Conditional stochastic model for generating daily precipitation time series

Aristita Busuioc[1,*], Hans von Storch[2]

[1]**National Institute of Meteorology and Hydrology, Sos. Bucuresti-Ploiesti 97, Sector 1, Bucharest, Romania**
[2]**Institute for Coastal Research, GKSS Research Center, Max-Planck Straße 1, 21502 Geesthacht, Germany**

ABSTRACT: The purpose of this paper is the construction of a conditional stochastic model to generate daily precipitation time series. The model is a mixture of a 2-state first-order Markov chain and a statistical downscaling model based on canonical correlation analysis (CCA). The CCA model links the large-scale circulation, represented by the European sea-level pressure (SLP) field, with 4 precipitation distribution parameters: i.e. 2 transition probabilities and 2 gamma distribution parameters. This model is tested for the Bucharest station, for which long observed daily time series were available (1901–1999). The comparison of the capabilities of the conditional stochastic model and an unconditional stochastic model (based only on a Markov chain) is presented using ensembles of 1000 runs of the 2 models. The performance of the conditional stochastic model is analyzed in 2 steps. First, the ability of the CCA model for estimating the 4 precipitation distribution parameters is assessed. Second, the performance of both stochastic models in reproducing the statistical features of the observed precipitation time series is analyzed. The CCA model is most accurate for winter and autumn (transition probabilities), less accurate for the mean precipitation amount on wet days and inaccurate for the shape parameter. There are no significant dissimilarities between the conditional and unconditional models regarding their performance except for the linear trend and interannual variability, which are better captured by the conditional model. Some statistical features are well reproduced by both stochastic models for all seasons, such as mean and expected maximum duration of wet/dry intervals, daily mean of precipitation for wet days. Other statistical features are only partially reproduced by both models or are better reproduced by one of the models, such as mean duration of dry interval, standard deviation of daily precipitation amount, seasonal mean of rainy days and expected maximum daily precipitation. For all seasons, generally, the frequency of shorter dry intervals is underestimated and that of longer dry intervals (greater than 9 d) is overestimated. In conclusion, the conditional stochastic model presented in this paper can be used to generate daily precipitation time series for winter and autumn. For the other seasons, the unconditional model can be used to reproduce some statistical features.

KEY WORDS: Stochastic model · Precipitation · CCA · Markov chain · Monte Carlo experiments

1. INTRODUCTION

Hydrological and crop models usually require daily precipitation time series as input. To evaluate the sensitivity of these models to long-term changes in the precipitation regime, an ensemble of input data sets is needed. The observed sequences provide only one realization of the weather process. In impact studies, which use as input data precipitation time series derived from regional/global climate model (RCM/GCM) simulations, the number of these sequences is limited due to the high computational cost. To evaluate the range of results obtained with other statistically equivalent series, it is desirable to generate synthetic sequences of precipitation data based on the stochastic structure of the meteorological process. Richardson (1981) presented such a technique to simulate daily values of precipitation, maximum and minimum tem-

*Email: busuioc@meteo.inmh.ro

perature, and solar radiation. For the precipitation component, a 2-state first-order Markov chain has been used to describe the precipitation occurrence, and the exponential distribution has been used to approximate the distribution of rainfall amount. This model is termed Richardson's model or WGEN. Katz (1996) presented the theoretical statistical properties of a simplified version of Richardson's model and found that when its parameters are varied certain unanticipated effects can be produced. This model has also been used by Wilks (1992) with the gamma distribution instead of the exponential distribution for rainfall amount, adapting the model for climate change studies.

Such models can be conditioned on large-scale meteorological conditions, which incorporate cause-and-effect information about the probability of wet or dry conditions. There are various ways to define the large-scale conditions. One way is to use large-scale circulation indices. Such a conditional stochastic model to generate daily precipitation sequences has been developed by Katz & Parlange (1993) using a large-scale circulation index defined by 2 states (above normal and below normal) of the monthly sea-level pressure (SLP) at a specified grid point. Other frequently used approaches are circulation classification (Zorita et al. 1995, Goodess & Palutikof 1998) and analogs (Zorita et al. 1995). Lettenmaier (1995) showed the advantages and disadvantages of various models and presented a synthesis of the stochastic models of precipitation (conditional and unconditional). Non-linear approaches, such as neural networks, have recently been developed (Cavazos 1999, Zorita & von Storch 1999). In a recent synthesis of empirical downscaling methods used in synoptic climatology, Yarnal et al. (2001) reviewed the advantages and disadvantages of stochastic models. They found, in the case of climate change, that the conditioning of the stochastic parameters in a physically meaningful way is difficult to achieve.

In this paper, a stochastic model conditioned upon large-scale climate characteristics to generate daily precipitation amount is presented. The model uses a first-order Markov chain combined with a downscaling model. To link the precipitation distribution parameters with the large-scale circulation, represented by SLP on the European scale, a regression model based on canonical correlation analysis (CCA) is used (von Storch et al. 1993, Heyen et al. 1996, Busuioc et al. 1999, 2001). In this way, an adjustment of stochastic parameters in a physically meaningful way is proposed. The model is tested for the Bucharest station, for which long daily observations are available (1901–1999). The ability of the downscaling model for estimating the parameters of precipitation distribution,

and the capability of the conditional stochastic model for reproducing the most important statistical features of the observed precipitation time series, are shown in Sect. 3. The comparison between the abilities of the conditional and the unconditional stochastic models is also presented. Compared with similar models (e.g. Katz & Parlange 1993) the procedure presented in this paper gives confidence intervals of the precipitation distribution parameters derived from 1000 Monte Carlo runs of both models.

2. METHODOLOGY

The model presented in this paper is a combination between a first-order Markov chain and a statistical downscaling model. In the following it is referred to as a *conditional stochastic model.* Additionally, an *unconditional stochastic model*, based only on a first-order Markov chain, is used to assess the performance of the conditional model for climate change purposes. Both models are tested for daily precipitation amount at the Bucharest station in the southern part of Romania. Observational data refer to the period 1901–1999, and they are seasonally stratified: winter (December–February), spring (March–May), summer (June–August) and autumn (September–November).

2.1. Unconditional stochastic model

Precipitation occurrence is described by a 2-state, first-order Markov chain. Precipitation either occurs or it does not (the 2 states), and the conditional probability of precipitation occurrence depends only on the occurrence on the previous day. There are 2 parameters describing the precipitation occurrence process: the transition probabilities p_{01}, the probability of a wet day following a dry day, and p_{11}, the probability of a wet day following a wet day. As a wet-day threshold, a daily precipitation amount of >0.1 mm is used in this paper. The choice of the optimum precipitation threshold is an important decision. Dobi-Wantuch et al. (2000) analysed the threshold influence on the results for 2 Hungarian stations (one of them being near the Romanian border) and found the 0.1 mm threshold to be appropriate.

The variation of precipitation amount on wet days is described by the gamma distribution, which has 2 parameters: the shape parameter (k) and the scale parameter (β) (Coe & Stern 1982, Wilks 1992). In terms of the 2 distribution parameters, the mean precipitation amount (considering only wet days) is $\mu = k\beta$. In this paper μ and k are defined as the gamma distribution parameters. μ is estimated as the sample mean

from the observed data set, and k is derived as the solution of:

$$\ln k - \psi(k) = \overline{\ln x} - \overline{\ln x}$$

where

$$\overline{x} = \left(\sum_{i,t} x_i(t)\right)\Big/n$$

and $\psi(k)$ is the first derivative of the log gamma function and is obtained using a computational subroutine by Amos (1983); in our case, $x(t)$ represents the daily precipitation amount for wet days.

The p_{01}, p_{11} transition probabilities were estimated from the observed data set. Therefore, the stochastic model to generate daily precipitation depends on 4 parameters (p_{01}, p_{11}, μ and k) and is referred to as the *unconditional model*. The 4 parameters were computed over the complete 1901–1999 period for every month, and their seasonal variation evaluated. Based on this, the data set was divided into the 4 seasons, and the model built for every season.

In conclusion, daily precipitation for wet days is generated (in the case of the unconditional model) using the 4 parameters computed from observations over a fixed period following the procedure presented by Wilks (1992). Knowing whether precipitation occurred on the previous simulated day, the appropriate transition probability, p_{01} or p_{11}, is compared to a newly generated uniform [0, 1] random number. A wet day is simulated if the random number is less than the transition probability. If this is the case, a random precipitation amount is generated for the current day using the appropriate gamma distribution. Since the 4 parameters are kept constant, stationary precipitation time series are generated. This is a disadvantage of the unconditional model when used in the climate change context. However, this drawback can be overcome by using the conditional stochastic model, which links unconditional stochastic model parameters to large-scale circulation parameters as described below.

2.2. Conditional stochastic model

To link the 4 parameters of the unconditional stochastic precipitation model to the large-scale circulation, a regression model is constructed with the help of CCA (von Storch et al. 1993, Heyen et al. 1996, Werner & von Storch 1993, Busuioc et al. 1999, 2001). European-scale SLP anomalies are chosen as representing the large-scale circulation. Monthly SLP data with a resolution of $5° \times 5°$ were provided by the National Center of Atmospheric Research (NCAR, USA) (Trenberth & Paolino 1980). The SLP area of $5–50°\,E$, $30–55°\,N$ was selected, so that the accuracy of the downscaling model linking seasonal precipitation over

Romania and SLP is maximized, since seasonal total precipitation in Romania is strongly connected to the European SLP distribution, especially for winter and autumn (Busuioc & von Storch 1996, Busuioc et al. 1999). Therefore, a strong connection between the 4 parameters characterizing the daily precipitation distribution and the SLP field is assumed to exist. In Sect. 3, it will be shown that this assumption is correct, especially for winter and autumn.

The stochastic parameters (p_{01}, p_{11}, μ and k) are computed for every season using 90 to 92 daily precipitation amounts in every year. In this way, a time series of the parameters is obtained. Prior to the CCA, the original data are standardized by subtracting the mean from the original value and dividing by the SD. The CCA determines pairs of patterns of 2 time-dependent variables (the large-scale SLP and the 4 stochastic parameters) so that their time components are optimally correlated. The SLP data are projected onto their EOFs (empirical orthogonal functions) to eliminate noise (small-scale features) and to reduce the dimensions of the data. Since the time coefficients are normalized to unity, the canonical correlation patterns represent the typical strength of the signals. A subset of CCA pairs is then used in a regression model to estimate the 4 stochastic parameters from large-scale SLP. The precipitation distribution parameters (p_{01}, p_{11}, μ, k) estimated through the CCA model are then used in the stochastic model in order to generate daily precipitation amounts. These time series are generated for every season in every year. Since the 4 parameters should satisfy the conditions $0 \leq p_{01}$, $p_{11} \leq 1$ and μ, $k > 0$, the CCA model outputs are processed by applying the reversed standardization procedure before being used in the stochastic model.

2.3. Accuracy of the stochastic models

The full data set, 1901–1999, was split into 2 intervals, 1901–1949 and 1950–1999. Then, both unconditional and conditional models were fitted with one interval and validated with the other interval, so that the 2 models were fitted and validated independently. The accuracy of the CCA downscaling model is expressed through the variance explained by the reconstructed values as a fraction from the total variance of the observed values or, alternatively, by the correlation between observed and reconstructed values. Additionally, the performance of the stochastic models was determined in terms of how well the model reproduces the following statistical features of the observed precipitation time series:
- Appearance/nonappearance of precipitation quantified by mean and expected maximum duration of wet and dry intervals.

- Daily mean and standard deviation of precipitation for wet days, expected maximum daily precipitation amount and frequency distributions of daily precipitation.
- Changes (linear trend) in the seasonal precipitation amount induced. The significance of the linear trend is estimated by the Mann-Kendall statistic (Sneyers 1975).
- Interannual variability of seasonal precipitation amount quantified by the SD.

These statistical features were computed for the 2 subintervals.

An ensemble of 1000 simulations was generated for each model, and the statistical parameters listed above were expressed as ensemble means with their 90% confidence intervals computed with a bootstrapping procedure.

3. RESULTS

3.1. Precipitation distribution parameters

CCA identified pairs of patterns in the SLP fields and in combined vectors of the transition probabilities (p_{01}, p_{11}) and gamma distribution parameters (μ, k) whose time series share a maximum of correlation. The correlation coefficients R_1 and R_2 associated with the first 2 CCA pairs for the 4 seasons are presented in Table 1. The explained variance of the seasonal SLP anomalies and 4 parameter anomalies is also presented. The strongest link was found for winter and autumn. Similar results were achieved when the direct relationship between the SLP field and seasonal total precipitation in Romania was analyzed (Busuioc & von Storch 1996, Busuioc et al. 1999). Figs. 1 to 5 show the patterns of the first 2 CCA pairs for all seasons.

The CCA was performed for 2 subintervals: 1901–1949 and 1950–1999. The first CCA pair is almost identical for the 2 subintervals for all seasons except for a slight shift and spatial extension of the pattern nucleus. Thus only the first subinterval is shown for spring to autumn (Figs. 3 to 5).

Southerly/northwesterly circulation over Romania in winter is associated with above/below-normal daily mean precipitation on wet days and higher/lower transition probabilities to wet days at Bucharest (Figs. 1 & 2). Southwesterly circulation brings moist air masses from the Mediterranean basin to Romania (especially in the southern part, where Bucharest is located), and more precipitation is recorded. As a result, the daily mean precipitation on wet days (μ) and the wet day probability ($p_{01} + p_{11}$) are higher. However, the link with μ is mostly weak in all seasons, except autumn. This seems to contradict the presence of a strong link between SLP and total seasonal precipitation amount (Busuioc & von Storch 1996, Busuioc et al. 1999). However, these 2 observations may be reconciled by noting that the seasonal total is dominated by the number of wet days and less dependent on the mean amount on wet days. In fact, the correlation between μ and the first SLP EOF (very similar to the first CCA pattern) time series is low, while the correlation with the number of wet days is high. The link between SLP variations and k variability is unclear, with the sign of the link changing when different fitting periods are used (compare Figs. 1a and 2a).

A negative/positive SLP pattern centered over the northwest of the study area (spring) or southeastern Europe (autumn) is associated with positive anomalies for p_{01}, p_{11} and μ and negative anomalies for k (Figs. 3a & 5a). In summer, a similar structure (centered in the southeastern part of the Black Sea) is associated with positive anomalies of transition probabilities and slightly negative anomalies of daily mean precipitation amount. These results are not in agreement with those obtained previously through analysis of the relationship between SLP and summer Romanian precipitation totals (Busuioc & von Storch 1995). The reason could be that either the high variability of summer daily precipitation in Romania is not that strongly linked with

Table 1. Canonical correlation coefficients of the first 2 CCA pairs (R_1, R_2) and proportion of variance explained for patterns of seasonal mean sea-level pressure (SLP) and the 4 precipitation distribution parameters described by the first 2 CCA pairs

		1901–1949						1950–1999			
	R_1	R_2	Explained variance (%)				R_1	R_2	Explained variance (%)		
			SLP		Parameters				SLP		Parameters
			CCA_1	CCA_2	CCA_1	CCA_2			CCA_1 CCA_2		CCA_1 CCA_2
Winter	0.77	0.49	45	7	29	27	0.78	0.42	38 21		33 9
Spring	0.66	0.53	31	6	24	25	0.51	0.40	22 8		29 22
Summer	0.67	0.55	11	10	30	22	0.65	0.44	13 7		30 21
Autumn	0.82	0.48	16	34	34	18	0.80	0.52	10 37		29 32

a **cca1 slp cor = 0.77 var = 54%**   **cca2 slp cor = 0.49 var = 7%** **b**
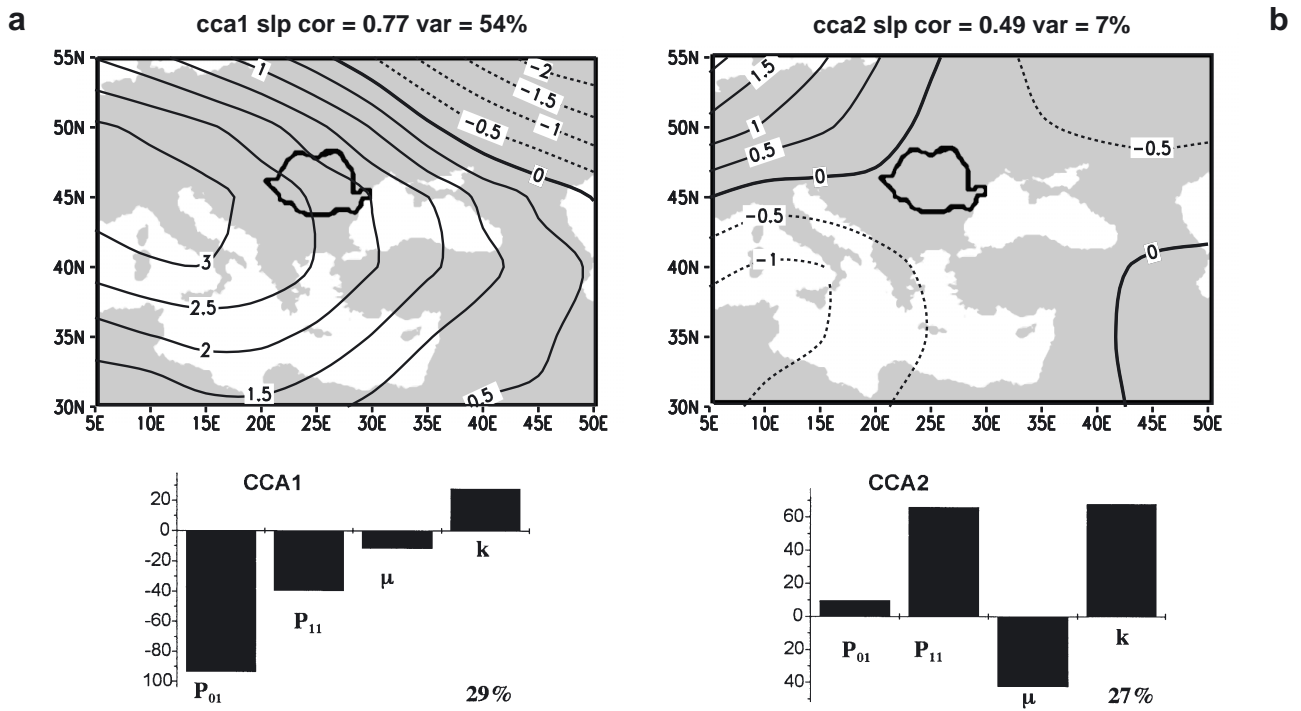
Fig. 1. Patterns of the first 2 CCA pairs of winter mean SLP and winter parameters of precipitation distribution derived from the first half of the observations (1901–1949)
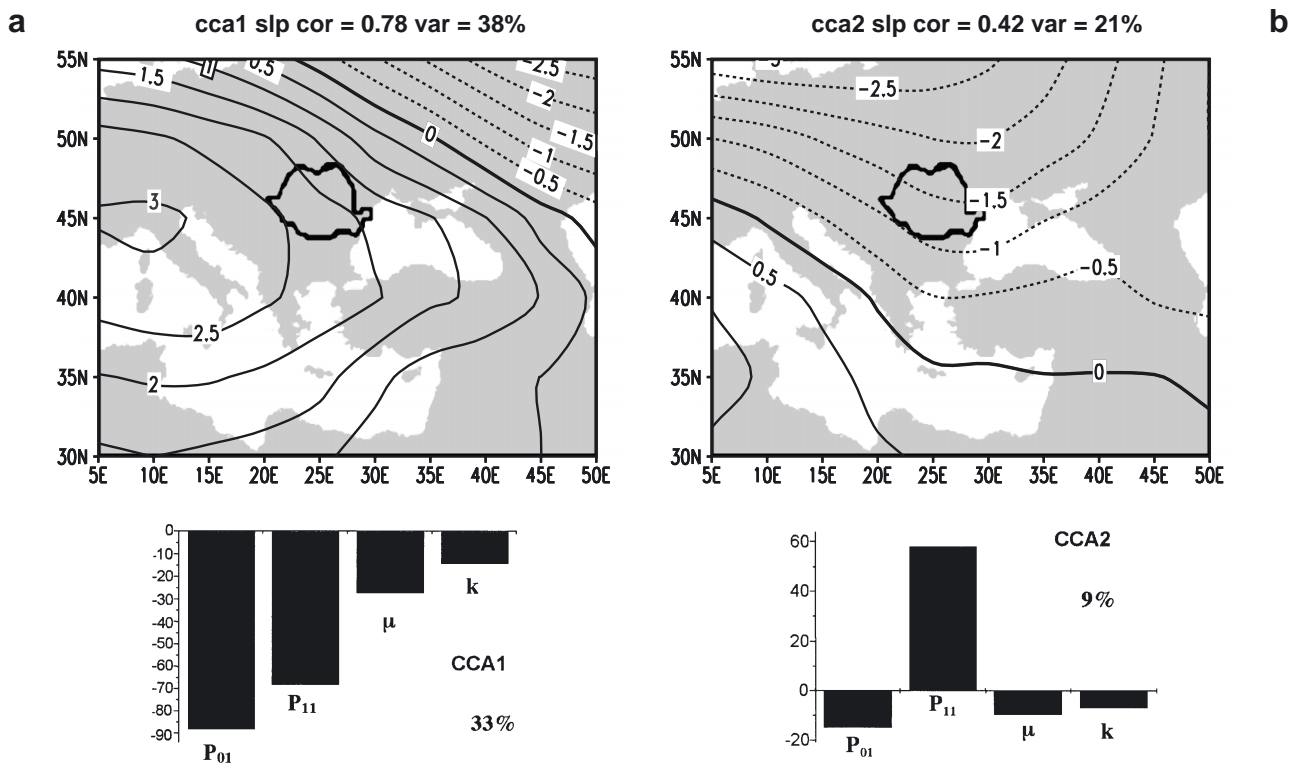
a **cca1 slp cor = 0.78 var = 38%**   **cca2 slp cor = 0.42 var = 21%** **b**

Fig. 2. Patterns of the first 2 CCA pairs of winter mean SLP and winter parameters of precipitation distribution derived from the second half of the observations (1950–1999)
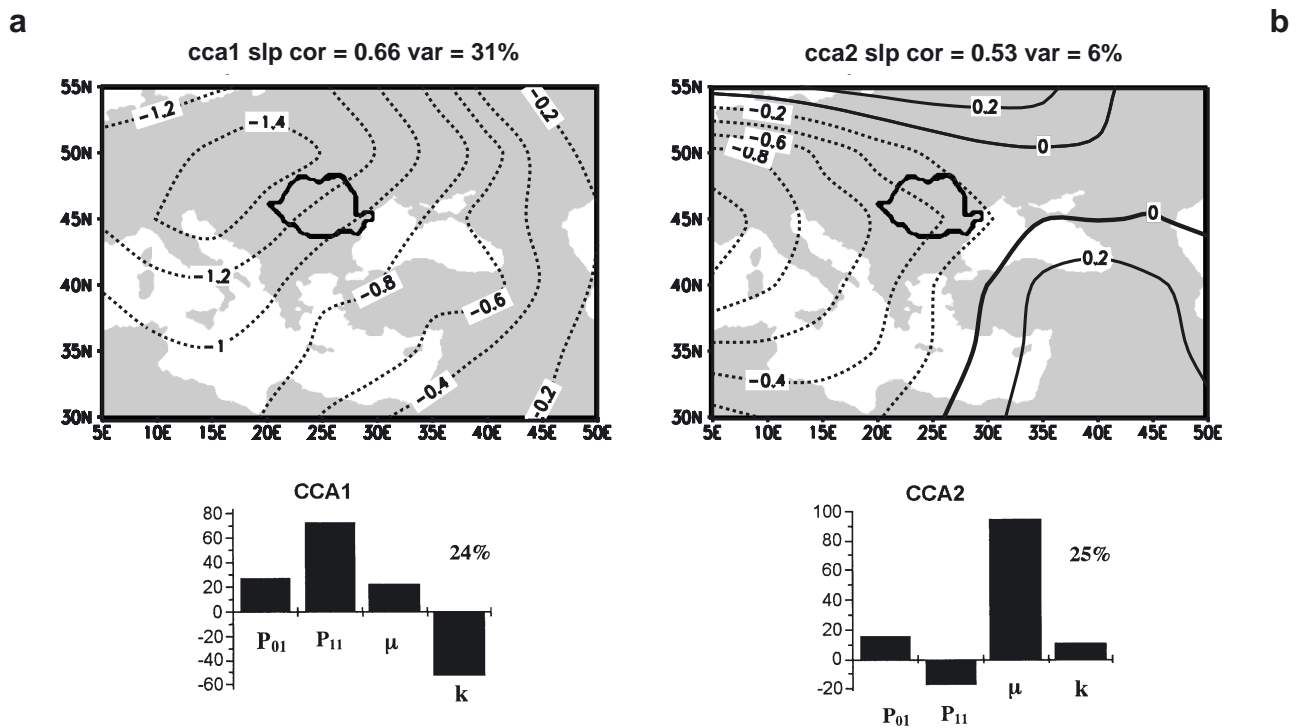
Fig. 3. Patterns of the first 2 CCA pairs of spring mean SLP and spring parameters of precipitation distribution derived from observations (1901–1949)
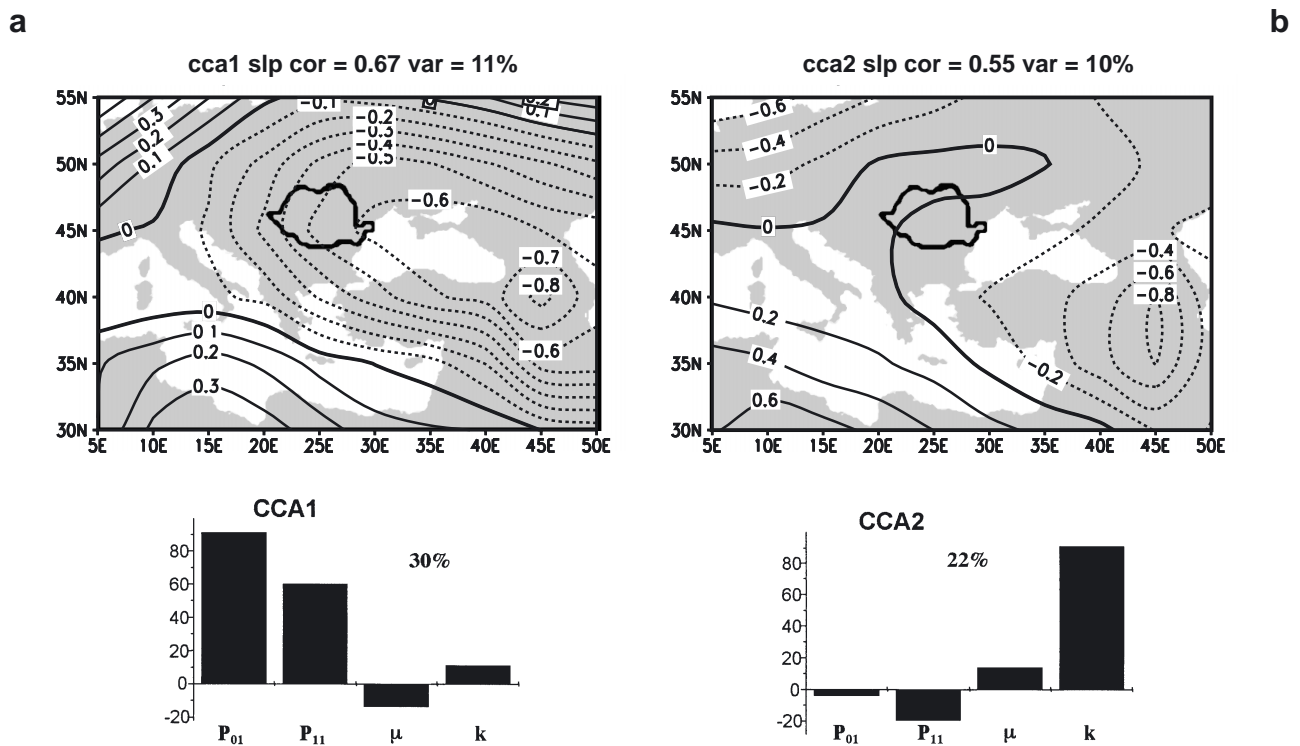

Fig. 4. Patterns of the first 2 CCA pairs of summer mean SLP and summer parameters of precipitation distribution derived from observations (1901–1949)
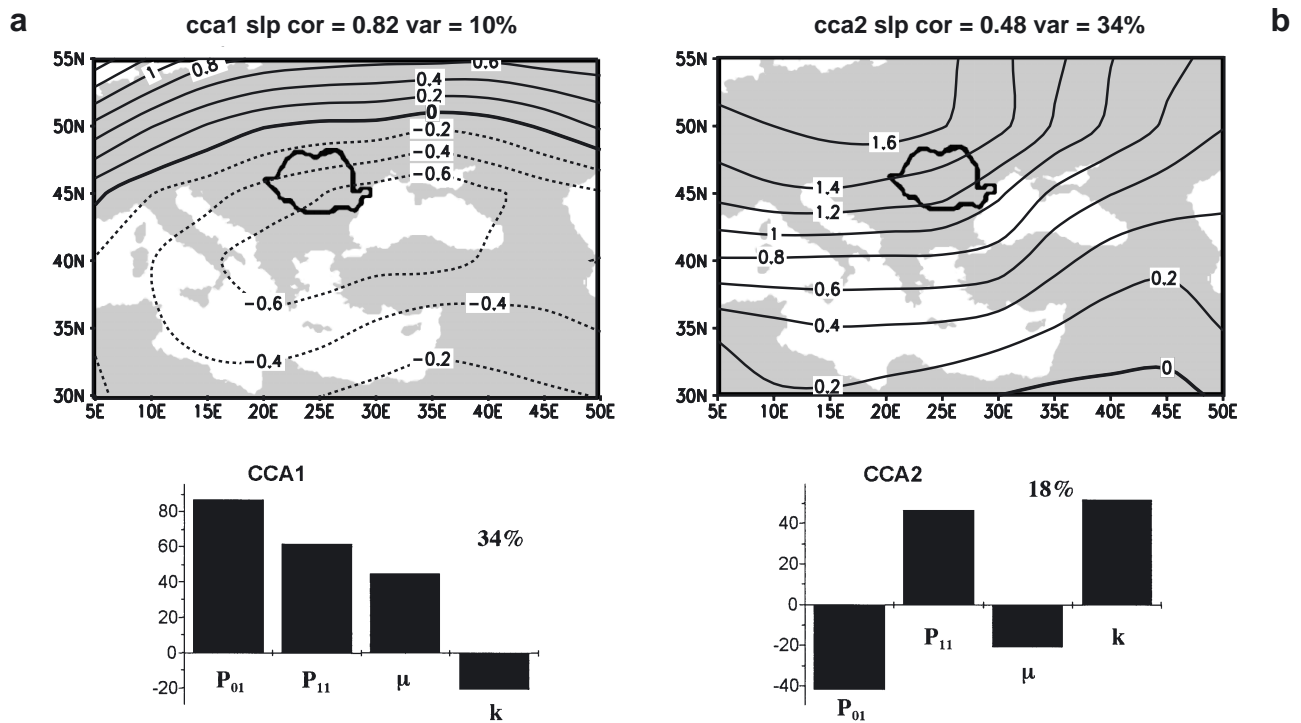
Fig. 5. Patterns of the first 2 CCA pairs of autumn mean SLP and autumn parameters of precipitation distribution derived from observations (1901–1949)

SLP variability, or it is controlled by other physical mechanisms, such as convective processes. In conclusion, the k variability does not seem to be linked with SLP variability in any season.

The SLP pattern of the second CCA pair is different over the 2 subintervals for winter (compare Figs. 1b and 2b), but for the other seasons they are stable, apart from slight shifts in the center of the SLP pattern (not shown). The mechanisms given by the second CCA pair cannot be physically interpreted in the same manner as those presented for the first CCA pair, since there is no coherence between the variability of the precipitation distribution parameters and SLP variability. For example, the positive winter SLP anomalies (Fig. 2b) are associated with the positive anomalies of p_{11} and negative anomalies of the others, which is not consistent with the physically reasonable mechanism presented above.

The skill of the statistical downscaling model constructed using the time coefficients associated with the 4 CCA pairs is summarized in Table 2. The skill is calculated for the independent subintervals not used to fit the statistical model. The model is most skillful for winter and autumn (transition probabilities), slightly skillful for μ (winter, especially for the 1950–1999 period) and unskillful for k. Figs. 6 & 7 show, as an example, the temporal evolution of the observed and estimated standardized anomalies for winter and autumn. The 2 curves vary coherently for the transition probabilities. The amplitude is sometimes different, but the year-to-year evolution is quite good for μ, but for k the dissimilarities are substantial.

Table 2. Accuracy of the CCA model (expressed as percentage of explained variance/correlation coefficient) for estimation of the 4 parameters (p_{01}, p_{11}, μ, k) from the SLP field over the 2 subintervals used as independent data sets

| | 1901–1949 | | | | 1950–1999 | | | |
	p_{01}	p_{11}	μ	k	p_{01}	p_{11}	μ	k
Winter	45/0.67	16/0.43	4/0.23	–	39/0.63	24/0.49	11/0.33	–
Spring	–	7/0.26	–7/0.26	–7/0.21	9/0.30	–	–3/0.18	–
Summer	19/0.14	8/0.30	–	–	2/0.39	6/0.34	–	–
Autumn	50/0.71	21/0.47	–14/0.14	–	51/0.72	15/0.38	–12/0.8	–

Fig. 6. Winter standardized anomalies of the precipitation distribution parameters for 1901–1999 derived from observations (solid line) and derived indirectly from the observed European-scale SLP anomalies using the downscaling model (dashed line) fitted to the 1901–1949 data
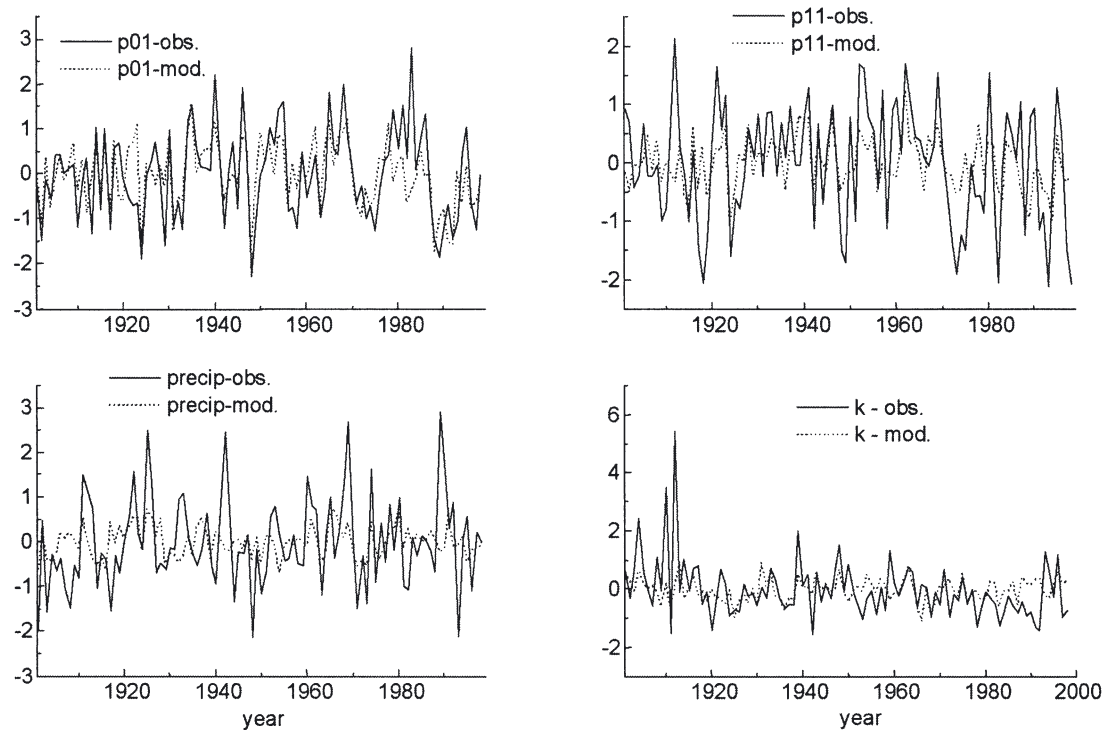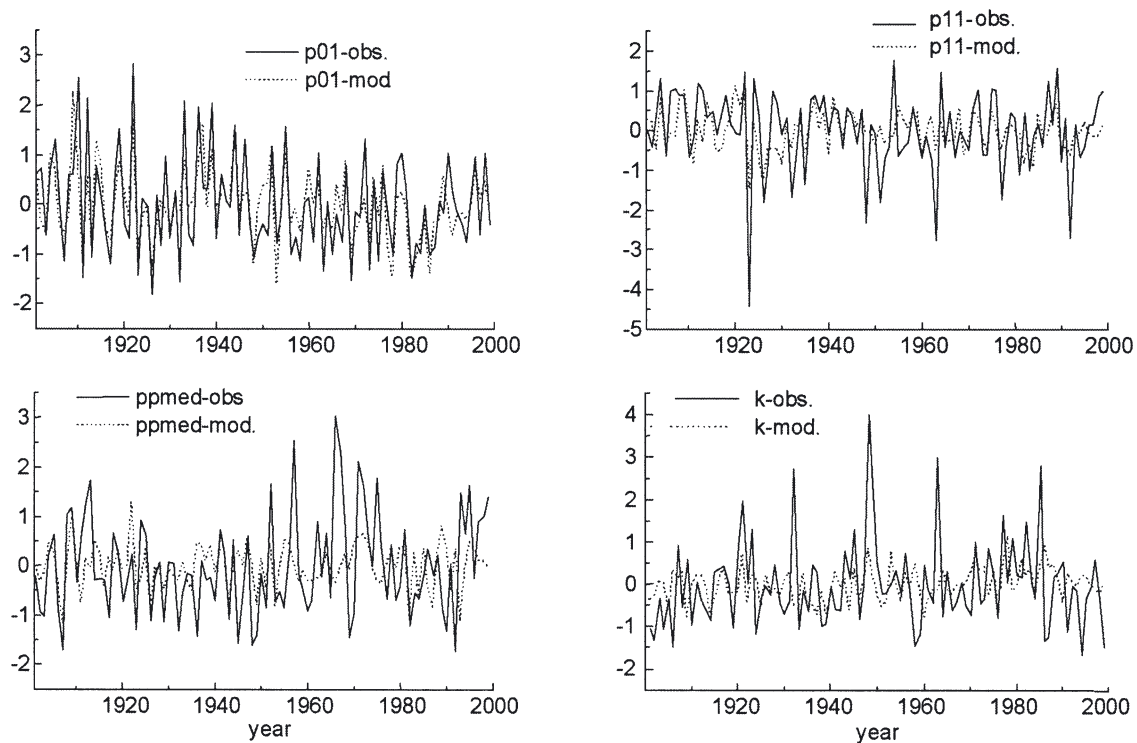

Fig. 7. Autumn standardized anomalies of the precipitation distribution parameters for 1901–1999 derived from observations (solid line) and derived indirectly from the observed European-scale SLP anomalies using the downscaling model (dashed line) fitted to the 1901–1949 data

3.2. Accuracy of the stochastic models

The 4 parameters (p_{01}, p_{11}, μ and k), estimated directly from precipitation data (in the case of the unconditional model) and indirectly from European SLP through the statistical downscaling model (in the case of the conditional model) are used in the Markov chain model to generate daily time series of precipitation occurrence. The daily precipitation amount is randomly generated on wet days using a gamma distribution. The performance of these stochastic models is assessed in terms of how well they reproduce the statistical features of the observed precipitation time series listed in Section 2.3. These features are represented by: maximum duration of dry and wet intervals (d_{dry}^{max}, d_{wet}^{max}), mean duration of dry and wet intervals (d_{dry}^{mean}, d_{wet}^{mean}), daily mean/standard deviation of precipitation on wet days (pp_{mean}, pp_{SD}), mean number of wet days (nr), Mann-Kendall statistic (t), expected maximum of daily precipitation amount (pp_{max}), standard deviation of seasonal precipitation amount and frequency distributions of daily precipitation within various intervals. After running the unconditional and conditional models 1000 times, a distribution of these parameters was achieved. Then, the ensemble mean and 90% confidence intervals of the respective parameters were computed.

These statistics derived from both stochastic models and directly from the observations are presented in Tables 3 to 5. They were computed separately for the 2 subintervals. d_{wet}^{mean} and d_{wet}^{max} are very well simulated for all seasons, both subintervals and both models (unconditional and conditional). There are no statistically significant differences between the unconditional and conditional models. However, d_{dry}^{max} is better estimated by the unconditional model and overestimated by the conditional model, although for both models the observed values are generally covered by the simulated 90% confidence intervals. d_{dry}^{mean} is well simulated by both models for winter and summer (1901–1949) and autumn (1950–1999). Both models except for spring (both intervals) overestimate the observed values. nr is generally well reproduced by both stochastic models, except for spring (1901–1949, both models, and 1950–1999, conditional model) and autumn (1950–1999, both models), when it is underestimated by the conditional model and overestimated by the unconditional model. pp_{mean} is generally well reproduced by both stochastic models (with small differences for 1950–1999), while pp_{SD} is better estimated for 1901–1949 (both models), except for summer, when it is underestimated. For 1950–1999 both models underestimate pp_{SD} for all seasons (but less underestimated in summer). Expected maximum daily precipitation is generally underestimated for all seasons

(both models) but the observed values are covered by the 90% confidence intervals, except for summer (1901–1949) and winter and autumn (1950–1999).

In order to assess whether the stochastic models reproduce the observed linear trends, the Mann-Kendall statistic (t) was computed for the seasonal precipitation amount as derived directly from observations and indirectly from generated daily time series. A t-value greater than 1.96 allows the rejection of the null hypothesis with a risk of 5% or less, provided that the data are not serially correlated. This t-statistic was computed for every experiment of the 1000 run ensemble. Table 3 gives the ensemble means of this statistic with their 90% confidence intervals.

As expected, the unconditional model does not show any significant trend. The observed data set reveals a statistically significant linear trend at the 5% level only for the winter season over 1901–1949, namely an increasing one. For 1950–1999 a slightly decreasing trend is identified. The conditional stochastic model correctly simulates the direction of these trends, but the magnitude is underestimated for the first period and overestimated for the second one. The simulated 90% confidence intervals of the t-statistic generally cover the observed values. For the other seasons the observed trends are not statistically significant and the conditional stochastic model reproduces this characteristic well.

The main reason for these results could be the urban effect of Bucharest, which leads to more precipitation there than at stations located in similar geographical conditions (Neacsa et al. 1974). This is supported by the analysis of the large-scale circulation's variability connected to the seasonal precipitation regime in Romania. Previous papers (Busuioc & von Storch 1995, 1996, Busuioc et al. 1999) have shown that there is a strong link between the large-scale SLP and seasonal total precipitation at 14 Romanian stations distributed over the entire country. For the winter season the first SLP EOF gives this link, and the shifts found in the Romanian precipitation are similar to those found in the large-scale circulation. For example, the southwestern stations reveal a precipitation increase around 1933 (due to more frequent southwesterly circulation over Romania) and a decrease after 1969–1970 (due to less frequent southwesterly circulation). Only the Bucharest station does not fit into this pattern, as precipitation has increased since 1922. This result is related to the urban effect. For the present study, the time series associated with the first SLP EOF were analyzed for the 2 subintervals. For 1901–1949 a slightly decreasing trend (not significant using the Mann-Kendall test) led to a simulated precipitation close to normal. For 1950–1999 the situation is reversed: the time series associated with the first SLP EOF has a very

strong decreasing trend (associated with less frequent southwesterly circulation over Romania) that induces less precipitation. In the observations, the urban effect, associated with an increase in precipitation, is superimposed on the natural mechanism given by large-scale circulation, leading to more precipitation than simulated using the large-scale SLP.

Using the standard deviations of the generated seasonal precipitation amount, which were computed for every experiment of the 1000-run ensemble, the accuracy of the stochastic models related to the interannual variability was assessed. Table 4 gives the ensemble means and simulated 90% confidence intervals along with those derived from observations. The standard

Table 3. Statistics of precipitation regime at the Bucharest station derived directly from observations and indirectly from the stochastic conditional and unconditional models: maximum duration (d) of dry and wet intervals d_{dry}^{max}, d_{wet}^{max}, mean duration of dry and wet intervals d_{dry}^{mean}, d_{wet}^{mean}, daily mean/standard deviation of precipitation (mm) on rainy days pp_{mean}, pp_{SD}; mean number of rainy days nr; Mann-Kendall statistic t, expected maximum of daily precipitation amount pp_{max} (mm^{d-1}). These statistics are computed as ensemble means for 1000 runs and are derived over the 2 subintervals. 90% confidence intervals in parentheses

		d_{dry}^{max}	d_{wet}^{max}	d_{dry}^{mean}	d_{wet}^{mean}	pp_{mean}	pp_{SD}	nr	t	pp_{max}
1901–1949										
Winter	Observations	24	10	4.3	2.1	3.8	5.4	28	2.0	58.6
	Conditional	35	12	4.3	2.1	4.1	5.1	30	1.0	44.1
		(25, 51)	(9, 16)	(4.1, 4.5)	(2.0, 2.2)	(3.9, 4.4)	(4.7, 5.5)	(28, 31)	(−0.2, 2.2)	(33.6, 60.4)
	Unconditional	27	12	4.3	2.1	4.2	5.2	29	0.02	45.1
		(21, 36)	(9, 16)	(4.1, 4.6)	(2.0, 2.2)	(4.0, 4.4)	(4.9, 5.7)	(29, 31)	(−1.6, 1.7)	(34.5, 60.1)
Spring	Observations	25	13	3.7	2.0	4.6	6.4	31	−0.6	61.6
	Conditional	33	12	5.0	2.1	4.9	6.3	27	1.4	57.8
		(25, 45)	(9, 16)	(4.7, 5.3)	(2.0, 2.2)	(4.6, 5.2)	(5.8, 6.9)	(26, 29)	(0.2, 2.7)	(42.8, 81.9)
	Unconditional	24	12	3.9	2.2	4.9	6.3	34	0.01	54.8
		(19, 33)	(10, 16)	(3.7, 4.1)	(2.1, 2.3)	(4.7, 5.2)	(5.9, 6.7)	(32, 3.5)	(−1.6, 1.7)	(41.6, 72.8)
Summer	Observations	29	9	4.0	1.9	7.0	10.8	28	0.1	136.6
	Conditional	27	10	3.9	1.8	6.7	8.6	29	0.9	75.6
		(20, 35)	(8, 13)	(3.8, 4.1)	(1.8, 1.9)	(6.3, 7.1)	(7.9, 9.3)	(28, 30)	(−0.6, 2.3)	(56.7, 101.6)
	Unconditional	25	10	4.0	1.8	6.7	8.7	29	0.5	74.4
		(20, 34)	(8, 13)	(3.8, 4.2)	(1.8, 1.9)	(6.3, 7.1)	(8.0, 9.3)	(28, 31)	(−0.9, 2.0)	(55.8, 99.7)
Autumn	Observations	30	12	4.7	2.0	5.2	7.7	25	−1.1	59.6
	Conditional	40	10	6.0	1.9	6.0	7.6	23	0.1	66.8
		(29, 54)	(8, 14)	(5.3, 5.6)	(1.8, 2.0)	(5.6, 6.3)	(6.9, 8.4)	(22, 25)	(−1.0, 1.2)	(49.0, 91.9)
	Unconditional	36	10	5.6	1.9	5.9	7.6	23	0.1	63.6
		(27, 47)	(8, 13)	(5.3, 6.0)	(1.8, 2.0)	(5.5, 6.3)	(7.0, 8.2)	(22, 25)	(−1.3, 1.5)	(48, 85.7)
1950–1999										
Winter	Observations	31	11	4.0	2.1	4.2	6.0	29	−1.0	60.9
	Conditional	36	12	4.5	2.1	3.8	4.4	29	−1.8	39.1
		(26, 51)	(9, 16)	(4.3, 4.7)	(2.0, 2.2)	(3.6, 4.0)	(4.1, 4.8)	(28, 30)	(−3.0, −0.6)	(29.0, 54.2)
	Unconditional	29	12	4.6	2.2	3.8	4.6	29	0.02	38.5
		(23, 39)	(9, 16)	(4.3, 4.8)	(2.1, 2.3)	(3.6, 4.0)	(4.3, 4.9)	(28, 31)	(−1.7, 1.7)	(30.0, 51.2)
Spring	Observations	28	12	3.7	2.1	4.9	7.2	33	−0.4	58.8
	Conditional	32	12	5.1	2.0	4.6	5.5	26	−0.5	47.6
		(25, 43)	(9, 16)	(4.8, 5.3)	(1.9, 2.1)	(4.3, 4.8)	(5.1, 5.9)	(25, 28)	(−2.0, 1.0)	(35.3, 64.8)
	Unconditional	25	11	3.9	2.0	4.6	5.7	32	0.02	49.0
		(19, 33)	(9, 15)	(3.7, 4.1)	(2.0, 2.1)	(4.4, 4.8)	(5.3, 6.1)	(30, 33)	(−1.7, 1.6)	(37.1, 65.0)
Summer	Observations	28	10	3.7	1.8	6.7	9.7	29	1.1	88.2
	Conditional	29	11	4.2	1.9	7.0	8.7	29	−0.2	75.3
		(22, 40)	(8, 15)	(4.0, 4.5)	(1.8, 2.0)	(6.6, 7.4)	(8.1, 9.4)	(27, 30)	(−1.6, 1.4)	(57.1, 101.9)
	Unconditional	27	11	4.3	1.9	7.0	8.8	29	−0.02	75.5
		(21, 36)	(8, 14)	(4.1, 4.5)	(1.9, 2.9)	(6.6, 7.4)	(8.2, 9.5)	(27, 30)	(−1.6, 1.5)	(58.2, 100.2)
Autumn	Observations	31	11	5.1	1.9	5.9	8.8	23	1.1	93.0
	Conditional	39	11	5.0	2.0	5.0	6.2	26	−0.3	53.1
		(28, 56)	(8, 14)	(4.8, 5.3)	(1.9, 2.1)	(4.7, 5.3)	(5.7, 6.7)	(24, 27)	(−1.6, 0.9)	(39.5, 71.2)
	Unconditional	33	12	5.2	2.1	5.2	6.8	27	0.02	58.0
		(25, 44)	(9, 15)	(4.9, 5.5)	(2.0, 2.2)	(4.9, 5.6)	(6.2, 7.4)	(25, 28)	(−1.5, 1.6)	(43.7, 78.4)

Table 4. Standard deviations of seasonal precipitation (mm season^{-1}) derived from the observed data set and from the generated time series through conditional and unconditional stochastic models. The values are computed as ensemble means of 1000 runs over the 2 subintervals. 90% confidence intervals for the conditional and unconditional stochastic models in parentheses

	1901–1949				1950–1999			
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn
Observations	42.6	48.7	65.0	67.3	56.7	66.9	66.4	65.6
Conditional	47.3	58.8	67.8	71.3	42.5	43.0	68.0	54.4
	(38.7, 56.1)	(47.6, 71.1)`	(56.0, 80.9)	(58.6, 84.7)	(35.0, 50.0)	(34.6, 52.1)	(54.3, 81.4)	(45.1, 64.2)
Unconditional	34.2	40.4	54.4	53.4	31.5	36.8	57.3	39.6
	(26.4, 43.5)	(31.9, 50.1)	(41.8, 67.5)	(43.7, 63.3)	(24.9, 39.7)	(28.8, 45.4)	(44.7, 72.1)	(31.4, 49.7)

Table 5. Seasonal mean frequencies of precipitation (d season^{-1}) derived from the observed data set and from the generated time series through conditional and unconditional stochastic models. The values are computed as ensemble means of 1000 runs over the 2 subintervals. 90% confidence intervals for the conditional and unconditional stochastic models in parentheses

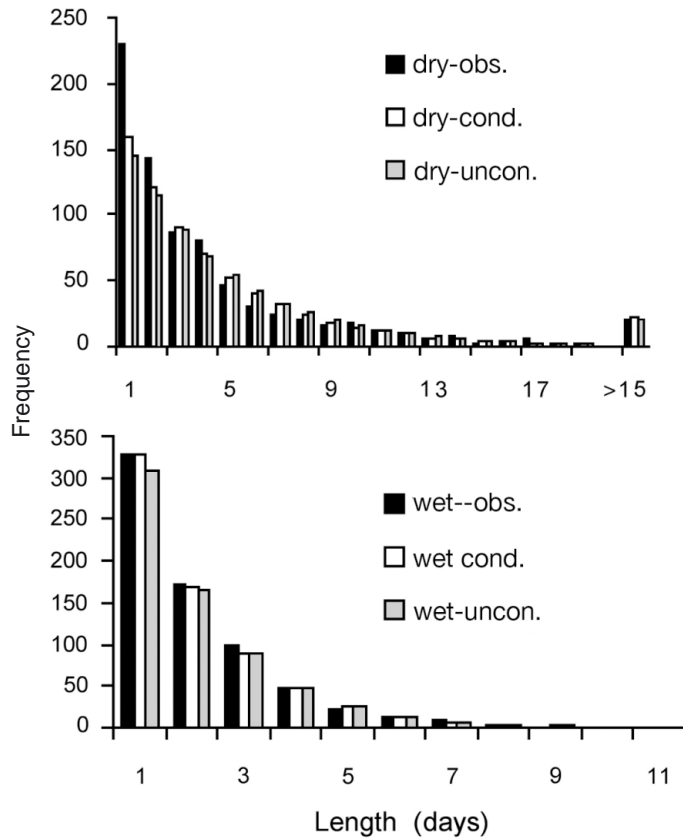|		≤5 mm	(5 mm, 10 mm)	(10 mm, 15 mm)	>10 mm	>15 mm	>20 mm
1901–1949							
Winter	Observations	21.4	4.1	1.4	2.8	1.4	0.5
	Conditional	21.5	5.2	1.9	3.2	1.3	0.6
		(20.4, 22.7)	(4.6, 5.8)	(1.6, 2.2)	(2.8, 3.7)	(1.1, 1.6)	(0.4, 0.8)
	Unconditional	21.5	5.1	2.0	3.4	1.5	0.6
		(20.4, 22.8)	(4.5, 5.7)	(1.6, 2.3)	(3.0, 3.9)	(1.2, 1.7)	(0.5, 0.8)
Spring	Observations	22.1	5.1	1.9	4.0	2.1	1.0
	Conditional	18.6	4.8	2.0	3.9	1.9	0.9
		(17.5, 19.8)	(4.3, 5.4)	(1.7, 2.4)	(3.4, 4.4)	(1.5, 2.2)	(0.7, 1.2)
	Unconditional	22.7	5.9	2.6	5.0	2.4	1.2
		(21.5, 23.9)	(5.3, 6.5)	(2.2, 2.9)	(4.4, 5.5)	(2.1, 2.8)	(0.9, 1.4)
Summer	Observations	17.0	5.1	2.4	6.2	3.8	2.3
	Conditional	17.1	5.5	2.8	6.5	3.7	2.2
		(16.2, 18.0)	(5.0, 6.1)	(2.4, 3.2)	(5.8, 7.1)	(3.2, 4.1)	(1.8, 2.5)
	Unconditional	17.4	5.4	2.8	6.5	3.8	2.2
		(16.4, 18.4)	(4.8, 6.0)	(2.4, 3.2)	(5.9, 7.1)	(3.3, 4.2)	(1.9, 2.6)
Autumn	Observations	17.3	3.9	1.7	4.0	2.3	1.3
	Conditional	14.3	4.3	2.0	4.4	2.3	1.3
		(13.3, 15.2)	(3.9, 4.9)	(1.7, 2.4)	(3.8, 4.9)	(1.9, 2.7)	(1.0, 1.6)
	Unconditional	14.7	4.3	2.1	4.5	2.4	1.4
		(13.7, 15.7)	(3.8, 4.8)	(1.7, 2.4)	(4.0, 5.0)	(2.1, 2.8)	(1.1, 1.7)
1950–1999							
Winter	Observations	21.7	3.9	1.7	3.5	1.8	1.1
	Conditional	21.5	5.0	1.6	2.5	0.9	0.3
		(20.4, 22.6)	(4.5, 5.6)	(1.3, 1.9)	(2.1, 2.9)	(0.7, 1.1)	(0.2, 0.5)
	Unconditional	21.7	4.9	1.7	2.7	1.0	0.4
		(21.5, 23.9)	(4.6, 5.6)	(1.4, 2.1)	(2.4, 3.2)	(0.8, 1.3)	(0.3, 0.5)
Spring	Observations	22.9	5.1	2.1	4.9	2.8	
	Conditional	18.2	4.9	1.9	3.3	1.5	0.6
		(17.2, 19.4)`	(4.3, 5.4)	(1.6, 2.2)	(2.9, 3.8)	(1.2, 1.7)	(0.5, 0.8)
	Unconditional	21.9	5.7	2.3	4.2	1.9	0.9
		(20.9, 23.1)	(5.2, 6.3)	(2.0, 2.7)	(3.7, 4.7)	(1.6, 2.2)	(0.6, 1.1)
Summer	Observations	18.4	4.3	2.0	6.1	4.1	2.3
	Conditional	16.3	5.6	2.9	6.8	3.9	2.3
		(15.3, 17.2)	(5.1, 6.2)	(2.5, 3.3)	(6.2, 7.4)	(3.4, 4.3)	(1.9, 2.6)
	Unconditional	16.5	5.5	2.8	6.8	3.9	2.3
		(15.6, 17.5)	(4.9, 6.0)	(2.4, 3.2)	(6.2, 7.4)	(3.5, 4.4)	(2.0, 2.79
Autumn	Observations	14.8	3.4	1.8	4.3	2.5	1.3
	Conditional	17.2	4.7	2.0	3.8	1.8	0.9
		(16.1, 18.3)	(4.2, 5.2)	(1.7, 2.3)	(3.3, 4.3)	(1.5, 2.1)	(0.7, 1.1)
	Unconditional	17.6	4.7	2.1	4.3	2.2	1.1
		(16.4, 18.7)	(4.1, 5.2)	(1.8, 2.5)	(3.8, 4.9)	(1.8, 2.6)	(0.9, 1.4)

Fig. 8. Frequencies of winter dry (top) and wet (bottom) intervals with various lengths as derived from observations and from generated time series (conditional and unconditional models) at the Bucharest station. The results are obtained for the independent data set over 1950–1999 with the model fitted to 1901–1949 as ensemble means of 1000 runs. The rightmost column for the dry intervals refers to lengths >15 d

cially for wet intervals). For all seasons, the dry intervals of 1–2 d length are less frequent in the simulations than in reality. Generally, the shorter dry intervals are underestimated and the longer dry intervals (greater than 9 d) are overestimated. The frequency of the extreme events is very well reproduced for winter and summer.

The performance of the stochastic models in generating daily precipitation time series was also quantified for frequencies of days with precipitation amount within or exceeding selected thresholds. These results are presented in Table 5. For 1901–1949, in spite of some failures of the CCA model for spring and summer, the mean observed frequencies of days with precipitation exceeding 20 mm are covered by the 90% confidence intervals of both models for all seasons. For other thresholds, the conditional model is generally better, except for winter (5–15 mm), when the values are overestimated by both models, and autumn (≤5 mm), when the values are underestimated. For 1950–1999, the results are not so good, and the observed frequencies are generally underestimated.

deviations are considerably underestimated by the unconditional model for all seasons and both subintervals, while the conditional model is obviously better, especially for summer. However, the conditional model slightly overestimates the interannual variability of the other seasons for the first subinterval and underestimates it for the second one (especially for spring).

Figs. 8 to 11 show the frequencies of dry and wet intervals with various lengths derived from observations and from the unconditional and conditional stochastic models as ensemble means over 1000 runs. The frequencies of the extreme events such as dry intervals longer than 15 d are presented separately. Generally, there are no significant differences between the results achieved with unconditional and conditional models, except for spring (short intervals), when the unconditional model is better. For the other seasons, the unconditional model is slightly better than the conditional one for shorter dry intervals. The best agreement with observations is obtained for winter and autumn (espe-
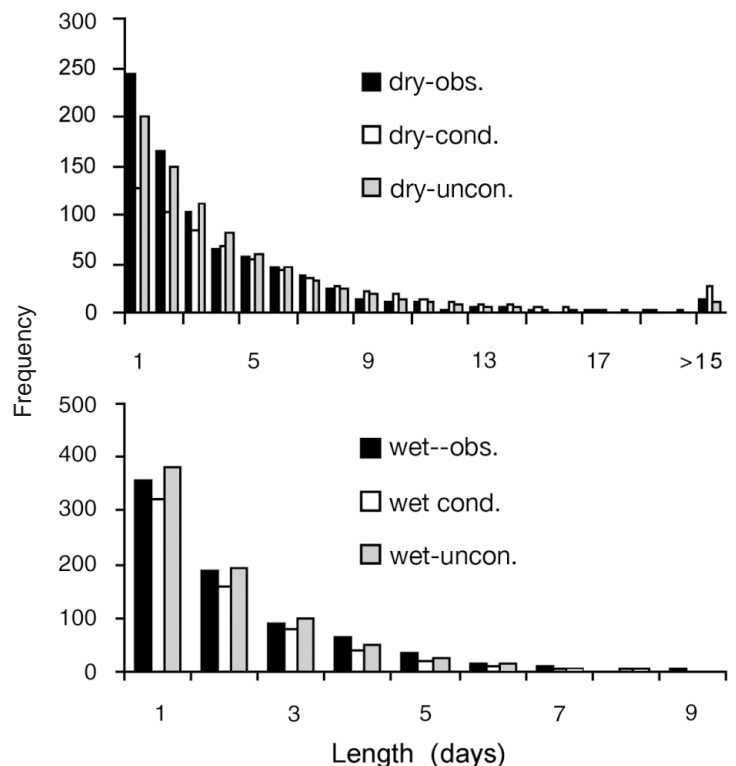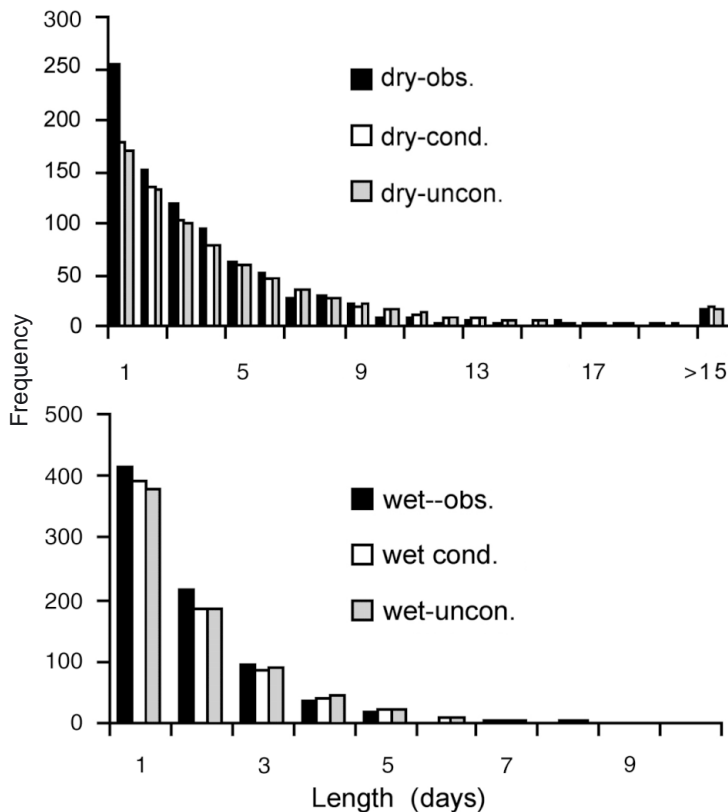

Fig. 9. Frequencies of spring dry (top) and wet (bottom) intervals with various lengths as derived from observations and from generated time series (conditional and unconditional models) at the Bucharest station. The results are obtained for the independent data set over 1950–1999 with the model fitted to 1901–1949 as ensemble means of 1000 runs. The rightmost column for the dry intervals refers to lengths >15 d

4. CONCLUSIONS

The CCA model is most accurate for winter and autumn (transition probabilities), slightly accurate for μ (winter, especially for 1950–1999) and inaccurate for *k*. This result is in agreement with previous studies (Busuioc & von Storch 1996, Busuioc et al. 1999) on the connection between seasonal precipitation totals at 14 Romanian stations (including Bucharest) and the large-scale circulation. The unexpected low accuracy for μ suggests that the strong link of seasonal precipitation (winter and autumn) with the SLP field is given by the strong link between the number wet days and SLP, a fact supported by the direct correlation between them. The results also suggest that *k* does not depend on the large-scale circulation. One reason for the model failure regarding μ and *k* could be that the 2 subintervals have different statistics of precipitation at Bucharest, i.e. 2 different precipitation regimes. Urban influence could be one reason for this behavior, but it is difficult to

Fig. 10. Frequencies of summer dry (top) and wet (bottom) intervals with various lengths as derived from observations and from generated time series (conditional and unconditional models) at the Bucharest station. The results are obtained for the independent data set over 1950–1999 with the model fitted to 1901–1949 as ensemble means of 1000 runs. The rightmost column for the dry intervals refers to lengths >15 d
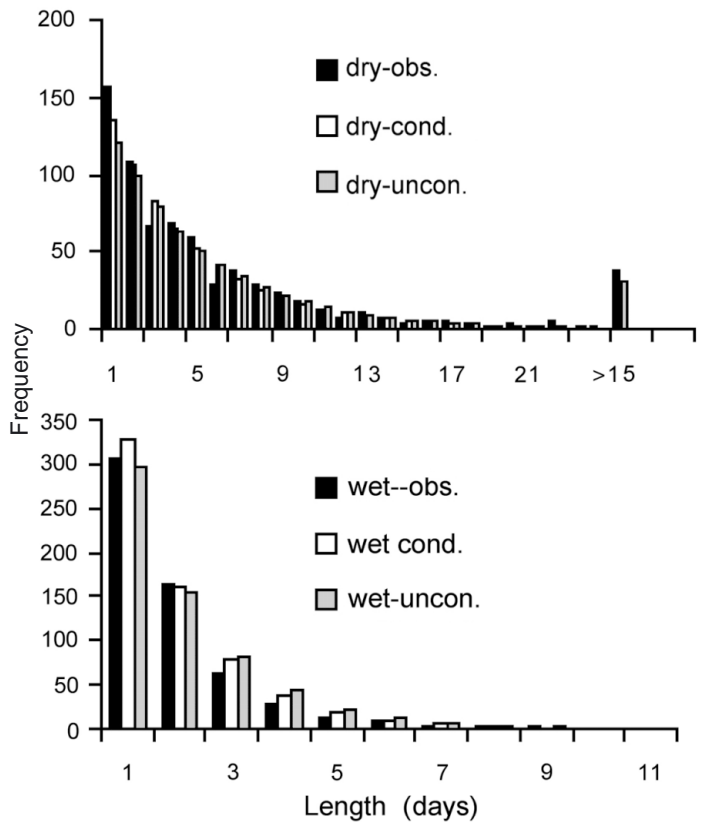
Fig. 11. Frequencies of autumn dry (top) and wet (bottom) intervals with various lengths as derived from observations and from generated time series (conditional and unconditional models) at the Bucharest station. The results are obtained for the independent data set over 1950–1999 with the model fitted to 1901–1949 as ensemble means of 1000 runs. The rightmost column for the dry intervals refers to lengths >15 d

separate it from natural variability, since daily precipitation time series as long as that for Bucharest are not available for other stations in Romania.

The conditional and unconditional stochastic models were similarly accurate in reproducing the statistical properties of precipitation, except for the trend and interannual variability, which inherently only the conditional model can deal with (through SLP forcing). Therefore, for climate change applications, especially in the case of the transient scenarios, the conditional model is recommended. The unconditional models could be used (especially in equilibrium scenarios) but the problem comes in determining how to perturb the parameters in some sensible and consistent way. The advantage of the conditional model is that this can be done more automatically using GCM/RCM large-scale circulation data.

In this study, some statistical features were well reproduced by both stochastic models for all seasons, such as:

- mean and expected maximum duration of wet intervals;
- daily mean of precipitation for rainy days;
- expected maximum duration of dry intervals is well covered by the 90% confidence intervals of both models, but better estimated by the unconditional model;
- frequency of days with precipitation amount greater than 20 mm for the 1901–1999 period.

Other statistical features of the generated precipitation time series are only partially reproduced by both models, or are better reproduced by one of the models, such as:

- The mean duration of dry intervals for winter and summer (1901–1949) and autumn (1950–1999) are reproduced well by both models; in the other cases, both models overestimate the observed values, except for spring, when the unconditional model is better.
- The daily standard deviation is better estimated for 1901–1949, except for summer, when it is underestimated; for 1950–1999 the standard deviation is slightly underestimated for all seasons (less for summer).
- The expected maximum daily precipitation is generally underestimated for all seasons (both models), but the observed values are covered by the 90% confidence intervals, except for summer (1901–1949) and winter and autumn (1950–1999).
- The frequency of wet intervals (winter and autumn) and of extreme dry events (winter and autumn) well reproduced; in the rest of cases, generally, the frequency of shorter dry intervals is underestimated and the longer dry intervals (>9 d) are too frequent.
- The seasonal mean of rainy days is generally well reproduced by both stochastic models, except for spring and autumn, when it is underestimated by the conditional model and overestimated by the unconditional model.
- The linear trend of winter precipitation is well identified by the conditional stochastic model, but the increase in 1901–1949 is underestimated and the decrease in 1950–1999 is overestimated.
- The interannual variability is better captured by the conditional stochastic model in all seasons (especially for summer) and both subintervals, except for spring (1950–1999), when it is underestimated.

In conclusion, the conditional stochastic model presented in this paper can be used to generate daily precipitation time series, especially for winter and autumn. Compared to the unconditional model this model has the advantage of capturing the observed changes in the local seasonal precipitation induced by changes in the large-scale circulation, represented by the SLP field. This makes it useful for the construction of climate change scenarios based on GCM output (especially transient versions). Another important advantage of the conditional stochastic model is that it captures better the interannual variability. This stochastic model could be improved (especially for spring and summer seasons) by adding other large-scale parameters. Moisture variables could be important large-scale predictors (Heimann & Sept 2000), but unfortunately they are not available for long time series. The NCEP reanalysis are only available for 1948 onwards. For climate change scenarios, this problem could be solved by developing conditional stochastic models over shorter time periods using a cross-validation procedure (Barnett & Preisendorfer 1987).

LITERATURE CITED

Amos DE (1983) A portable Fortran subroutine for derivates of the PSI function, ACM Trans Math software, Algorithm 610

Barnett TP, Preisendorfer R (1987) Origin and levels of monthly and seasonal forecast skill for United States surface air temperature determined by canonical correlation analysis. Mon Weather Rev 115:1825–1850

Busuioc A, von Storch H (1995) The connection between summer precipitation anomalies in Romania and large-scale atmospheric circulation. In: Proceedings of atmospheric physics and dynamics in the analysis and prognosis of precipitation fields, Rome, 15–16 November 1994, p 369–373

Busuioc A, Von Storch H (1996) Changes in the winter precipitation in Romania and its relation to the large-scale circulation. Tellus 48A:538–552

Busuioc A, von Storch H, Schnur R (1999) Verification of GCM generated regional seasonal precipitation for current climate and of statistical downscaling estimates under changing climate conditions J Clim 12:258–272

Busuioc A, Chen D, Hellström C (2001) Performance of statistical downscaling models in GCM validation and regional climate change estimates: application for Swedish precipitation. Int J Climatol 21:557-578

Cavazos T (1999) Large-scale circulation anomalies conductive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas. J Clim 12:1506–1523

Coe R, Stern RD (1982) Fitting models to daily rainfall data. J Appl Meteorol 21:1024–1031

Dobi-Wantuch I, Mika J, Szeidl L (2000) Modelling wet and dry spells with mixture distributions. Meteorol Atmos Phys 73:245–256

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

Heimann D, Sept V (2000) Climate change estimates of summer temperature and precipitation in the Alpine region. Theor Appl Climatol 66 1–12

Heyen H, Zorita E, von Storch H (1996) Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. Tellus 48A:312–323

Katz RW (1996) Use of conditional stochastic models to generate climate change scenarios. Clim Change 32:237–255

Katz RW, Parlange MB (1993) Effects of an index of atmospheric circulation on stochastic properties of precipitation. Water Resour Res 29:2335–2344

Lettenmaier D (1995) Stochastic modelling of precipitation with application to climate model downscaling. In: von Storch H, Navarra A (eds) Analysis of climate variability: application of statistical techniques. Springer-Verlag, Berlin, p 197–212

Neacsa O, Popovici C, Tuinea P, Popa G (1974) Contribution to study of Bucharest city climate. In: Climatological studies, Vol I. National Institute of Meteorology and Hydrology, Bucharest, p 5–55 (in Romanian)

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Sneyers R (1975) Sur l'analyse statistique des series d'observation. WMO Note Tech no. 143

Trenberth KE, Paolino DA (1980) The Northern Hemisphere sea-level pressure data set: trends, errors and discontinuities. Mon Weather Rev 108:855–872

von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scale: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Werner P, von Storch H (1993) Interannual variability of Central European mean temperature in January–February and its relation to large-scale circulation. Clim Res 3:195–207

Wilks D (1992) Adapting stochastic weather generation algorithms for climate change studies. Clim Change 22:67–84

Yarnal B, Comrie AC, Frakes B, Brown D (2001) Development and prospects in synoptic climatology. Int J Climatol 21:1923–1950

Zorita E, von Storch H (1999) A survey of statistical downscaling techniques. J Clim 2:2474–2489

Zorita E, Hughes JP, Lettemaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 8:1024–1042