand & Brandsma (2001) for multi-site generation of daily precipitation and temperature.

In the nearest-neighbour method weather variables such as precipitation and temperature are sampled simultaneously with replacement from the historical data. To incorporate autocorrelation, resampling is conditioned on the days in the historical record that have similar characteristics as those of the previously simulated day. One of these nearest neighbours is randomly selected and the observed values for the day subsequent to that nearest neighbour are adopted as the simulated values for the next Day t . A feature vector (or state vector) \mathbf{D}_t is used to find the nearest neighbours in the historical record. \mathbf{D}_t was based on the standardized weather variables generated for Day $t-1$ in Rajagopalan & Lall (1999) and on summary statistics of precipitation and temperature in Buishand & Brandsma (2001). Summary statistics are particularly necessary for multi-site simulations in order to avoid problems with the high-dimensional data space. As in earlier papers the k nearest neighbours of \mathbf{D}_t were selected in terms of a weighted Euclidean distance. For 2 q -dimensional vectors, \mathbf{D}_t and \mathbf{D}_u , the latter is defined as:

$$\delta(\mathbf{D}_t, \mathbf{D}_u) = \left(\sum_{j=1}^q w_j (v_{tj} - v_{uj})^2 \right)^{\frac{1}{2}} \quad (1)$$

where v_{tj} and v_{uj} are the j th components of \mathbf{D}_t and \mathbf{D}_u respectively and w_j scaling weights.

A discrete probability distribution or kernel is required to select 1 of the k nearest neighbours. Lall & Sharma (1996) recommended a kernel that gives higher weight to the closer neighbours. For this decreasing kernel the probability, p_j , that the j th closest neighbour is resampled is given by:

$$p_j = \frac{1/j}{\sum_{i=1}^k 1/i}, \quad j = 1, \dots, k \quad (2)$$

This probability kernel was also adopted in earlier applications of nearest-neighbour resampling for the Rhine basin (Buishand & Brandsma 2001).

For the simulation of weather variables conditional on the atmospheric circulation (or CNNR: conditional nearest-neighbour resampling) the procedure is slightly different. In that type of simulation, one searches for days in the historical record that have atmospheric circulation characteristics similar to those of the conditioning day. Again 1 of these nearest neighbours is randomly selected and the observed values of that nearest neighbour are adopted as the simulated values for the conditioning Day t . The feature vector \mathbf{D}_t should therefore at least consist of circulation characteristics of the conditioning Day t . In addition, simulated weather variables and/or circulation characteristics of Day $t-1$ and earlier days could be included in the feature vector.

Apart from creating a feature vector, the number k of nearest neighbours and the weights w_j have to be specified. The choice of k depends on the type of probability kernel $\{p_j\}$, the number, n , of daily values from which the nearest neighbours are selected, and the dimension, q , of the feature vector. Lall & Sharma (1996) recommended for the decreasing kernel (Eq. 2) $k = n^{1/2}$ provided that $1 \leq q \leq 6$ and $n \geq 100$. Young (1994) recommended $k = 3$ using a uniform kernel, while q was 3 and $n \approx 1200$. A sensitivity analysis in Buishand & Brandsma (2001), with the decreasing kernel and similar values for n and q as in our application, gave best results for $k = 2$ and $k = 5$. In this study the decreasing kernel with $k = 5$ was adopted. To obtain an equal contribution of all feature vector elements to the Euclidean distance, the weights w_j should be inversely proportional to the variance of the feature vector elements. This is usually a good starting point, and Brandsma & Buishand (2001) showed that variation of the weights generally has little effect on the statistical properties of the simulated data. In Wójcik & Buishand (2003) an alternative approach was introduced that avoids specification of the weights by using the Mahalanobis distance instead of the Euclidean distance.

2.2. The analog method

The analog method (e.g. Zorita et al. 1995, Zorita & von Storch 1999) is basically a special case of CNNR. In nearest-neighbour resampling, 1 of the k nearest neighbours is randomly selected from the historical record, whereas in the analog method, the closest one is always selected. The analog method is therefore identical to CNNR with $k = 1$.

Zorita et al. (1995) and Zorita & von Storch (1999) based the search for analog days on characteristics of a single conditioning day or a sequence of conditioning days. In those papers the conditioning characteristics referred to the atmospheric circulation only. Since in the analog method no randomness is involved, this

method is in essence deterministic. There is thus only 1 realisation of the simulated time series for each conditioning time series. Consequently, for simulation conditional on the historical time series of circulation indices, the conditioning day itself must be excluded, because otherwise the historical time series of weather data would be generated.

In this paper the 5 best analog days were extracted from the historical record, and 1 of these analogs was randomly selected using the decreasing kernel (Eq. 2) with $k = 5$. This stochastic version is better comparable with the CNNR models than the originally deterministic analog method.

2.3. Data

As in Buishand & Brandsma (2001) daily precipitation and temperature data from 25 German stations in the Rhine basin for 1961–1995 were used (see Fig. 1). For the 22 stations that lie below 500 m, the mean annual precipitation ranges from 542 mm (Geisenheim) to

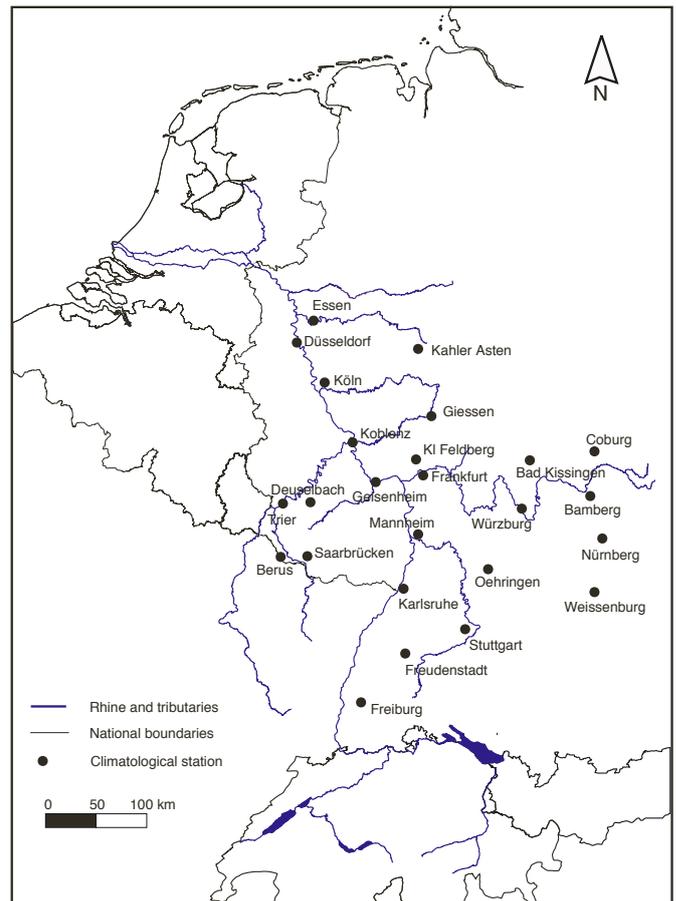


Fig. 1. Locations of the 25 German stations in the drainage basin of the river Rhine used in this study

944 mm (Freiburg) and the mean annual temperature lies between 8.2°C (Coburg) and 10.9°C (Freiburg). The 3 remaining stations, at an altitude of about 800 m, have relatively lower mean annual temperatures and higher mean annual precipitation, the latter due to orographic enhancement. The lowest mean annual temperature (5.0°C) is observed for Kahler Asten, and the largest mean annual precipitation (1691 mm) for Freudenstadt.

For the same 35 yr period, 3 daily indices of the atmospheric circulation were used: (1) strength of the westerly flow, W ; (2) strength of the southerly flow, S ; and (3) relative vorticity, Z . As in Jones et al. (1993) these circulation indices were derived from daily mean sea level pressure data from the UK Meteorological Office on a 5° latitude by 10° longitude grid, except that the grid was centered at the Rhine basin instead of the British Isles. In a number of studies the same circulation indices were used to obtain an objective version of the Lamb classification (e.g. Jenkinson & Collinson 1977, Jones et al. 1993, Goodess & Palutikof 1998, Linderson 2001).

2.4. Standardization procedure

To reduce the seasonal variation in the feature vector elements, precipitation, temperature and circulation indices were standardized. The daily temperatures and circulation indices were standardized by subtracting an estimate, m_d , of the mean and dividing by an estimate, s_d , of the standard deviation for the calendar day d of interest:

$$\tilde{x}_t = (x_t - m_d)/s_d, \quad t = 1, \dots, 365J \quad \text{and} \quad d = (t-1) \bmod 365 + 1 \quad (3)$$

where x_t and \tilde{x}_t are the original and standardized variables for Day t , respectively, and J is the total number of years in the record. The estimates m_d and s_d were obtained by smoothing the sample mean and standard deviation of the successive calendar days in a similar way as in Brandsma & Buishand (1998) and Wójcik & Buishand (2003).

Daily precipitation was standardized by dividing by a smooth estimate, $m_{d,\text{wet}}$, of the mean wet-day precipitation amount:

$$\tilde{x}_t = x_t/m_{d,\text{wet}}, \quad t = 1, \dots, 365J \quad \text{and} \quad d = (t-1) \bmod 365 + 1 \quad (4)$$

with wet days defined as days with 0.1 mm precipitation or more.

To facilitate the reproduction of seasonally varying weather characteristics the search for nearest neighbours was restricted to days within a moving window, centered on the calendar day of interest. The width of this window was 61 d, as in Brandsma & Buishand (1998) and Buishand & Brandsma (2001). Thus for the 35 yr

historical record the nearest neighbours are selected from $n = 2135$ d. At the end of the resampling procedure the simulated standardized variables are re-transformed to their original scale using the inverse of Eqs. (3) & (4).

2.5. Summary statistics

For the 25 stations in Fig. 1 precipitation, P , and temperature, T , observations were available for each day. Parsimony of feature vector elements requires that the P and T fields are described by a small number of summary statistics, like the 3 circulation indices were used to characterize the mean sea level-pressure field (i.e. the atmospheric circulation). Otherwise, considerable differences between the k nearest neighbours may occur because of the large dimension of \mathbf{D}_t . Computer time also increases with the dimension of \mathbf{D}_t .

Two important summary statistics are the arithmetic means of the standardized values of the P and T fields:

$$\tilde{P} = \frac{1}{25} \sum_{i=1}^{25} \tilde{P}_i \quad (5)$$

$$\tilde{T} = \frac{1}{25} \sum_{i=1}^{25} \tilde{T}_i \quad (6)$$

where \tilde{P}_i and \tilde{T}_i are the standardized P and T values, respectively, for the i th station. Because of the relatively large spatial variation of the P field, there is some need for a more complete summary of this field than just \tilde{P} . An additional statistic to summarize the P field is the fraction, F , of stations with precipitation above some threshold, as suggested in Buishand & Brandsma (2001). Here F was used with a threshold of 0.1 mm. The statistic F helps to distinguish between large-scale and convective precipitation. Buishand & Brandsma (2001) also considered 2 alternatives to the \tilde{P} and F combination: a vector consisting of the daily averages of the standardized values over 5 different sub-regions, and a vector consisting of the 5 leading principal components obtained from the sample covariance matrix of the \tilde{P}_i . With respect to the reproduction of the standard deviation and the autocorrelation coefficients of both precipitation and temperature, these 2 alternatives did not give better results than the simulations with the combination of \tilde{P} and F (Buishand & Brandsma 2001).

3. MODEL IDENTIFICATION AND SIMULATION RESULTS

3.1. Models used

Six resampling models were considered. Three analog-type models were distinguished: a first-order, a

third-order and a fifth-order model. In the first-order analog model (analog 1) the circulation indices of the conditioning day (Day t) were used to find the analog days. In the third- and fifth-order analog models (analog 3 and analog 5) the search for analog days was based on the circulation indices of respectively 3 and 5 consecutive conditioning days (Days $t-2$, $t-1$, t and $t-4$, ..., t respectively). In the higher-order analog models thus also a part of the evolution of the atmospheric circulation is taken into account. Zorita et al. (1995) refer to the fifth-order model as a '5-Day-segment' model. The remaining 3 resampling models are CNNR models. The first of these models (CNNR 1) contains the circulation indices of Day t and simulated precipitation and temperature characteristics of Day $t-1$ as feature vector elements. The second model (CNNR 2C) additionally contains the circulation indices of Day $t-1$, yielding a second-order model in terms of the atmospheric circulation. Model CNNR 2F, finally, uses in addition to CNNR 1 the fraction F of stations with precipitation of Day $t-2$ to determine the nearest neighbours, resulting in a second-order model in terms of precipitation. Fig. 2 schematically presents the feature vectors of these 6 models.

For a fair comparison of both types of models, a stochastic version of the analog method was used (see Section 2.2). Further, the selection of the conditioning day was excluded in the CNNR models as in the analog models. Allowing the selection of the conditioning day is considered to generate 'artificial skill' (Zorita et al. 1995). Consequently, the only difference between the analog models and the CNNR models examined here is the composition of the feature vector (see Fig. 2).

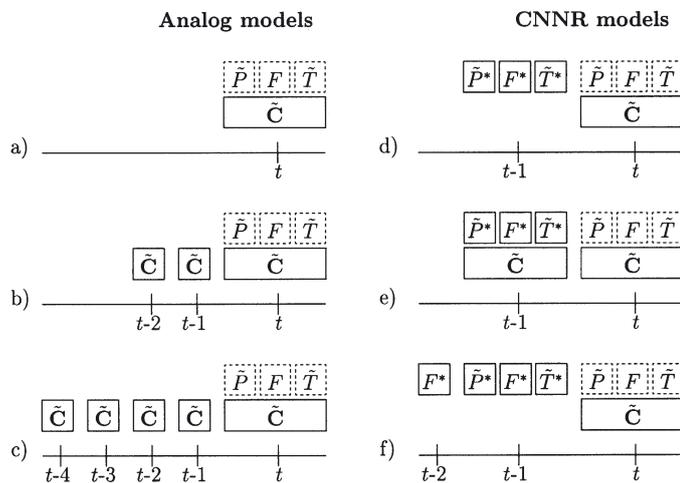


Fig. 2. Elements of the feature vector (solid boxes) for conditional simulations of new variables (dashed boxes): (a) analog 1; (b) analog 3; (c) analog 5; (d) CNNR 1; (e) CNNR 2C; (f) CNNR 2F. The vector C contains the 3 circulation indices Z , W and S ; asterisk indicates that the corresponding variable was simulated in a previous time step; tilde refers to a standardized value

The weights w_j in Eq. (1) are in the CNNR models approximately equal to the reciprocal of the variance of the feature vector elements. The weights for \tilde{P} , F and \tilde{T} were rounded to 2, 5 and 1 respectively, and for \tilde{Z} , \tilde{W} and \tilde{S} the weights are 1 as a result of the standardization. For the analog models all weights equal 1 since the feature vector elements involve only standardized circulation indices.

In Section 3.2, 2 types of conditional simulations are investigated: simulations conditional on the 1961–1995 time series of circulation indices and simulations conditional on simulated time series of circulation indices. The simulated time series of circulation indices were obtained with an unconditional nearest-neighbour resampling model. A description of that model is given in Appendix 1. Time series of simulated circulation indices are needed to generate longer time series of P and T than the historical time series of circulation indices (see Section 3.3). In the simulations conditional on *simulated* time series of circulation indices the conditioning day itself was not excluded, since no artificial skill can be inherited from a simulated time series of circulation indices.

3.2. Model results

With all 6 models, two 980 yr simulations were performed, a simulation consisting of 28 runs of 35 yr conditional on the same 35 yr record of circulation indices, and a single 980 yr simulation run conditional on 980 yr of simulated circulation indices. For comparisons with the historical data, the latter was split into 28 independent 35 yr records. Given the application of rainfall-runoff modelling for the river Rhine, the presented statistics refer to the winter half-year (October–March). Second-order moments (such as standard deviations and autocorrelation coefficients) were first calculated for each calendar month separately as in Buishand & Brandsma (2001) and then averaged over the 6 calendar months, the 25 stations and the 28 periods of 35 yr in order to reduce the influence of the seasonal cycle in the mean on these statistics.

3.2.1. Mean and second-order moments

Table 1 gives an overview of the reproduction of the means, the standard deviations of the monthly and daily values, s_M and s_D respectively, and the lag-1 and lag-2 autocorrelation coefficients of the daily values, $r(1)$ and $r(2)$ respectively.

The first part of the table refers to simulations conditional on historical circulation indices, and the second part to simulations conditional on simulated circulation indices. The table also gives the historical estimates

Table 1. Differences in means and second-order moments between the 980 yr simulations and the historical data for the winter (October–March), averaged over the 25 stations and the 28 (35 yr) runs. For the mean precipitation (monthly totals), the mean temperature and the mean lag-1 and lag-2 autocorrelation coefficients, $r(1)$ and $r(2)$, the absolute differences are given, and for the mean standard deviations of monthly and daily values (s_M and s_D) the percentage differences are given. Bottom lines: average historical (1961–1995) estimates and their standard error, SE. Mean and standard deviations are in mm for precipitation and in °C for temperature. SE are in mm for mean precipitation, in °C for mean temperature, in % for standard deviations and dimensionless for the autocorrelation coefficients. Values in **bold** refer to statistically significant differences

Model	Mean		s_M		s_D		$r(1)$		$r(2)$	
	<i>P</i>	<i>T</i>	<i>P</i>	<i>T</i>	<i>P</i>	<i>T</i>	<i>P</i>	<i>T</i>	<i>P</i>	<i>T</i>
Historical circulation indices (1961–1995)										
CNNR 1	1.0	0.27	-5.6	-19.9	-0.1	-6.8	-0.047	-0.096	-0.028	-0.069
CNNR 2C	-5.3	0.31	-8.8	-24.6	-4.6	-7.8	-0.048	-0.123	-0.026	-0.120
CNNR 2F	0.2	0.29	-2.1	-22.0	0.0	-7.5	-0.042	-0.104	-0.012	-0.087
Analog 1	-1.7	0.02	-15.4	-42.4	-0.9	-1.4	-0.181	-0.493	-0.078	-0.397
Analog 3	-7.9	0.20	-14.8	-34.4	-5.4	-2.5	-0.156	-0.352	-0.057	-0.288
Analog 5	-12.1	0.21	-18.5	-28.5	-9.4	-2.2	-0.161	-0.326	-0.063	-0.248
Simulated circulation indices										
CNNR 1	-0.8	0.26	-8.7	-23.5	-1.2	-7.3	-0.049	-0.099	-0.027	-0.077
CNNR 2C	-6.1	0.24	-13.6	-27.6	-5.0	-7.2	-0.054	-0.118	-0.030	-0.118
CNNR 2F	-1.8	0.28	-7.6	-26.5	-2.0	-7.8	-0.048	-0.108	-0.016	-0.096
Analog 1	-2.9	0.04	-18.6	-42.2	-2.1	-1.3	-0.168	-0.453	-0.073	-0.364
Analog 3	-6.8	0.16	-17.7	-36.2	-4.4	-1.7	-0.147	-0.336	-0.062	-0.282
Analog 5	-9.3	0.20	-20.2	-33.9	-6.1	-1.9	-0.152	-0.325	-0.067	-0.256
Historical	65.0	3.54	35.9	2.2	4.2	4.2	0.287	0.825	0.148	0.639
SE	3.8	0.17	4.8	6.2	2.6	2.5	0.009	0.007	0.010	0.015

and their standard errors. The standard errors of the autocorrelation coefficients were obtained with the jackknife method of Buishand & Beersma (1993). The standard errors of the standard deviations of the daily and the monthly values were calculated in a similar way, following Buishand & Beersma (1996) and Beersma & Buishand (1999) respectively. Differences larger than twice the standard error of the historical data are referred to as statistically significant (this corresponds approximately to a 2-sided test at the 5% level).

For the 3 analog models the average winter precipitation is underestimated due to the selection effects discussed in Section 3.2.2. The underestimation increases with the order of the model and becomes significant for the simulation with the analog-3 model based on historical circulation indices and for both simulations with the analog-5 model. The largest underestimation is 12.1%, whereas for the CNNR models the differences in monthly mean precipitation are not more than 6.1%. The standard deviations s_M and s_D are generally underestimated. The underestimation of the monthly standard deviations for precipitation and temperature in the analog models is about twice as large as in the CNNR models. In the analog models the underestimation of the daily standard deviation for precipitation is also somewhat larger. But for temperature, the underestimation of the daily standard deviations is at least 2 times smaller in the analog models. The biases in the lag-1 and lag-2 autocorrelation coef-

ficients for precipitation and temperature are in the analog models about 3 times as large as in the CNNR models. But even for the simulations with the latter models the autocorrelation coefficients are significantly underestimated. The bias in the autocorrelation coefficients is the main cause for the underestimation of the standard deviations of the monthly values.

For the CNNR models the biases are generally somewhat larger for the simulations based on simulated circulation indices than for those based on historical indices. For the statistics in Table 1, the best performing analog model on the whole is analog 3. Its performance is, however, still below that of the weakest CNNR model (CNNR 2C).

Time series of the area-average winter precipitation totals and the area-average winter temperatures for the simulations conditional on the historical circulation indices are compared with the observed 1961–1995 winter area-average precipitation and temperature in Figs. 3 & 4 respectively. For each simulation the coloured symbols represent the averages of the 28 runs in each winter. For the simulations with the CNNR-1 and analog-1 models, the whiskers represent the range of the 28 runs. The skill score S in the figures is defined as:

$$S = 1 - \frac{\sum_j (y_j - \hat{y}_j)^2}{\sum_j (y_j - \bar{y})^2} \quad (7)$$

where the y_j are the historical winter precipitation totals or temperature averages, \bar{y} is the overall histori-

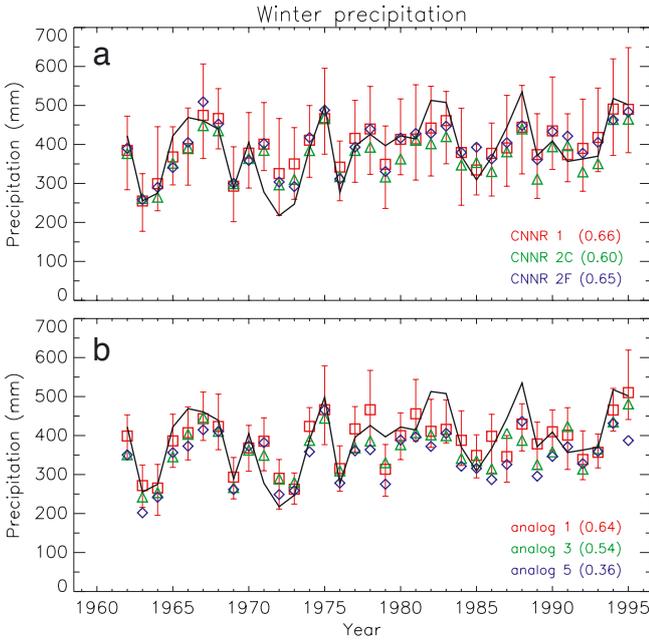


Fig. 3. Observed and simulated area-average winter precipitation totals. (a) CNNR models; (b) analog models. Black line: historical values; coloured symbols: simulated values averaged over 28 runs; red whiskers: range of values in 28 runs (CNNR-1 and analog-1 models only). Values in parentheses are skill scores (see text)

cal average and the \hat{y}_j are the simulated values for each winter, averaged over the 28 runs. Note that $S = 1$ for a perfect predictor, and $S = 0$ if \bar{y} is taken as predictor.

The temporal variation of the area-average winter precipitation is well described by the CNNR and analog-1 models, with S ranging between 0.60 and 0.66. The higher-order analog models exhibit less skill due to the significant underestimation of the mean precipitation amounts (see Table 1). This underestimation is also visible in Fig. 3. For most winters the whiskers of the simulated precipitation amounts are considerably wider for the CNNR models than for the analog models, as shown for the CNNR-1 and analog-1 models. The difference in whisker width between the CNNR and analog models is even larger for the area-average winter temperature (Fig. 4). The larger width of the whiskers in the CNNR models is likely due to a larger variation in the potential analogs (see Section 3.2.2).

The simulations overestimate the average temperature of the coldest winters (1963 and during the mid-1980s) and underestimate the temperature of the warm winters (around 1990). This conditional bias is weakest for the analog-5 model. As a result, this model has the highest skill score ($S = 0.71$). So for the analog models, the highest order is favourable for the predictive skill of the winter temperature, while the lowest order is favourable for the predictive skill of

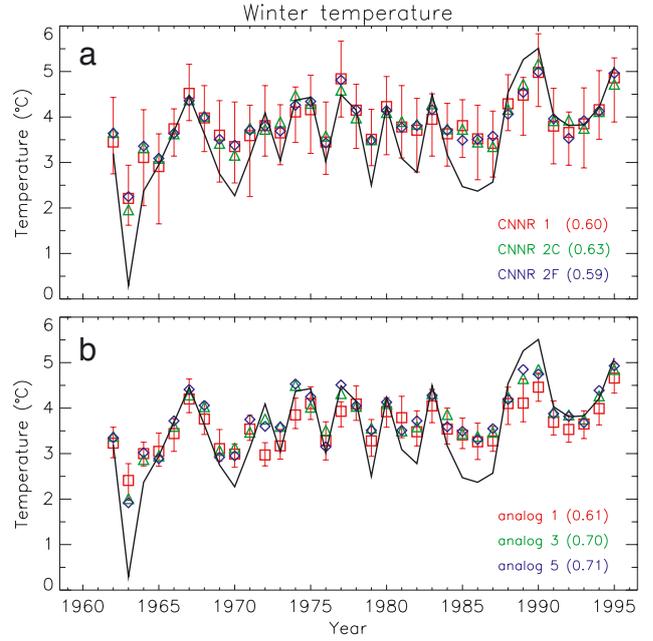


Fig. 4. Observed and simulated area-average winter temperatures. Details as in Fig. 3

the winter precipitation due to the large underestimation of the mean precipitation in the higher-order models.

3.2.2. Selection effects

As a result of random sampling with replacement, some historical days will appear more frequently in a simulation run than other days. In the standard bootstrap such differences are purely random. Nearest-neighbour resampling may, however, also lead to a systematic underselection of certain days and an overselection of other days (Young 1994). This explains for instance why the mean and the daily standard deviation s_D of the historical data are not necessarily reproduced in the simulations.

The selection effects of the simulations in this paper are studied in the same way as in Buishand & Brandsma (2001). Let K_t be the number of times that Day t ($t = 1, \dots, 365J$) appears in a simulation run of J^* yr. In the case of random sampling, K_t has a binomial distribution which can be approximated by a Poisson distribution with parameter $v = J^*/J$:

$$\Pr(K_t = r) \approx \frac{v^r e^{-v}}{r!} \quad (8)$$

Note that the distribution of K_t does not depend on the use of a moving window. For nearest-neighbour resampling the number of historical days that is drawn

Table 2. Number of historical days drawn r times in 980 yr 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	1–10	11–20	21–30	31–40	41–50	>50	
Historical circulation indices (1961–1995)								
CNNR 1	44	1096	2995	3857	2683	1220	880	152
CNNR 2C	155	1814	2830	2913	2301	1443	1319	131
CNNR 2F	37	1128	2949	3795	2671	1324	871	131
Analog 1	107	903	2787	3820	3004	1511	643	92
Analog 3	716	1993	2576	2413	1922	1378	1777	133
Analog 5	1078	2330	2498	2041	1528	1187	2113	179
Simulated circulation indices								
CNNR 1	13	1021	2999	4002	2636	1270	834	144
CNNR 2C	17	1432	3127	3272	2429	1394	1104	120
CNNR 2F	12	1135	2952	3809	2698	1314	855	127
Analog 1	0	224	2187	5620	3792	843	109	65
Analog 3	6	1335	3162	3254	2464	1520	1034	94
Analog 5	20	2305	3256	2523	1771	1218	1682	154
Bootstrap	0	1	928	7890	3797	158	1	–

r times can be compared with the number expected from the Poisson distribution with parameter v . The latter equals $365J \times \Pr(K_t = r)$.

In Table 2 the frequency distributions of r for the 980 yr simulations are compared with the frequencies for the Poisson distribution with $v = 28$. The frequency distributions are wider than the theoretical frequency distribution for random sampling as a result of selection effects. In the CNNR models the selection effect is slightly larger for the simulations based on historical circulation indices than for those based on simulated indices. Of the CNNR models, CNNR 2C (second order in the atmospheric circulation) has the largest selection effect. In the analog models the selection effect increases considerably with the order of the model. Thus, the more information of the evolution of the atmospheric circulation is used, the larger the selection effects tend to grow. Further, the differences in the selection effects between the simulations based on the historical circulation indices and those based on simulated indices are much larger than in the CNNR models. This is mainly due to the relatively large selection effect in the analog simulations conditional on historical circulation indices. Recall that in that case each simulation consists of 28 runs based on the same 35 yr. In the analog models the nearest neighbours (or analogs) are determined by the circulation characteristics of the historical record only. A particular conditioning day therefore has in each of the 28 runs the same nearest neighbourhood, i.e. the same potential analogs. In the CNNR models the nearest neighbourhood of a particular conditioning day varies among the 28 runs, since the precipitation and temperature characteristics of the previ-

ously simulated day determine the potential analogs as well. This larger variation in potential analogs is probably responsible for the smaller selection effect.

Surprisingly, large differences in the selection effects do not necessarily lead to large differences in the standard deviations and autocorrelation coefficients. In particular for the analog-1 model the selection effect for the simulation based on historical indices is much larger than for the simulation based on simulated indices, but the differences between the second-order moments (Table 1) are relatively small. In the simulations with the largest selection effects, however, the average monthly precipitation amount is significantly underestimated, indicating an underselection of days with large rainfall and an overselection of relatively dry days.

3.2.3. Temporal dependence of spatial patterns

A resampling technique automatically preserves the spatial patterns of the daily precipitation and temperature fields, but it does not necessarily reproduce the dependence between the patterns of successive days. Two measures of the temporal dependence between the spatial patterns were considered: the pattern correlation of 2 days that are l days apart (for precipitation, days with no rainfall were excluded), and the length of the difference vector of 2 days that are l days apart. The pattern correlation correlates the spatial fields relative to their respective spatial means (centered statistic). The reproduction of the pattern correlation and the difference vector of the precipitation and the temperature fields is presented in Table 3.

Table 3. Differences in the lag-1 and lag-2 pattern correlations, $r_p(1)$ and $r_p(2)$ respectively, and the lengths of the lag-1 and lag-2 difference vectors, $d(1)$ and $d(2)$, between the 980 yr simulations and the historical data for the winter (October–March), averaged over the 28 (35 yr) runs. Absolute differences are given for $r_p(1)$ and $r_p(2)$, and percentage differences for $d(1)$ and $d(2)$. Bottom line: average historical (1961–1995) estimates. $r_p(1)$ and $r_p(2)$ are dimensionless, and $d(1)$ and $d(2)$ are given in mm for precipitation and in °C for temperature

Model	$r_p(1)$		$r_p(2)$		$d(1)$		$d(2)$	
	P	T	P	T	P	T	P	T
Historical circulation indices (1961–1995)								
CNNR 1	-0.089	-0.128	-0.044	-0.065	6.2	18.4	3.2	3.4
CNNR 2C	-0.083	-0.121	-0.041	-0.070	0.5	22.0	-2.3	8.1
CNNR 2F	-0.087	-0.122	-0.040	-0.060	5.2	19.3	1.2	4.7
Analog 1	-0.100	-0.167	-0.040	-0.087	16.0	90.1	5.6	42.2
Analog 3	-0.089	-0.141	-0.039	-0.082	7.6	66.6	-1.2	30.6
Analog 5	-0.088	-0.145	-0.040	-0.082	3.7	61.9	-4.9	26.9
Historical	0.271	0.774	0.196	0.672	86.3	53.0	98.3	75.2

There is always an underestimation of the pattern correlation. The biases in the lag-1 precipitation pattern correlation are similar in both types of models. For temperature the biases are somewhat larger, in particular for the analog models. The biases in the lag-2 pattern correlation coefficients are about half of those in the lag-1 coefficients. The simulations based on simulated circulation indices have similar biases (not shown).

Most lagged difference vectors are overestimated. Overall, the overestimation is worse for temperature than for precipitation, and worse for the analog models than for the CNNR models. This overestimation is mainly due to the underestimation of the autocorrelation coefficients (Table 1). The underestimation of the latter also contributes to the underestimation of the pattern correlation. However, unlike the difference vector,

the pattern correlation also depends on lagged cross-correlations between the daily temperatures (or daily precipitation) at different locations. From a first-order approximation of the expected value of the pattern correlation, it can be shown that the effect of the underestimation of the autocorrelation is partly compensated by biases in the lagged cross-correlations. As a result the pattern correlation looks less sensitive to biases in the temporal dependence than the difference vector.

3.2.4. Dry spell counts and dry spell lengths

Table 4 presents the relative biases of the average number of dry days ($P = 0$ mm); the average number of dry spells (i.e. series of consecutive dry days); the average dry spell length, DSL; the longest dry spell in a 35 yr period, $DSL_{\max 35}$; and the bias of the lag-1 wet-dry autocorrelation coefficient, $r_{wd}(1)$. These statistics were again calculated for the winter half-year (October–March). The number of dry days is nearly correct in the CNNR and first-order analog models, but the third- and fifth-order analog models overestimate this by 4 to 6%. This overestimation partly explains the underestimation of the mean precipitation (see Table 1) and is due to the previously discussed selection effects. In both types of models the number of dry spells is overestimated: in the CNNR models somewhat more than 10%, and in the analog models more than 30%. As a result the average spell length is underestimated: in the CNNR models slightly more than 10%, and in the ana-

Table 4. Differences in the number of dry days, the number of dry spells, the average dry spell length, DSL, the maximum dry spell length in 35 yr, $DSL_{\max 35}$, and the lag-1 wet–dry autocorrelation coefficient, $r_{wd}(1)$, between the 980 yr simulations and the historical data in winter (October–March). Absolute differences are given for $r_{wd}(1)$, and relative differences (%) for the other statistics. Absolute and relative differences are averaged over the 25 stations and the 28 runs of 35 yr. Bottom line: historical (1961–1995) estimates; counts are per winter, lengths are in days and $r_{wd}(1)$ is dimensionless

Model	No. of dry days	No. of dry spells	DSL	$DSL_{\max 35}$	$r_{wd}(1)$
Historical circulation indices (1961–1995)					
CNNR 1	-1.6	13.3	-13.1	-8.8	-0.082
CNNR 2C	1.1	12.5	-10.1	-4.7	-0.074
CNNR 2F	-1.0	13.2	-12.4	-5.3	-0.081
Analog 1	1.1	38.1	-26.8	-28.1	-0.227
Analog 3	3.8	33.7	-22.3	-16.0	-0.198
Analog 5	5.8	35.4	-21.8	-15.6	-0.208
Historical	83.8	26.6	3.1	24.4	0.411

log models slightly more than 20%. Apart from the analog-1 model, the relative underestimation of $DSL_{\max 35}$ is considerably smaller than that of the average dry spell length.

Like the lag-1 autocorrelation coefficient of the daily precipitation amounts in Table 1, the lag-1 wet-dry autocorrelation coefficient is significantly underestimated in all simulations (the jackknife SE of the lag-1 wet-dry autocorrelation for the historical data is 0.010). The underestimation of both $r_{wd}(1)$ and DSL can be understood from the relation (Buishand 1978):

$$r_{wd}(1) = 1 - 1/DSL - 1/WSL \quad (9)$$

where WSL is the average wet spell length. If, as in Table 4, the number of dry spells is overestimated, then the number of wet spells is also overestimated (since by definition a wet spell follows a dry spell and vice versa). As a result both DSL and WSL are underestimated and consequently $r_{wd}(1)$ is underestimated.

In terms of spell counts and spell lengths there is very little difference between the simulations based on historical circulation indices and those on simulated circulation indices (not shown).

3.3. Long-duration simulations

Monte Carlo techniques enable us to produce synthetic time series of precipitation and temperature that are much longer than the observed records. Using such long-duration simulations as input into a rainfall-runoff model offers the opportunity to get more reliable estimates of the probabilities of extreme river discharges.

The 980 yr simulations conditional on simulated circulation indices that were split into 28 independent 35 yr records earlier are now used as single long-duration simulations. The distribution of the extreme 10 d area-average precipitation amounts in these simulations is examined in this section. An interval of 10 d was chosen because flooding of the river Rhine is often caused by large rainfall in winter over periods of about 10 d. An analysis of the January 1995 Rhine flood in Germany (Fink et al. 1996) demonstrated that in parts of the basin the monthly totals were more than 3 times as large as the climatological averages and that about 70 to 80% of these high monthly totals fell within a period of only 10 d.

The largest 10 d area-average precipitation amounts (average of all 25 stations) in each winter (October–March) were extracted from the 980 yr simulations and the 35 yr historical data. Fig. 5 presents Gumbel plots of these winter maxima (the horizontal scale in these plots is such that the ordered maxima follow a straight line in the case of a Gumbel distribution).

The Gumbel plots show that much larger 10 d area-average amounts (up to 35%) are simulated than the historical (1961–1995) maximum. Such unprecedented rainfall events can be very useful for hydrological design. Fig. 5 further shows that all models underestimate the 10 d area-average precipitation amounts for return periods between 5 and 20 yr. The underestimation is largest for the analog models and the CNNR-2C model. The analog models systematically underestimate the 10 d area-average precipitation amounts for all return periods, indicating that these models are not very suitable for applications where the extreme multi-day precipitation amounts are of interest.

4. SUMMARY AND CONCLUSIONS

Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation has been studied for 25 stations in the German part of the Rhine basin using CNNR and a stochastic version of the analog method. To fully explore the differences between the CNNR and the analog models, the simulations were divided into simulations conditional on historical time series of circulation indices and simulations conditional on simulated time series of circulation indices. A second resampling model was used to generate long-duration time series of circulation indices.

All conditional simulation models have a tendency to underestimate the standard deviations and autocorrelation coefficients of daily precipitation and temperature and the standard deviations of the monthly precipitation totals and the monthly mean temperatures.

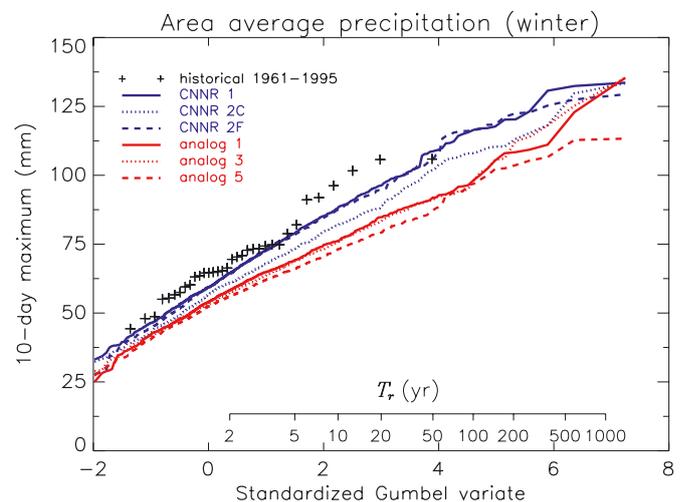


Fig. 5. Gumbel plots of the 10 d winter precipitation maxima for 980 yr simulations conditional on simulated circulation indices and for the historical 1961–1995 data. T_r represents the return period in years.

In general, the underestimation is larger for the analog models than for the CNNR models, with the exception of the underestimation of the standard deviation of the daily temperatures, which is much smaller in the analog models. The number of days that is never or almost never selected is relatively large for the analog models conditional on historical circulation indices. The mean precipitation amounts are significantly underestimated in the simulations with the largest selection effects. All simulations conditional on the historical circulation indices underestimate the average temperature in the warmest winters and overestimate the average temperature in cold winters. The reproduction of the temporal dependence of the spatial patterns of precipitation and temperature by both types of models turned out favourably for the CNNR models, although the differences were not as large as for the univariate autocorrelations. The biases in the dry spell counts and the dry spell lengths are for the analog models often more than twice as large as for the CNNR models. The CNNR models also reproduced the extreme-value distribution of the 10 d area-average winter precipitation amounts better than the analog models. Despite an underestimation of the quantiles of this distribution, the highest 10 d area-averages were in most 980 yr simulations much larger than the highest observed 10 d area-average.

Since, the observed weather of historical days is resampled, the dependence between daily precipitation at different sites and the dependence between daily precipitation and temperature is automatically preserved. These dependencies often have a complicated structure, which may not be adequately described by parametric models. For many hydrological applications the spatial dependency is of crucial importance. This makes multivariate resampling models particularly suitable for hydrological purposes. The comparison between the analog models and the CNNR models, however, demonstrates that besides the circulation characteristics of the target day, also the precipitation and temperature characteristics of the previously simulated day should be taken into account.

A few potential limitations of the methodology can also be identified. The method is rather data intensive and, resampling of multivariate data may become problematic if data are missing (which is quite common in observational records). The method does not produce new daily precipitation and temperature fields but merely reshuffles the historical days to form realistic new sequences of those fields. As a result daily rainfall amounts cannot be larger than those observed. Similarly, daily temperatures for a particular location cannot be lower or higher than the observed minimum or maximum value for that location. The latter limitation may seriously bias the

results of CNNR in climate change applications. To overcome this limitation, Lall & Sharma (1996) suggested evaluating the means of the required variables first from the selected nearest neighbours and then to perturb these values by a residual using nearest-neighbour resampling. For conditional simulation on atmospheric predictors this strategy may need extension to allow for predictor values outside their range in the historical data.

It should further be noted that several studies of climate change simulations with General Circulation Models (GCMs) have revealed that changes in precipitation usually cannot be explained by changes in the atmospheric circulation alone. Consequently it becomes recognized that the simulation of precipitation should also be conditioned on (large-scale) predictors directly related to the atmospheric moisture and the temperature (for an overview, see Giorgi et al. 2001). For similar reasons the simulation of temperature should include predictors like the large-scale (2 m) temperature, geopotential height or the thickness of an atmospheric layer (e.g. Huth et al. 2001, Benestad 2002).

Acknowledgements. The authors would like to thank 2 anonymous reviewers for their useful comments. The UK Meteorological Office gridded mean sea level pressure 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/KHR). The work was performed in co-operation with the Institute for Inland Water Management and Waste Water Treatment (RIZA).

LITERATURE CITED

- Al-Awadhi S, Jolliffe I (1998) Time series modelling of surface pressure data. *Int J Climatol* 18:443–455
- Bárdossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. *Wat Resour Res* 28:1247–1259
- Beersma JJ, Buishand TA (1999) A simple test for equality of monthly variances in climate time series. *J Clim* 12: 1770–1779
- Bellone E, Hughes JP, Guttorp P (2000) A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Clim Res* 15:1–12
- Benestad RE (2002) Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble. *Clim Res* 21:105–125
- Brandsma T, Buishand TA (1998) Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrol Earth Syst Sci* 2:195–209 (Corrigendum, *Hydrol Earth Syst Sci* 3:319, 1999)
- Buishand TA (1978) The binary DARMA(1,1) process as a model for wet-dry sequences. Technical Note 78–01, Department of Mathematics, Agricultural University Wageningen

- Buishand TA, Beersma JJ (1993) Jackknife tests for differences in autocorrelation between climate time series. *J Clim* 6: 2490–2495
- Buishand TA, Beersma JJ (1996) Statistical tests for comparison of daily variability in observed and simulated climates. *J Clim* 9:2538–2550 (Corrigendum, *J Clim* 10:818, 1997)
- Buishand TA, Brandsma T (2001) Multi-site simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbour resampling. *Wat Resour Res* 37: 2761–2776
- Charles SP, Bates BC, Hughes JP (1999) A spatiotemporal model for downscaling precipitation occurrence and amounts. *J Geophys Res* 104:31567–31669
- Conway D, Wilby RL, Jones PD (1996) Precipitation and air flow indices over the British Isles. *Clim Res* 7:169–183
- Corte-Real J, Xu H, Qian B (1999) A weather generator for obtaining daily precipitation scenarios based on circulation patterns. *Clim Res* 13:61–75
- Fink A, Ulbrich U, Engel H (1996) Aspects of the January 1995 flood in Germany. *Weather* 51:34–39
- Fowler HJ, Kilsby CG, O'Connell PE (2000) A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions. *Hydrol Earth Sys Sci* 4:263–282
- Giorgi F, Hewitson B, Christensen J, Hulme M, von Storch H, Whetton P, Jones R, Mearns L, Fu C (2001) Regional climate information—evaluation and projections. In: Houghton, JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds) *Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, p 583–638
- Goodess CM, Palutikof JP (1998) Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *Int J Climatol* 18:1051–1083
- Hughes JP, Lettenmaier DP, Guttorp P (1993) A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation. *Wat Resour Res* 29:3303–3315
- Huth R, Kyselý J, Dubrovský M (2001) Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series. *J Clim* 14: 4047–4061
- Jenkinson AF, Collinson FP (1977) An initial climatology of gales over the North Sea. *Synoptic Climatology Branch Memorandum 62*, Meteorological Office, Bracknell
- Jones PD, Hulme M, Briffa KR (1993) A comparison of Lamb circulation types with an objective classification scheme. *Int J Climatol* 13:655–663
- Lall U, Sharma A (1996) A nearest neighbor bootstrap for resampling hydrologic time series. *Wat Resour Res* 32: 679–693
- Linderson ML (2001) Objective classification of atmospheric circulation over southern Scandinavia. *Int J Climatol* 21: 155–169
- Palutikof JP, Goodess CM, Watkins SJ, Holt T (2002) Generating rainfall and temperature scenarios at multiple sites: examples from the Mediterranean. *J Clim* 15: 3529–3548
- Qian B, Corte-Real J, Xu H (2002) Multisite stochastic weather models for impact studies. *Int J Climatol* 22: 1377–1397
- Rajagopalan B, Lall U (1999) A *k*-nearest-neighbor simulator for daily precipitation and other variables. *Wat Resour Res* 35:3089–3101
- Schubert S (1994) A weather generator based on the European 'Grosswetterlagen'. *Clim Res* 4:191–202
- Stehlik J, Bárdossy A (2002) Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. *J Hydrol* 256:120–141
- Wilby RL, Hassan H, Hanaki K (1998) Statistical downscaling of hydrometeorological variables using general circulation model output. *J Hydrol* 205:1–19
- Wilson LL, Lettenmaier DP, Skillingstad E (1992) A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *J Geophys Res* 97:2791–2809
- Wójcik R, Buishand TA (2003) Simulation of 6-hourly rainfall and temperature by two resampling schemes. *J Hydrol* 273:69–80
- Young KC (1994) A multivariate chain model for simulating climatic parameters from daily data. *J Appl Meteorol* 33: 661–671
- Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated models. *J Clim* 12:2474–2488
- Zorita E, Hughes JP, Lettenmaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *J Clim* 8:1023–1042
- Zwiers F, von Storch H (1990) Regime-dependent autoregressive time series modeling of the Southern Oscillation. *J Clim* 3:1347–1363

Appendix 1. Simulation of daily circulation indices

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 cannot be reproduced by classical autoregressive (AR) processes. Al-Awadhi & Jolliffe (1998) therefore studied the use of threshold autoregressive (TAR) models to describe time series of surface pressure in the UK. Zwiers & von Storch (1990) applied this class of 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 usually more than one index is needed to characterize the atmospheric circulation. Lall & Sharma (1996) showed that nearest-neighbour resampling is able to reproduce the nonlinear behaviour of a TAR model. Because the extension of nearest-neighbour resampling to the multivariate situation is straightforward, this method was used to simulate time series of the daily circulation indices Z , W and S .

Three unconditional nearest-neighbour simulation models for generating Z , W and S were examined. The different feature vectors \mathbf{D}_t are schematically shown in Fig. A1. In the first-order model (CIRC 1) the feature vector contains the 3 standardized circulation indices on Day $t-1$

Appendix 1 (continued)

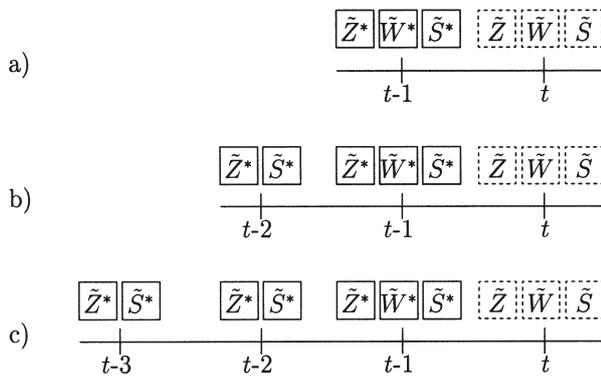


Fig. A1. Elements of the feature vector (solid boxes) for unconditional simulation of circulation indices Z , W , S (dashed boxes): (a) CIRC 1; (b) CIRC 2 and (c) CIRC 3. Asterisk indicates that the circulation index was simulated in a previous time step; tilde refers to a standardized value

with equal weights w_i . The feature vector of the second-order model (CIRC 2) contains standardized circulation indices for Days $t-1$ and $t-2$. 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, 0 and 1 respectively. The weight for \tilde{W}_{t-2} was set to zero because the autocorrelation structure of the W index closely resembles that of a first-order AR process. Finally, there is a third-order model (CIRC 3) for which \tilde{Z}_{t-3} and \tilde{S}_{t-3} are included in the feature vector both with unit weight (and the weight for \tilde{W}_{t-1} was set to 3). All models made again use of the decreasing kernel in Eq. (2) with $k = 5$, and the search for nearest neighbours was restricted to days within a 91 d moving window, centered on the calendar day of interest.

With the 3 models, 980 yr simulations were performed by resampling from the historical circulation indices of the 35 yr period 1961–1995. Table A1 presents the differences between the lag-1 and lag-3 autocorrelation coefficients of the simulated circulation indices and those of the historical record. The historical estimates and their jackknife standard errors (Buishand & Beersma 1993) are also given.

In the first-order model, the lag-3 autocorrelation coefficients of the Z and S indices are significantly underestimated. The other autocorrelation coefficients differ only slightly from the historical ones. For the second-order model there are no significant differences. The third-order model is no improvement compared to the second-order model because the lag-1 autocorrelation coefficients of Z and S are significantly underestimated.

Additionally, the reproduction of the average run length of 6 typical circulation

types was examined. Days were classified as cyclonic, strong westerly and southerly if the standardized values of respectively Z , W and S were larger than 1.0, and as anticyclonic, easterly and northerly if \tilde{Z} , \tilde{W} and \tilde{S} were smaller than 1.0. Table A2 presents the percentage differences between the average run lengths of the simulated indices and those of the historical indices.

The first-order model somewhat underestimates the average run lengths. The underestimation of the run lengths in the second-order model is worse but significant only for northerly flows. The third-order model significantly underestimates the average run lengths for the cyclonic and anticyclonic flows as well as for the southerly and northerly flows. In the historical record the average run length for easterly flows is about 25% larger than for strong westerly flows. All models are able to reproduce this asymmetry between the average duration of strong westerly and easterly flows.

In conclusion, the third-order model (CIRC 3) performs worse both in terms of lag-1 autocorrelation coefficients and average run lengths of circulation types. The second-order model performs better with respect to reproduction of the autocorrelation coefficients and the first-order model reproduces the run length statistics somewhat better. For the simulations conditional on simulated circulation indices in Sections 3.2 and 3.3, the second-order model (CIRC 2) was used to generate 980 yr of daily circulation indices.

Table A1. Differences between the lag-1 and lag-3 autocorrelation coefficients of daily circulation indices in 980 yr simulations and the historical data. Bottom lines: historical (1961–1995) estimates of $r(1)$ and $r(3)$ with their standard errors, SE. Z , W and S denote the relative vorticity, the west component and the south component of the flow respectively. Values in **bold** refer to statistically significant differences

Model	$r(1)$			$r(3)$		
	Z	W	S	Z	W	S
CIRC 1	-0.008	-0.008	-0.008	-0.091	0.008	-0.086
CIRC 2	-0.018	-0.008	-0.021	-0.014	0.017	-0.013
CIRC 3	-0.033	-0.010	-0.036	-0.015	0.011	-0.022
Historical	0.498	0.755	0.521	0.182	0.393	0.190
SE	0.012	0.006	0.011	0.010	0.012	0.011

Table A2. Relative differences (%) between average run lengths of 6 typical circulation types in the 980 yr simulations and the historical data. Bottom lines: historical (1961–1995) estimates of the average run lengths (d) with their relative standard errors, SE (%). Values in **bold** refer to statistically significant differences

Model	Cyclonic	Strong westerly	Southerly	Anti-cyclonic	Easterly	Northerly
	CIRC 1	-1.20	-1.00	-0.65	-0.39	-2.17
CIRC 2	-2.77	-2.17	-2.62	-2.47	-2.00	-4.53
CIRC 3	-5.26	-4.25	-4.00	-4.71	-3.50	-6.87
Historical	1.74	2.14	1.69	1.62	2.70	1.81
SE (%)	1.97	2.32	1.83	1.74	2.86	2.10