

Research Article

A Hidden Markov Model Applied to the Daily Spring Precipitation over the Danube Basin

Constantin Mares,¹ Ileana Mares,¹ Heike Huebener,² Mihaela Mihailescu,³
Ulrich Cubasch,⁴ and Petre Stanciu¹

¹ National Institute of Hydrology and Water Management, 013686 Bucharest, Romania

² Hessian Centre on Climate Change, 65203 Wiesbaden, Germany

³ University of Agronomic Sciences and Veterinary Medicine, 011464 Bucharest, Romania

⁴ Institute of Meteorology, Free University Berlin, 12165 Berlin, Germany

Correspondence should be addressed to Constantin Mares; constantin_mares_ro@yahoo.com

Received 12 August 2013; Revised 20 November 2013; Accepted 26 December 2013; Published 19 February 2014

Academic Editor: Klaus Dethloff

Copyright © 2014 Constantin Mares et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main goal of this study is to obtain an improvement of the spring precipitation estimation at local scale, taking into account the atmospheric circulation on the Atlantic-European region, by a statistical downscaling procedure. First we have fitted the precipitation amounts from the 19 stations with a HMM with 7 states. The stations are situated in localities crossed by the Danube or situated on the principal tributaries. The number of hidden states has been determined by means of BIC values. A NHMM has been applied then to precipitation occurrence associated with the information about atmospheric circulation over Atlantic-European region. The atmospheric circulation is quantified by the first 10 components of the decomposition in the EOFs or MEOFs. The predictors taking into account CWTs for SLP and the first summary variable from a SVD have also been tested. The atmospheric predictors are derived from SLP, geopotential, temperature, and specific and relative humidity at 850 hPa. As a result of analyzing the multitude of the predictors, a statistical method of selection based on the informational content has been achieved. The test of the NHMM performances has revealed that SLP and geopotential at 850 hPa are the best predictors for precipitation.

1. Introduction

Hidden Markov models (HMM) were first described in [1] and in other series of statistical papers in the second half of the 1960s. The most cited publication regarding the solutions of practical and theoretical problems that appear related to HMM application is that of Rabiner [2]. Although the first applications of HMM in meteorology, hydrology, or climate change studies appeared relatively late after the 1960s, a substantial increasing of the investigations based on fitting by HMM has been noted lately.

Considering that precipitation, as it is known, represents the most important predictor in hydrological studies, we will focus on the modeling of these meteorological variables. If at the beginning of the 1990s the precipitation modeling was achieved using a HMM, then the number of investigations where precipitation modeling using HMM is associated with precipitation modeling using nonhomogenous HMM

(NHMM) by introducing atmospheric factors as inputs has increased. The atmospheric factors (or other predictors) modify the transition matrix of HMM, and this modeling clearly approaches the reality best.

Among the first applications of HMM and NHMM for generating precipitation, there is the one described in [3]. Paper [4] presents the relation between the transition matrix of a NHMM which is dependent of time and the transition matrix of a HMM which is independent on time. The variable that realizes this transformation is the input variable. As it is shown in [5], the HMMs are utilized both as diagnostic and as prognostic tools.

Regarding the predictors (inputs) in NHMM, their selection is a rather difficult problem. As it is presented in [6] the selection of suitable predictors is crucial when developing a statistical downscaling model. The basic demand for a predictor refers to the given information that may be emphasized

by statistical analyses mainly by correlative analysis between predictor and predictand.

The atmospheric fields must be filtered in order to reduce the data volume but, at the same time, we must keep sufficient information about those fields for the association with the precipitation; namely, we must filter or compress data [7, 8]. On compressing the atmospheric variables, the fact that circulation regimes have intrinsic time scales of several days to a week and exert control on local weather needs to be taken into account. Also the study of the atmospheric variables on components as wave-like allows the connection between oscillatory phenomena [9].

Among the filtered techniques in the NHMM applications, the most used are singular value decomposition (SVD) and the principal component analysis. Bellone et al. [10] apply the SVD technique for finding the relation between the HMM states for precipitation amounts in 17 winters in Washington state and the atmospheric variables represented by geopotential at 1000 and 850 hPa and relative humidity at 1000 and 850 hPa. Paper [11] presents the application of a NHMM model with inputs taken from general circulation models (GCMs) and limited area models (LAMs) quantified by an index that characterizes temperature and humidity at 850 hPa.

As Zheng and Katz [12] showed, the modeling of the relation between precipitation and the atmospheric or oceanic indices is very important because depending on this relation we will be able to estimate the impact of the climatic changes on the hydropower generation capacity of the studied region. Establishing a classification of the atmospheric circulation in accordance with the study purpose, for example, the determination of atmospheric conditions (circulation types) that produce precipitation between some limits, is also very important for obtaining the relation between local and synoptic scale.

The first investigations related to the HMM applications to the discharges in the Danube lower basin in association with circulation atmospheric types (CWTs) over Europe have been presented in [13].

A method which is often applied for the precipitation downscaling is the analogy technique. Cubasch et al. [14] achieved a study based on the deterministic analogy in which generated rainfall was considered equal to that observed on the day from the recorded observations which had the greatest similitude with the circulation characteristics on a large scale. An attractive technique of analogy for the study of hydrometeorological extreme events is presented in [15, 16]. The method of resampling of the historical hydrometeorological time series presented by the authors is a nonparametric method which offers the opportunity to simulate different variables (multivariate) for different locations (multisite) simultaneously; therefore, it is possible to simulate much longer time series than historical records from which it is resampled.

There are also other applications of HMM besides the precipitation field. In [17] the authors apply a HMM for studying regimes and metastability in atmospheric flows.

In paper [18], a HMM with 7 states has been applied to spring daily discharges at Orsova. The simulations of

HMM states and of the atmospheric states have led to an improvement of the information on the sequences of discharge states in the Danube lower basin, associated with atmospheric circulation.

A long range forecasting of monthly discharge of the Danube in Bratislava has been obtained in [19], where two types of models have been applied, a hidden Markov model derived from Fourier series and a stochastic seasonal autoregression model of moving averages.

In the present paper we have applied first a HMM to daily precipitation in the spring time, recorded at 19 stations situated along the Danube basin. After the fitting with HMM, the next step is the application of NHMM by introducing some atmospheric predictors. The method used for estimating HMM and NHMM parameters is that described in [20–22].

Regarding the problem of precipitation modeling at multistations, the problem of the dependence between the stations appears. This dependence can influence to a certain extent the results depending on the stations distribution, the local conditions, and the period of the year which is analyzed. Kirshner et al. [23] have shown that an improvement of the results can be obtained when considering the discrete-valued vector time series data, modeling using extensions of Chow-Liu tree models, to capture both dependencies across time and dependencies across variables.

For the atmospheric data compressing we have used both known methods such as decomposition in EOF, singular value decomposition (SVD), circulation types, and some new ones such as decomposition in multivariate EOF (MEOF) of several atmospheric fields and an analysis with MEOF simultaneously of one or more atmospheric fields and precipitation. This decomposition in MEOF with including the precipitation field in the covariance matrix is similar to the SVD technique in which information about the correlations between the atmospheric and precipitation fields is also introduced in order to obtain the predictors. For all the analyzed cases of decomposition in EOF or MEOF the first 10 principal components (the temporal components of the decomposition) have been considered.

The novelty of the present study consists in the selection of the predictors which are introduced in a NHMM and mostly in the application of a method of decomposition in MEOF of the atmospheric variables associated with the precipitation field.

The present study has the following structure: the data and methodology used are described in Section 2, the results are discussed in Section 3, and the conclusions are presented in the last paragraph.

2. Materials and Methods

2.1. Data. Daily precipitation amounts for the spring period from 19 stations (Table 1) that can contribute to the Danube discharge are subject to analysis in association with the atmospheric circulation for the 1958–2001. So, in each of the 44 years (1958–2001), we have a time series of 92 days (the months of March, April, and May). This period has been

TABLE 1: Geographical information about the analysed stations.

No.	Station	CN	LONG	LAT	Height
1	AUGSBURG	GE	10.56	48.26	463
2	NURNBERG	GE	11.03	49.30	319
3	INNSBRUCK	AT	11.24	47.16	577
4	SONNBLICK	AT	12.57	47.03	3106
5	SALZBURG	AT	13.00	47.48	437
6	KREDARICA	SI	13.51	46.22	2514
7	LJUBLJANA	SI	14.31	46.04	299
8	GRAZ	AT	15.27	47.05	366
9	ZAGREB	HR	15.58	45.49	156
10	WIEN	AT	16.21	48.14	198
11	SARAJEVO	BA	18.23	43.51	577
12	OSIJEK	HR	18.38	45.32	88
13	BEOGRAD	YU	20.28	44.48	132
14	ARAD	RO	21.21	46.08	117
15	DROBETA-TURNU SEVERIN	RO	22.38	44.38	77
16	TG JIU	RO	23.16	45.02	203
17	TURNU MAGURELE	RO	24.53	43.45	31
18	BUCHAREST	RO	26.06	44.25	82
19	CALARASI	RO	27.20	44.12	19

selected in order to use the atmospheric fields from ERA-40. The precipitation time series have been downloaded from European climate assessment and dataset (ECA&D) website. Details about homogeneity of the precipitation over Europe are given in [24].

For the same period, over the Atlantic European (AE) region (50°W - 40°E; 30°-65°N), the atmospheric data (ERA-40, Re-analyses) for three fields, daily sea level pressure (SLP), geopotential at 850 hPa, and specific humidity at 850 hPa, has been considered.

ERA-40 is a reanalysis of meteorological observations from September 1957 to August 2002 produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) in collaboration with many institutions [25].

For the region mentioned above, circulation weather types (CWT) are calculated from daily SLP. The atmospheric circulation is quantified by ten circulation weather types (CWTs; two circular types: cyclonic and anticyclonic, eight directional types: NE, E, SE, S, SW, W, NW, and N). More details are given in [13].

For a smaller sector (E) delimited by (0°-40°E; 30°-65°N) besides the abovementioned fields, the following daily fields have been analyzed: 850 hPa temperature and the relative humidity at 850 hPa.

2.2. Methods

2.2.1. Homogenous Hidden Markov Model (HMM). In the following description of HMM we will take into account the notation used in [20]. Let $R_t = (R_t^1, \dots, R_t^M)$ be a multivariate random vector of the precipitation occurring for

a network of M rain stations. Let us consider the observed value $r_t^i = 1$ if there is observed precipitation in the day t at the station i and $r_t^i = 0$ if there was not observed any precipitation. Let us take S_t , the hidden state with rain for the day t , and i is one of the values from 1 to K states. We note $R_{1:T}$ and $S_{1:T}$ are the daily sequences for the precipitation occurrence and the precipitation hidden states. A HMM for precipitation consists of two assumptions of conditional independence. The first assumption is that the observation of the multivariate precipitations R_t at the moment t is independent of all the other variables from the model until the moment t , conditioned by the weather state S_t at moment t , which is

$$P(R_t/S_{1:t}, R_{1:t-1}) = P(R_t/S_t). \quad (1)$$

The second assumption is that the process of the hidden state $S_{1:T}$ is a first-order Markov:

$$P(S_t/S_{1:t-1}) = P(S_t/S_{t-1}) \quad (2)$$

and that this first order Markov chain is homogeneous in time, which in fact signifies that the transition probabilities matrix of $K \times K$ dimension from (2) does not change in time.

2.2.2. Nonhomogeneous Hidden Markov Model (NHMM). Let us consider a D -dimensional column vector X of the predictors for the day t . Let us note through $X_{t:T}$ the sequence X_t, \dots, X_T .

In this case (2) becomes

$$P(S_t/S_{1:t-1}, X_{t:T}) = P(S_t/S_{t-1}, X_t), \quad (3)$$

such that the hidden state from the day t depends on the predictor vector X_t for the t day and also on the hidden state value S_{t-1} from the day $t - 1$.

Therefore, if X_t varies in time, the transition probability matrix will vary also in time and this makes the Markov chain model nonhomogeneous (NHMM).

A simple way to parameterize this probability is

$$P(S_t = j/S_{t-1} = i, X_t = x) \propto \gamma_{ij} f, \quad (4)$$

where f denotes a probability density function and γ_{ij} gives the baseline transition probability from state i to state j .

For a fixed number of hidden states K we can find a parameter Θ so as to have, for a data set, the best adjustment according to the NHMM states sequence. Therefore we search the Θ parameter which maximizes the conditioned probabilities of the observed data as a Θ function. In fact, this is a maximum likelihood probability applied at an observation set from the past (learning algorithm). The Θ parameter which maximizes $L(\Theta)$,

$$\begin{aligned} L(\Theta) &= P(R_{1:T}/X_{1:T}, \Theta) \\ &= \sum_{S_{1:T}} P(S_i/X_i) \prod_{t=2}^T P(S_t/S_{t-1}, X_t) \prod_{t=1}^T P(R_t/S_t), \end{aligned} \quad (5)$$

can be obtained through the well-known Baum-Welch method.

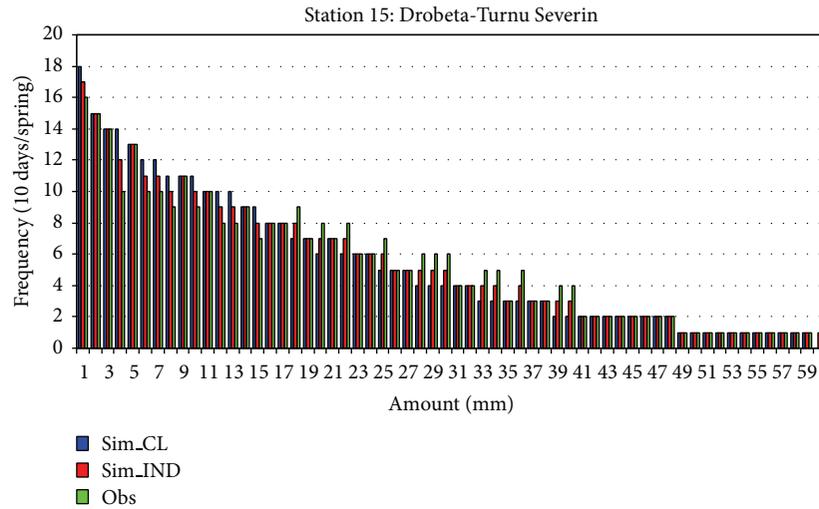


FIGURE 1: The histogram of the observed spring daily precipitation amount (total over 10 days) and simulated precipitation, considering spatial independence (IND) and dependence (CL) for Drobeta-Turnu Severin station.

The process is iterative on a number of cases. In order to establish the best approximation of the maximum likelihood, the Bayesian information criterion (BIC) is applied. Details about the procedure of determination of the NHMM parameters are described in [21, 22].

In the present study, the precipitation amounts are modeled using a *delta-exponential* with three components, consisting of a delta-function at zero and a set of two exponentials. Estimations were performed using the Multivariate Nonhomogeneous Hidden Markov Model Toolbox (MVNHMM) (<http://www.stat.purdue.edu/~skirshne/MVNHMM/>) and details can be found in [22].

In compressing the atmospheric fields analyzed in the present study we have used mostly the decomposition in empirical orthogonal functions (EOF) for every field analyzed separately but also the decomposition in multivariate EOF (MEOF) when considering two or more fields simultaneously. For all EOF decomposition, we have selected the first 10 temporal components (principal components—PCs) as predictors.

In the modeling by NHMM, the first 10 PCs of the decomposition in EOF and also of the decomposition in MEOF (SLP, geopotential and specific humidity at 850 hPa) on the Atlantic-European region were taken as predictors. The verification of the results has been realized by comparing the observed precipitation with those simulated by NHMM. Because some of the results were not entirely satisfactory, we considered the atmospheric fields on a smaller sector, which we named (E). This sector is $0\text{--}40^\circ\text{E}$; $35\text{--}65^\circ\text{N}$.

We have also selected, as predictors for precipitation, the first summary variables obtained by SVD technique and the predictors in the form of CWT. The experiments with SVD considered only the first summary variable from the SLP and geopotential at 850 hPa fields on the European region. The number of summary variables needed to explain a certain portion of the correlation depends on the relative magnitude of the singular [10].

3. Results and Discussion

In the MVNHMM toolbox, two spatial models are available: (a) *Independent*, in which rainfall at each station is assumed independent of the other stations or (b) *Chow-Liu*, in which spatial dependence is modeled using tree models. We have applied the two methods to our data, and we have compared the results obtained by fitting the precipitation from 19 stations with HMM considering the independence between stations with the results considering the dependence between the stations by applying the algorithm Chow-Liu.

The comparison of the 100 simulations, sequences of 44 year and 92 days per year (springtime) with HMM parameters taking into account two situations: independence and dependence between stations, led to the conclusion that dependence does not improve results (Figure 1). As we can observe from Figure 1, the results differ slightly, and taking into account the fact that the computing time in the case of considering the dependence between the stations is substantially greater than in the case of independence, all the investigations using NHMM have been done considering the independence between the stations.

In Figure 1, the histogram is obtained from 396 values, which means 44 years, each year with 90 days (the spring period was truncated at 90 days), resulting in 3960 values which are taken in periods of 10 days. In Figure 1, as for all the simulations results presented in this paper, the ensemble mean on the 100 simulations was considered, in order to make the comparison with the observations.

Although seven states are difficult to interpret, this number of states was selected considering both the BIC values and the previous investigations related to seven states of Danube discharges in the lower basin [18]. The relative great number of states helps us to associate some states with extreme hydrological events in the next investigations. Spatial distribution of the precipitation amounts and the associated probabilities can be seen in Figure 7.

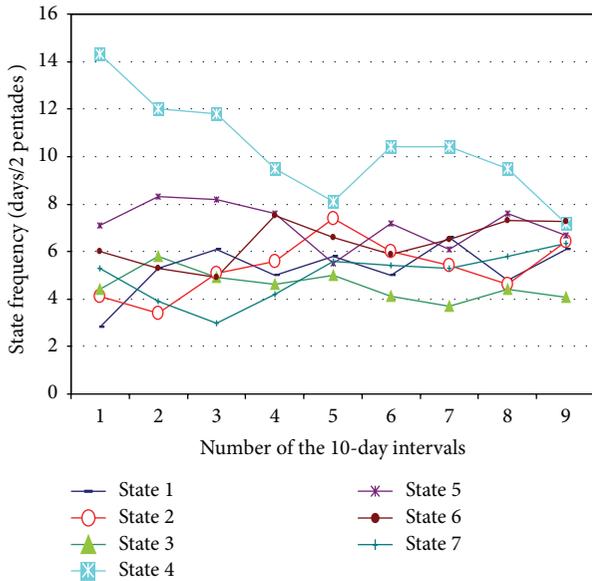


FIGURE 2: The state frequency in the springtime estimated by Viterbi algorithm.

The temporal variability of the precipitation field, estimated with the Viterbi algorithm, reveals the seasonal variability of the HMM states. Figure 2 shows the mean absolute frequency of each state on 10 days from the analyzed period of 44 years. The figures on *x*-axis indicate the number of the 10-day interval from the spring period.

The frequency of state 4 which represents, in HMM, the state with the weakest precipitation decreases from March to May with the convective precipitation increasing. State 5 which represents the greatest amounts in almost all the Danube basin has a variability less pronounced than that for state 4.

In the next stage of fitting with NHMM, the selection of the predictors was a difficult problem as many investigators reported. The difficulty consisted both in the selection of the atmospheric fields and in the modality of compressing the data.

In the present study, as shown before, we have considered as predictors the following atmospheric fields: sea level pressure (SLP) and four variables for the 850 hPa isobaric level, namely, geopotential, temperature, relative humidity, and specific humidity. Variables of the atmospheric fields (SLP, geopotential, and specific humidity at 850 hPa) were considered on two regions (on AE and a more restricted one E).

The NHMM performances have been tested by the correlation coefficients (Table 2) between observed precipitation and those obtained from the ensemble mean of the 100 simulations for each station.

In Table 2, the correlation coefficients between observed and simulated precipitation for the seasonal means have been calculated for the 44 years. From this table one can observe the improvement of the results for almost all the stations if a smaller sector is considered. The model is fitting very well

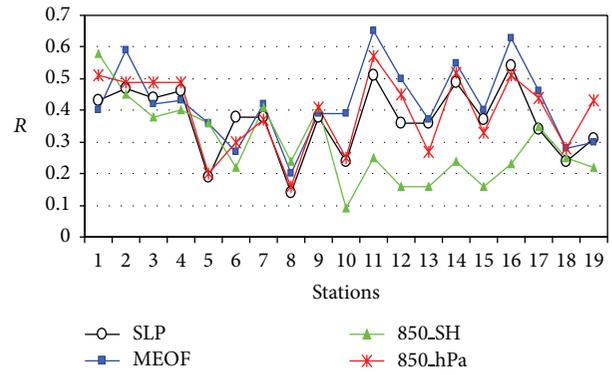


FIGURE 3: Correlation coefficients between observations and simulations (ensemble mean of 100 realizations) for four predictors, represented by 10-PCs of SLP, geopotential at 850 hPa, specific humidity at 850 hPa, and 10-PCs of MEOF decomposition.

the precipitation amounts (excepting some of them) provided that it uses the predictors resulted from the geopotential at 850 hPa and sea level pressure.

Figure 3, for the European region, presents the distribution on stations of the correlations for the four predictors, namely, SLP, geopotential, and specific humidity at 850 hPa and simultaneously three fields: SLP, geopotential, and specific humidity at 850 hPa.

Some characteristics of NHMM are given in Table 3. The first column represents the considered fields out of which the predictors were extracted (the first 10 PCs of the decomposition in EOF or MEOF). The second column contains the variance explained by the first 10 components of the decomposition in EOF or MEOF, the third column shows the log likelihood, and the last column shows the BIC values.

Taking into account the BIC's values, the most performing model should be the one which considers as predictors the first 10 PCs of the decomposition in MEOF, which starts from a covariance matrix in which both the geopotential field at 850 hPa and the precipitation field are simultaneously introduced. The BIC values indicate also good fitting of NHMM to precipitation with predictors from SLP field. The weakest results for the E sector have been obtained considering specific and relative humidity at 850 hPa. This can be explained by the fact that only 10 components which reduce about 55% of the total variance have been considered. We could make the experiment with more components but, as it is known, the increasing of the predictor number can reduce the model performance.

Figure 4 presents the results of the NHMM model performance for the case in which, for the predictor determination, the covariance between the atmospheric fields and those for precipitation at the 19 stations is introduced. For time series with a length of 44, the coefficients greater than 0.38 have a significance level of 1% and those greater than 0.29 have a significant level of 5%. From Figure 4, we note that the most significant results have been obtained by applying a NHMM having as predictors the first 10 components of the decomposition in EOF of the SLP and geopotential at the 850 hPa, fields associated with precipitation values at the

TABLE 2: Correlation coefficients between observed and simulated precipitation by NHMM taking into account atmospheric predictors (first 10 PCs) over the Atlantic European (AE) and the European (E) regions. MEOF here represent a combination between SLP, geopotential at 850 hPa, and specific humidity at 850 hPa.

Region	SLP		G-850 hPa		SH-850 hPa		MEOF (SLP, G, and SH-850 hPa)	
	AE	E	AE	E	AE	E	AE	E
1	0.43	0.54	0.51	0.48	0.58	0.39	0.40	0.52
2	0.47	0.64	0.49	0.60	0.45	0.33	0.59	0.60
3	0.44	0.37	0.49	0.46	0.38	0.13	0.42	0.45
4	0.46	0.32	0.49	0.39	0.40	0.28	0.43	0.37
5	0.19	0.16	0.20	0.25	0.36	0.25	0.36	0.35
6	0.38	0.37	0.30	0.45	0.22	0.26	0.27	0.31
7	0.38	0.50	0.37	0.49	0.41	0.30	0.42	0.40
8	0.14	0.18	0.16	0.22	0.24	0.23	0.20	0.13
9	0.38	0.47	0.41	0.48	0.41	0.26	0.39	0.41
10	0.24	0.18	0.25	0.16	0.09	0.06	0.39	0.16
11	0.51	0.49	0.57	0.57	0.25	0.49	0.65	0.55
12	0.36	0.44	0.45	0.53	0.16	0.32	0.50	0.46
13	0.36	0.32	0.27	0.46	0.16	0.14	0.37	0.29
14	0.49	0.45	0.52	0.50	0.24	0.19	0.55	0.50
15	0.37	0.35	0.33	0.43	0.16	0.23	0.40	0.34
16	0.54	0.50	0.51	0.51	0.23	0.20	0.63	0.51
17	0.34	0.51	0.44	0.61	0.35	0.36	0.46	0.58
18	0.24	0.40	0.28	0.50	0.25	0.28	0.28	0.39
19	0.31	0.49	0.43	0.58	0.22	0.33	0.30	0.48
Average	0.37	0.40	0.39	0.46	0.29	0.26	0.42	0.41

19 stations considered. Taking into account geopotential at 850 hPa as predictor, one can observe a good performance of the model for the stations Nurnberg, Ljubljana, Sarajevo, Osijek, Beograd, Drobeta-Turnu Severin, Tg. Jiu, Turnu Magurele, Bucharest, and Calarasi and a relatively bad one for the stations Salzburg, Graz, and Wien.

Table 4 presents the results obtained considering as predictors the first summary variables obtained by SVD technique and the results of introducing the predictors in the form of circulation types. The experiments with SVD considered only the first summary variable from the SLP and geopotential at 850 hPa fields on the European sector. The first summary variable represents the highest correlation coefficients between the two fields.

The first summary variable in our experiment explains a proportion of correlation with the predictand of 0.61 for SLP and 0.65 for the geopotential. Only the first variable was considered because the others produced a proportion of correlation smaller than 0.21. In this case, by comparing the BIC values, we can also observe that the geopotential at 850 hPa is a better predictor than SLP. But SVD performances in the presented experiments are inferior to those which consider PCs as predictors in NHMM. The weak results obtained in the present study by analysis with SVD comparatively with those obtained in other studies [10, 26] can derive from the fact that both precipitation and atmospheric variables are considered on a too extended spatial network. This aspect is evident also in the results of the present study; namely, the precipitation

simulation is better for the predictors considered on a limited area (E).

Related to the NHMM verifications by considering CWT on the AE sector, the best result was obtained considering ten CWTs and seven states of the precipitation in comparison with the introduction of three types of circulation (cyclonic, anticyclonic, and other) and 3 or 4 states of the precipitation. Regarding CWTs from Table 4, it must be mentioned that these have been calculated [13] to maximize the relation between SLP and precipitation around 25°E, 45°N (near the Bucharest station). For this reason, the results obtained must be considered only for the stations in the vicinity of the Bucharest station. Indeed, the correlation coefficient values between the observed data and those simulated with NHMM introducing as predictors CWTs have a high level of statistical significance for the stations situated in the lower Danube basin and the significance decreases as we go away from the central point (Bucharest) for which the circulation types have been calculated. CWT analysis for 25°E, 45°N shows significant positive (negative) correlation of rainfall with cyclonic (anticyclonic) days. During summer, rainfall originates for NE and N CWTs (low pressure over Black Sea), while during winter E and SE CWTs (low pressure over eastern Mediterranean) contribute to the most rainfall days. Westerly (SW, W, and NW) CWTs are negatively correlated with rainfall due to diffluence effects south of the Carpathians. The CWTs application for the precipitation in other region (e.g., Morocco) is given in [27].

TABLE 3: Fitting by NHMM with seven states of precipitation at the 19 stations, taking into account as predictors (first 10 PCs) for EOF or MEOF decomposition; $\log L$ is log likelihood and BIC represents Bayesian Information Criterion.

Predictors	Variance (%)	$\log L$	BIC
SLP (AE)	83	-125405	256125
SLP (E)	92	-125790	250563
G-850 hPa (AE)	86	-125799	256913
G-850 hPa (E)	94	-122435	250185
SH-850 hPa (AE)	38	-122999	251314
SH-850 hPa (E)	55	-122873	251062
RH-850 hPa (E)	55	-123494	251855
TT-850 hPa (E)	85	-125337	255990
MEOF (slp-G-850-SH-850 hPa)-(AE)	61	-126035	257385
MEOF (slp-G-850-SH-850)-(E)	71	-122651	250619
MEOF (slp-G-850-TT-850 hPa)-(E)	82	-122605	250527
SLP-pp (E)	87	-121812	248939
G-850 hPa-pp (E)	89	-121494	248304
TT-850 hPa-pp (E)	81	-122111	249539
MEOF (slp-G 850 hPa-pp)-(E)	89	-122208	249731

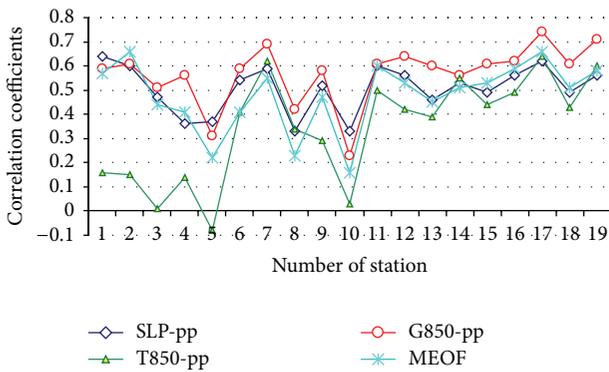


FIGURE 4: Correlation coefficients between observed and simulated mean spring precipitation by NHMM taking into account atmospheric predictors (first 10 PCs) for MEOF decomposition of two pairs of fields: atmospheric over the European region (E) and precipitation at the 19 stations.

Figures 5(a), 5(b), and 5(c) present the time evolution (44 years) of observed and simulated precipitation at the stations Ljubljana ($R = 0.69$), Wien ($R = 0.23$), and Turnu Magurele ($R = 0.74$). The simulated data represent a mean of the 100 simulations of 44 years, each with a sequence of 92 days (1 March–31 May).

The quantiles-quantiles plots between observed and simulated precipitation at the station Turnu Magurele are given in Figure 6, indicating also a very good concordance between the two time series.

It must be mentioned that, even in the case in which NHMM model performances are the highest, there are 1 or 2 stations where simulation with NHMM is not satisfactory. For the station with the highest performance of NHMM (Figure 5(c)), the correlation coefficient between observed and simulated precipitation is 0.74 and for the lowest performance the correlation is 0.23 (Figure 5(b)), a value with

TABLE 4: Fitting by NHMM with seven states of precipitation at the 19 stations, taking into account as predictors the first summary variable of the SVD decomposition of SLP and geopotential at 850 hPa and NHMM with different states, with 10 and 3 predictors of CWTs.

Predictors	Proportion of correlation	Np	$\log L$	BIC
SVD-SLP	0.61	586	-123346	251559
SVD-850 hPa	0.65	586	-123267	251402
CWT 10/7 S	—	586	-123431	251729
CWT 3/3 S	—	238	-127699	257375
CWT 3/4 S	—	392	-125925	254525

a lower statistical significance level. The quantiles-quantiles plots between observed and simulated precipitation indicate good concordance between the two time series for all stations, but with a very good fitting for the stations with high correlation (Figure 6).

Although the best results related to the model performance have been obtained with predictors from the geopotential at 850 hPa and SLP, in the following, we will present only the details obtained by fitting with NHMM considering as predictor SLP given the aspects that this field is easier to interpret and the model performance is relatively appropriate in this case. The distributions on the stations along the Danube basin of precipitation amounts estimated with NHMM having as input the first 10 PCs of the decomposition in MEOF of SLP associated with precipitation are shown in Figure 7(a). The corresponding probabilities are given in Figure 7(b). Because the 19 stations are distributed on a relatively extended area (Table 1), there is a great inhomogeneity that determines a certain difficulty in the results interpretation. Yet, we can associate state 1 with relative high precipitation almost in the entire Danube basin and state 5 with weak precipitation. The other states indicate combinations between great and small amounts disposed in

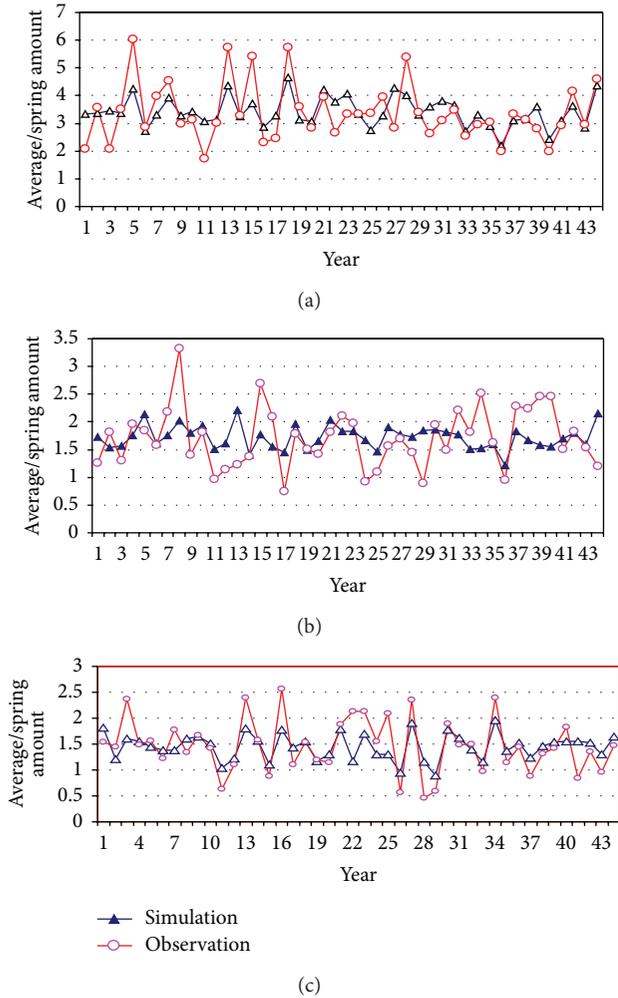


FIGURE 5: Observed and simulated mean seasonal precipitation (1958–2001) with predictors, the first 10 PCs—MEOF (G-850 hPa and pp) for three stations: (a) Ljubljana ($R = 0.69$), (b) Wien ($R = 0.23$), and (c) Turnu Magurele ($R = 0.74$).

different zones of the Danube basin. States 3 and 6 indicate great amounts in the higher and middle basin with great probabilities of occurrence and small amounts in the lower basin, while state 2 shows relative high precipitation in the middle and lower basin with high occurrence probabilities. After the NHMM with seven states including atmospheric circulation information has been fitted to the precipitation amounts, the examination of the states in terms of atmospheric composites has been achieved by means of the results given by Viterbi algorithm.

The atmospheric situations (composite maps) for each state are presented in Figure 8. The number of cases from which the *composite* maps have been calculated is presented in Table 5 and it takes the values from 406 for state 2 to 638 for state 7.

The atmospheric circulation types at the sea level associated with the 7 states of the precipitation explain rather well the synoptic conditions that determine the precipitation occurrence in the seven states considered.

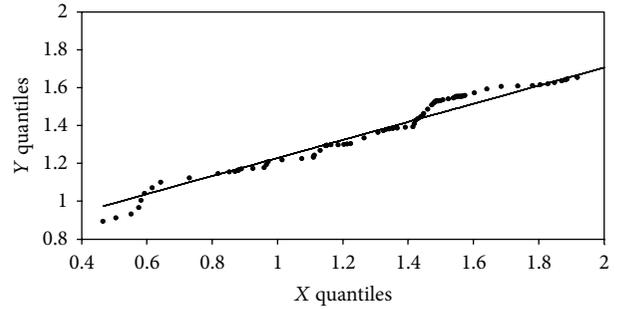


FIGURE 6: Quantiles-Quantiles (Y-Simulation, X-observations) with seasonal mean amounts for Turnu Magurele station.

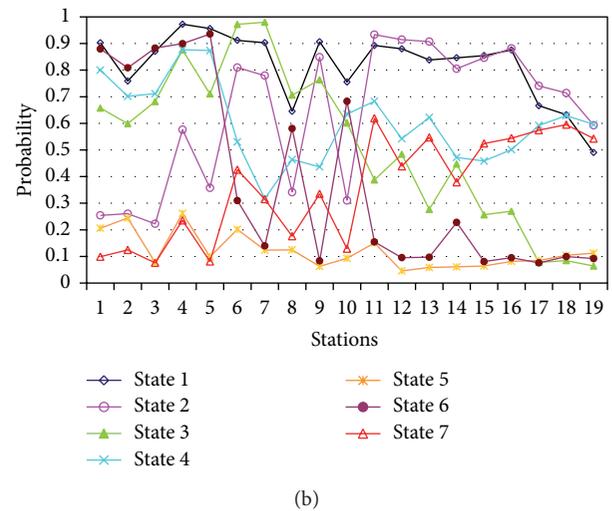
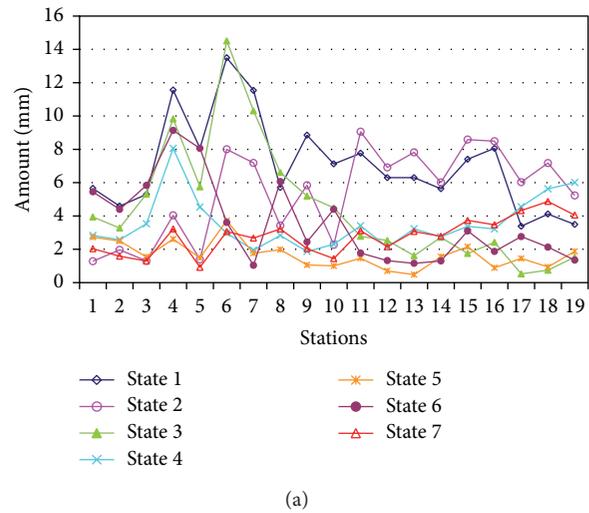


FIGURE 7: Distributions on stations of the NHMM seven states with the first 10 PCs of SLP as predictors: (a) rainfall amounts (mm); (b) corresponding probabilities of rain. Position of the stations along the Danube basin is presented in Table 1.

The association between the seven atmospheric states (composite maps) and the precipitation is evident particularly in the case of state 5 (the state with the weakest precipitation) for an anticyclonic field centered on Romania and extended

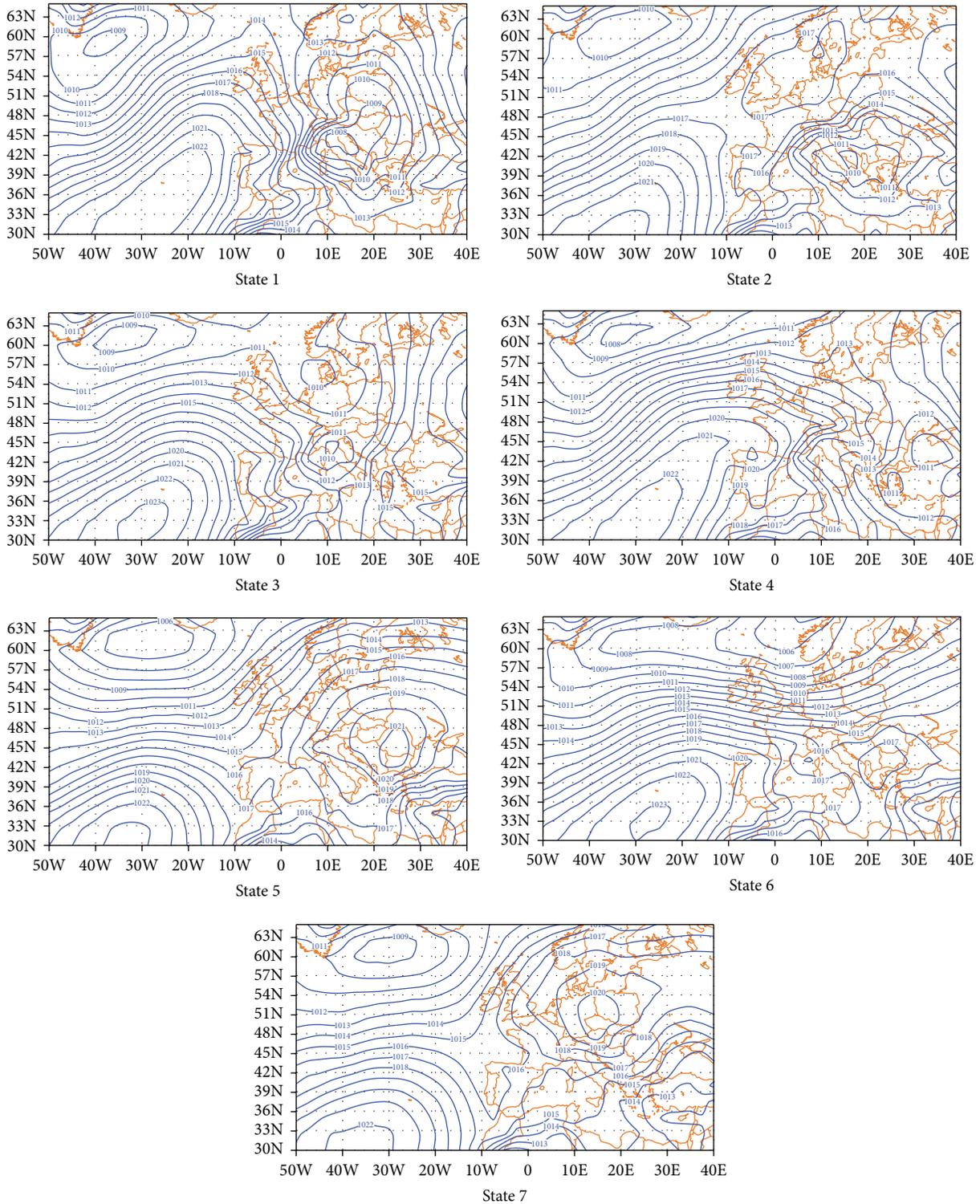


FIGURE 8: Composite maps of sea level pressure (SLP) corresponding to NHMM with seven states of precipitation.

on the entire Danube basin. Also, for the opposite states 1 and 2 a cyclonic field extended on the center and southeast of Europe is associated.

The most likely state sequence estimated by Viterbi algorithm helps us also to reevaluate the transition matrix, which has the form given in relation (6).

After this reevaluation of the transition matrix taking into account the atmospheric factor, namely, the sea level pressure and referring to the states persistence (the transition of a state of precipitation to the same state), the following aspects have been identified: for NHMM the greatest persistence is observed at state 5 and the weakest at state 2, while

TABLE 5: Number of days for each state, from which the composites maps are achieved.

	States						
	1	2	3	4	5	6	7
Days	428	406	581	482	912	601	638

the transition matrix for the homogeneous model (HMM), not shown there, presents the greatest persistence for state 4 and the weakest for state 3.

Transition matrix of NHMM:

$$M = \begin{pmatrix} 0.333 & 0.291 & 0.076 & 0.227 & 0.014 & 0.028 & 0.031 \\ 0.045 & 0.239 & 0.082 & 0.142 & 0.075 & 0.032 & 0.384 \\ 0.223 & 0.145 & 0.403 & 0.049 & 0.052 & 0.037 & 0.091 \\ 0.092 & 0.079 & 0.069 & 0.319 & 0.084 & 0.186 & 0.171 \\ 0.013 & 0.015 & 0.041 & 0.018 & 0.610 & 0.243 & 0.060 \\ 0.121 & 0.029 & 0.312 & 0.116 & 0.101 & 0.260 & 0.061 \\ 0.016 & 0.049 & 0.040 & 0.094 & 0.283 & 0.137 & 0.382 \end{pmatrix}. \quad (6)$$

4. Conclusions

We have modeled daily rainfall amount at 19 stations situated in the Danube basin during spring (1958–2001), by a hidden Markov model (HMM). Seven states have been identified by means of BIC values. Due to the uneven distribution of the stations, the seven states have very different spatial structures regarding the precipitation amounts. Of the seven states, two of them show a relative uniformity of precipitation amounts for the whole basin, highlighting the conditions for a weather regime with excessive or deficit of precipitation.

The influence of atmospheric circulation over Europe on precipitation over the Danube basin has been investigated using a NHMM. The predictors have been selected from the two isobaric levels: the sea level (pressure) and 850 hPa (geopotential, temperature, specific humidity, and relative humidity). Decompositions in EOF and MEOF have been obtained (combining two or more variables) and the first ten principal components (PCs) have been used as covariates in modeling precipitation.

Comparing the results of simulated and observed precipitation, the most significant results have been obtained in the case of considering predictors from SLP field and the 850 hPa geopotential field.

The composite maps associated with each state represent the synoptic situations that determine the rainfall occurrence over the Danube basin. Particularly in case of the state that defines the distribution of very small amounts of precipitation, there is an anticyclonic field centered on Romania and extended on the entire Danube basin. For the opposite states, with the highest amounts of precipitation, a cyclonic field extended on the center and southeast of Europe is associated.

The structure of the spatial distribution of precipitation on a restricted area (19 stations) as in Figure 7 could be emphasized first by selecting the predictors, and considering the set of significant predictors by MEOF and then by downscaling large scale circulation by the NHMM procedure.

This method of investigation can serve as an additional tool for the indirect evaluation of excess or deficit of precipitation in a given zone from the sequence of circulation types obtained from GCMs/RCMs.

These results represent an important starting point for the next investigations. It is necessary to analyze precipitation over a denser network. Also, the results of these investigations will be combined with analyses by extreme value theory of the Danube discharges obtained previously by the authors.

The results of the investigations on the observational data will be of real help in NHMM applications to outputs of GCMs/RCMs. The NHMMs are also useful tools for statistical downscaling for estimating the hydrological impact.

The investigations might be extended taking into account dependence with complete connections, which means a generalization of NHMM [28].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to the two anonymous reviewers for their very useful suggestions and comments. A part of this study has been achieved under the NATO Project (ESP EAP.CLG 982292) and the other part was achieved within the IS-ENES2 project, funded by the European Commission's 7th Framework Programme through Contract no. 312979.

References

- [1] L. E. Baum and T. Petrie, "Statistical inference for probabilities functions of finite state Markov chain," *Annals of Mathematical Statistics*, vol. 37, pp. 1553–1563, 1966.
- [2] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [3] J. P. Hughes and P. Guttorp, "Incorporating spatial dependence and atmospheric data in a model of precipitation," *Journal of Applied Meteorology*, vol. 33, pp. 1503–1515, 1994.
- [4] J. P. Hughes, P. Guttorp, and S. P. Charles, "A non-homogeneous hidden Markov model for precipitation occurrence," *Journal of the Royal Statistical Society C*, vol. 48, no. 1, pp. 15–30, 1999.
- [5] A. M. Greene, A. W. Robertson, and S. Kirshner, "Analysis of Indian monsoon daily rainfall on subseasonal to multidecadal time-scales using a hidden Markov model," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 633, pp. 875–887, 2008.
- [6] D. Maraun, F. Wetterhall, A. M. Ireson et al., "Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user," *Reviews of Geophysics*, vol. 48, no. 3, 2010.
- [7] A. H. Monahan, J. C. Fyfe, M. H. P. Ambaum, D. B. Stephenson, and G. R. North, "Empirical orthogonal functions: the medium is the message," *Journal of Climate*, vol. 22, no. 24, pp. 6501–6514, 2009.

- [8] S. Kravtsov and S. K. Gulev, "Kinematics of Eddy—mean flow interaction in an idealized atmospheric model," *Journal of the Atmospheric Sciences*, vol. 70, no. 8, pp. 2574–2595, 2013.
- [9] M. Ghil and A. W. Robertson, "Waves' vs. 'particles' in the atmosphere's phase space: a pathway to long-range forecasting?" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 1, pp. 2493–2500, 2002.
- [10] E. Bellone, J. P. Hughes, and P. Guttorp, "A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts," *Climate Research*, vol. 15, no. 1, pp. 1–12, 2000.
- [11] S. P. Charles, B. C. Bates, and J. P. Hughes, "A spatiotemporal model for downscaling precipitation occurrence and amounts," *Journal of Geophysical Research D*, vol. 104, no. 24, pp. 31657–31669, 1999.
- [12] X. Zheng and R. W. Katz, "Mixture model of generalized chain-dependent processes and its application to simulation of interannual variability of daily rainfall," *Journal of Hydrology*, vol. 349, no. 1-2, pp. 191–199, 2008.
- [13] H. Huebener, I. Mares, C. Mares, U. Cubasch, and P. Stanciu, "Estimating Romanian rainfall contribution to lower Danube discharge," *Geophysical Research Abstracts*, vol. 9, Article ID 08910, 2007.
- [14] U. Cubasch, H. Von Storch, J. Waszkewitz, and E. Zorita, "Estimates of climate change in Southern Europe derived from dynamical climate model output," *Climate Research*, vol. 7, no. 2, pp. 129–149, 1996.
- [15] J. J. Beersma, *Extreme hydro—meteorological events and their probabilities [M.S. thesis]*, Wageningen Universiteit, Wageningen, The Netherlands, 2007.
- [16] J. J. Beersma and T. A. Buishand, "Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation," *Climate Research*, vol. 25, no. 2, pp. 121–133, 2003.
- [17] C. Franzke, D. Crommelin, A. Fischer, and A. J. Majda, "A hidden Markov model perspective on regimes and metastability in atmospheric flows," *Journal of Climate*, vol. 21, no. 8, pp. 1740–1757, 2008.
- [18] I. Mares, C. Mares, and P. Stanciu, "A hidden Markov model for the Orsova discharge level in springtime," in *Proceedings of the International Conference on Water Observation and Information System for Decision Support*, Ohrid, Republic of Macedonia, 2008, http://balwois.com/balwois/administration/full_paper/ffp-940.pdf.
- [19] P. Pekárová, P. Miklánek, and J. Pekár, "Long-term Danube monthly discharge prognosis for the Bratislava station using stochastic models," *Meteorologický Časopis*, vol. 10, pp. 211–218, 2007.
- [20] A. W. Robertson, S. Kirshner, and P. Smyth, "Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model," *Journal of Climate*, vol. 17, no. 22, pp. 4407–4424, 2004.
- [21] A. W. Robertson, S. Kirshner, P. Smyth, S. P. Charles, and B. C. Bates, "Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland," *Quarterly Journal of the Royal Meteorological Society*, vol. 132, no. 615, pp. 519–542, 2006.
- [22] S. Kirshner, *Modeling of multivariate time series using hidden Markov models [Ph.D. thesis]*, University of California, Irvine, Calif, USA, 2005.
- [23] S. Kirshner, A. W. Robertson, and P. J. Smyth, "Conditional Chow-Liu tree structures for modeling discrete-valued vector time series," Tech. Rep. ICS-TR 04-04, Information and Computer Science, University of California, Irvine, Calif, USA, 2004.
- [24] A. M. G. Klein Tank, J. B. Wijngaard, G. P. Können et al., "Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment," *International Journal of Climatology*, vol. 22, no. 12, pp. 1441–1453, 2002.
- [25] S. M. Uppala et al., "The ERA-40 Re-Analysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, pp. 2961–3012, 2005.
- [26] H. von Storch and F. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, UK, 1999.
- [27] H. Huebener and M. Kerschgens, "Downscaling of current and future rainfall climatologies for southern Morocco. Part I. Downscaling method and current climatology," *International Journal of Climatology*, vol. 27, no. 13, pp. 1763–1774, 2007.
- [28] M. Iosifescu and S. Grigorescu, *Dependence with Complete Connections and Its Applications*, Cambridge University Press, Cambridge, UK, 1990.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

