Metastable Decomposition of High-Dimensional Meteorological Data with Gaps

Illia Horenko

Institute of Mathematics, Free University of Berlin, Berlin, Germany

STAMEN I. DOLAPTCHIEV

Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

ALEXEY V. ELISEEV AND IGOR I. MOKHOV

A.M. Obukhov Institute of Atmospheric Physics (IAP RAS), Moscow, Russia

RUPERT KLEIN

Institute of Mathematics, Free University of Berlin, Berlin, Germany

(Manuscript received 1 February 2008, in final form 26 March 2008)

ABSTRACT

This paper presents an extension of the recently developed method for simultaneous dimension reduction and metastability analysis of high-dimensional time series. The modified approach is based on a combination of ensembles of hidden Markov models (HMMs) with state-specific principal component analysis (PCA) in extended space (guaranteeing that the overall dynamics will be Markovian). The main advantage of the modified method is its ability to deal with the gaps in the high-dimensional observation data. The proposed method allows for (i) the separation of the data according to the metastable states, (ii) a hierarchical decomposition of these sets into metastable substates, and (iii) calculation of the state-specific extended empirical orthogonal functions simultaneously with identification of the underlying Markovian dynamics switching between those metastable substates. The authors discuss the introduced model assumptions, explain how the quality of the resulting reduced representation can be assessed, and show what kind of additional insight into the underlying dynamics such a reduced Markovian representation can give (e.g., in the form of transition probabilities, statistical weights, mean first exit times, and mean first passage times). The performance of the new method analyzing 500-hPa geopotential height fields [daily mean values from the 40-yr ECMWF Re-Analysis (ERA-40) dataset for a period of 44 winters] is demonstrated and the results are compared with information gained from a numerically expensive but assumption-free method (Wavelets-PCA), and the identified metastable states are interpreted w.r.t. the blocking events in the atmosphere.

1. Introduction

Many meteorological and climatological applications are characterized by the need to find some lowdimensional mathematical models for complex systems that undergo transitions between different phases. Such phases can be different circulation regimes in meteorology (Tsonis and Elsner 1990; Kimoto and Ghil

DOI: 10.1175/2008JAS2754.1

1993a,b; Cheng and Wallace 1993; Efimov et al. 1995; Mokhov and Semenov 1997; Mokhov et al. 1998; Corti et al. 1999; Palmer 1999) or glacial-interglacial sequences in climatology (Benzi et al. 1982; Nicolis 1982; Paillard 1998). Starting from the seminal paper by Charney and DeVore (1979), atmospheric blocking formation is also often associated with flip-flops between two states of atmospheric flow, one with strong (unblocked) and other with blocked zonal flow. Regimes of this kind can sometimes be not directly observable (i.e., "hidden") in many dimensions of the system's degrees of freedom and can exhibit persistent or metastable behavior (Majda et al. 2006; Franzke et al. 2008).

Corresponding author address: Illia Horenko, Institute of Mathematics, Free University of Berlin, Arnimallee 2–6, 14195 Berlin, Germany.

E-mail: horenko@math.fu-berlin.de

If knowledge about the system is present only in the form of observation or measurement data, the challenging problem of identification of those metastable states, together with construction of reduced low-dimensional models, becomes a problem of time series analysis and pattern recognition in many dimensions. The choice of the appropriate data analysis strategies (implying a set of method-specific assumptions on the analyzed data) plays a crucial role in correct interpretation of the available time series.

In their recent pioneering works, A. Majda and coworkers have demonstrated the presence of hidden persistent patterns in data generated by different atmospheric models on various scales and shown their connection to the blocking events in the atmosphere (Majda et al. 2006; Franzke et al. 2008). The strategy they used to identify those hidden patterns-a hidden Markov model (HMM) with Gaussian output, hereafter HMM-Gauss-implies the following assumptions about the underlying data: (i) The hidden process switching between the metastable states is Markovian (i.e., has no long term memory-effects) and (ii) the observed process in each of the metastable states is Gaussian and there is no causal dependence between the consecutive observations (i.e., the data points are assumed to be statistically independent of each other). Of particular interest in the present context is the numerical scaling of the expectation-maximization framework on which the HMM-Gauss strategy is based: (i) it scales as $O(n^3)$ w.r.t. the dimension n of the corresponding phase space of observation data (this reduces the applicability of the method to low-dimensional cases) and (ii) it scales as $O(K^2)$ w.r.t. the number K of the hidden states; (iii) the results are not unique because the expectation-maximization (EM) strategy finds only the local optima of the corresponding likelihood function (Baum 1972). On the other hand, the HMM-Gauss method scales linearly w.r.t. the length of the time series, thus making it possible to analyze very long time series.

The first attempts to develop more widely applicable generalizations of the HMM–Gauss approach resulted in construction of the following methods: (i) Wavelets– PCA (Horenko and Schuette 2008, manuscript submitted to *Econ. J.*, hereafter HoSc), (ii) HMM–PCA (hidden Markov models with principal component analysis (PCA; Horenko et al. 2006; HoSc) and (c) HMM– PCA–SDE [hidden Markov models with principal component analysis and stochastic differential equations (SDAs; Horenko et al. 2008)].

Wavelets–PCA is an "assumption free" approach, which means that no a priori knowledge about the properties of the underlying process is needed to identify the hidden persistent phases. The method is based on the minimization of the functional describing the weighted distance between the observed data and their projections on a finite set of *K* linear manifolds. As a result, the method provides the probabilities with which the data points can be assigned to *K* hidden states characterized by *K* specific sets of essential dimensions. However, the numerical cost of the method is scaling quadratically with number of transitions between the hidden states, which seriously restricts the applicability of the method to the relatively short time series with few (\approx 10–20) transitions between the hidden states (HoSc).

The HMM-PCA is based on the same idea (the minimization of the distance functional) as the Wavelets-PCA method except for two additional assumptions made for the analyzed data: (i) the process switching between the metastable states is assumed to be Markovian and (ii) in each of the metastable states the projections of the data onto the dominant state-specific dimensions are Gaussian. Compared with the HMM-Gauss approach in terms of its assumptions, the HMM-PCA allows only the weakening of the constraint regarding the Gaussianity of the observed process in all of the dimensions. However, concerning the numerical gains of the method, it scales as $kn \log(n)$, where n is the observation dimension and $k \ll n$ is the number of principal components (because instead of the full covariance matrix inversion as in the HMM-Gauss method, HMM-PCA requires only the identification of k dominant eigenvectors, which can be achieved by applying Raley-Ritz or Lanczos methods). This property, together with linearity of the method w.r.t. the length of the time series, makes HMM-PCA applicable for analysis of high-dimensional time series. However, the Markov assumption about the hidden process restricts the applicability of the method to data without memory.

If the structure of the data allows some insight into the type of the underlying dynamics [e.g., the type of the noise process (additive or multiplicative)], then this additional information can be used in the construction of more specific methods of data analysis. As was demonstrated in our recent paper, one can construct methods combining HMM–PCA with fitting of reduced stochastic differential equations (Horenko et al. 2008). As was demonstrated on historical temperature data in Europe, the resulting HMM–PCA–SDE method can be used for predictions and identification of the metastable states even in very high dimensions. However, this method inherits the drawback of the previous methods concerning the non-Markovianity of the analyzed data. Moreover, as was shown for the temperature data example, the metastability analysis of real meteorological data is "spoiled" with the seasonal trend that results in identification of four seasons as metastable states. The above-described numerical problems of the underlying EM algorithm prohibit reliable identification in cases in which many metastable states are involved, especially in cases in which the time series are relatively short, as in historical meteorological data.

In this paper we describe a hierarchical approach based on successive decomposition of the multidimensional time series in metastable states. Such an approach is especially useful for relatively short but multidimensional time series with many hidden states because simultaneous identification of all of the hidden states would be hampered by the large uncertainty of the parameter identification and the nonuniqueness of the EM optimization result. The resulting method is capable of dealing with data gaps (resulting from the separation of the data on the previous hierarchical level of analysis). We also demonstrate how to use the idea of extended space representation to cast processes with memory into the Markovian framework (thereby fulfilling the first assumption of the HMM–PCA method). We discuss the assumptions needed for the construction of a new likelihood model of the data with gaps and propose a modified EM algorithm for log-likelihood optimization. We explain how the quality of the resulting reduced representation of the data can be acquired, how it can help to estimate the number of the metastable states, and what kind of additional information about the analyzed process can be gained. We illustrate the performance of the new method analyzing non-Markovian 500-hPa geopotential height fields [daily mean values from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) dataset for a period of 44 winters] and compare the outcome to the results obtained with the Wavelets-PCA approach. We interpret the results w.r.t. the notion of blocking events in the atmosphere.

2. Topological dimension reduction in time series analysis

a. Memory in the data and Markovian representation

Let the observed data be given in the form of a timediscrete sequence $\{z_t\}_{t=1,...,T}$ of *c* dimensional data vectors that describe the observation or measurement of a process at *T* subsequent instances. We will say that the process underlying the observations has a memory depth $d \ge 0$ if the conditional probability distribution *P* of future states of the process, given the present state and all past states, depends only on the present state and *d* previous states but not on all past states. Mathematically, this property can be expressed as

$$P(z_{t+1} \mid z_1, z_2, \dots, z_t) = P(z_{t+1} \mid z_{t-d}, \dots, z_{t-1}, z_t).$$
(1)

We will call a process Markovian if d = 0. For $D \ge d > 0$ it is obvious that the extended stochastic process $x_t^{(D)} = (z_t, z_{t-1}, \dots, z_{t-D})$ (which we will call a *d*-frame recasting of the original process) is Markovian; i.e.,

$$P(x_{t+1}^{(D)} \mid x_1^{(D)}, x_2^{(D)}, \dots, x_t^{(D)}) = P(x_{t+1}^{(D)} \mid x_t^{(D)}).$$
(2)

We will further omit the upper index D to simplify the notation.

This means that any observed process with finite memory can be cast into the *Dc*-dimensional extended space and become Markovian (allowing us to apply Markovian techniques of time series analysis, such as HMMs).

There are two major problems associated with this strategy: (i) reliable estimation of the memory depth d is not a trivial task if the dimension c of the observation data is high and (ii) the numerical cost of the time series analysis increases significantly for large D because the dimension of the extended space is D times larger than the dimension of the original space.

The first of the abovementioned problems becomes even more serious if the physics of the underlying process is unknown, that is, if it is not a priori clear what kind of stochastic dynamics should be expected (linear or nonlinear, additive or multiplicative noise, etc.). Linear approaches, such as multivariate autoregressive processes (MVARs; Brockwell and Davis 2002), can be used for estimation of d in multiple dimensions. However, such analyses do not guarantee reliability because there are examples of systems with finite nonlinear memory (e.g., the time series of stock returns in finance) where linear analysis methods do not reveal any significant memory effects (Tsay 2005). Another problem of such methods is their high numerical cost: the MVAR method, for example, scales as $O(c^6)$. This prohibits the application of these methods to very highdimensional systems without making additional assumptions about the analyzed data (the single dimensions are statistically independent, etc.).

On the other hand, the reported examples of application of nonlinear memory estimation methods, like conditional heteroscedastic models [such as ARCH (Tsay 2005) or its generalizations], are limited to specific application areas (like econometrics and financial data analysis) and low-dimensional cases; in general they do not allow a robust estimation for very large datasets.

b. State-specific dimension reduction

All of the above arguments underline the importance of dimension reduction methods in time series analysis. To be able to find hidden metastable states in very high-dimensional data, one should be able to couple the problem of the identification of those states to an appropriate dimension reduction strategy. We will now briefly outline the main idea of one such approach, the topological dimension reduction (Horenko et al. 2006; HoSc; Horenko et al. 2008).

Let us assume that with the help of one of the methods described above, we were able to estimate the upper bound D of the memory depth for the given time series $\{z_i\}_{i=1,...,T}$. It is worthwhile to mention that we do not need to determine the memory depth exactly because all we are interested in later on is to cast the process into Markovian framework, as explained above. Therefore we need a lower bound on D. To account for memory effects in the analyzed data, we can extend the vector space of observables z_t at each time t with D previous observations $\{z_{t-1}, \ldots, z_{t-D}\}$. The resulting vector $x_t = \{z_t, z_{t-1}, \dots, z_{t-d}\}$ is a component in n = Dc-dimensional space. The idea of the method is to identify the *m* principal directions with the highest variance in *n*-dimensional data $x_i (m \ll n)$. In contrast to standard PCA, in which these principal directions are supposed to be global (i.e., valid for the whole time series x_t), the idea of state-specific topological dimension reduction consists of the assumption that the principal directions can vary in time and are defined with the help of a sequence of K linear projectors $\mathbf{T}_i \in$ $\mathbb{R}^{n \times m}$, $i = 1, \ldots, K$; that is, **T**_i is understood to project onto the subspace spanned by the local principal directions. Mathematically the problem of identifying \mathbf{T}_i can be stated as a minimization problem w.r.t. the residuum functional, describing the least squares difference between the original observation and its reconstruction by means of the *m*-dimensional projection

$$\mathbf{L}(x_{t}, \mathbf{T}_{i}, \mu_{i}) = \sum_{i=1}^{K} \sum_{t=1}^{T} \gamma_{i}(t) \| (x_{t} - \mu_{i}) - \mathbf{T}_{i} \mathbf{T}_{i}^{\mathrm{T}} (x_{t} - \mu_{i}) \|_{2}^{2},$$
(3)

where $\gamma_i(t)$ (i.e., the hidden path) denotes the probability to optimally describe the *n*-dimensional vector x_t at time *t* with the local projector \mathbf{T}_i and $\sum_{i=1}^{K} \gamma_i(t) = 1$ for all *t*. The quantity $\gamma_i(t)$ provides a relative weight to the statement that an observation x_t belongs to the *i*th hidden state. For the moment we assume the sequence of probabilities $\gamma_i(t)$ to be known and fixed; in the next section we will present a way to estimate this sequence from a given observation x_i . The functional **L** depends on the projector matrices \mathbf{T}_i and center vectors $\mu_i \in \mathbb{R}^n$. Moreover, the projectors \mathbf{T}_i are subject to the orthogonality condition

$$\mathbf{T}_i^{\mathrm{T}} \mathbf{T}_i = I d^{m \times m}. \tag{4}$$

The solution of the optimization problem (3) subjected to orthogonality constraints (4) is possible in three cases (HoSc).

1) CASE 1: KNOWN HIDDEN PATH

If the hidden path $\gamma_i(t)$ is known, then the minimum of the functional (3) can be found, analytically resulting in a state-specific version of the PCA:

7

$$\left[\sum_{t=1}^{T} \gamma_i(t) (x_t - \mu_i) (x_t - \mu_i)^{\mathrm{T}}\right] \mathbf{T}_i = \mathbf{T}_i \Lambda_i, \qquad (5)$$

$$\mu_i = \frac{\sum_{t=1}^{T} \gamma_i(t) x_t}{\sum_{t=1}^{T} \gamma_i(t)},$$
(6)

where Λ_i is a matrix with *m* dominant eigenvalues of the weighted covariance matrix $\Sigma_{t=1}^T \gamma_i(t)(x_t - \mu_i)(x_t - \mu_i)^T$ on the diagonal (nondiagonal elements are zero); that is, each of the *K* hidden states is characterized by a specific set of essential dimensions \mathbf{T}_i (which can be defined as corresponding dominant eigenvectors) and center vectors $\mu_i \in \mathbb{R}^n$ calculated from the conditional averaging of the time series w.r.t. corresponding occupation probabilities $\gamma_i(t)$ (Horenko et al. 2006).

2) CASE 2: HMM–PCA

Let us make the following two assumptions: (i) the unknown sequence of hidden probabilities $\gamma_i(t)$ can be assumed to be an output of the Markov process X_t with K states and (ii) the probability distribution $P(\mathbf{T}_i x_t | X_t)$ = i) (which is the conditional probability distribution of the projected data in the hidden state i) can be assumed to be Gaussian in each of the hidden states. If both of these assumptions hold then the HMM framework can be used and one can construct a special form of EM algorithm to find the minimum of the residuum functional (3) [for details of derivation and resulting algorithmic procedure, please refer to our previous works Horenko et al. (2006) and HoSc]. The resulting method is linear in T and scales as $O(mn^2)$ with the dimension of the problem and as $O(K^2)$ with the number K of the hidden states. However, as with all of the likelihoodbased methods in an HMM setting, HMM–PCA does not guarantee the uniqueness of the optimum because the EM algorithm converges toward a local optimum of the likelihood function.

3) CASE 3: WAVELETS-PCA

Because both of the HMM–PCA model assumptions are very difficult to check (especially for highdimensional data), we need to construct a method that is free of those assumptions and which we can use for a posteriori verification of the HMM–PCA results. Therefore, we assume that the unknown function $\gamma_i(t)$ can be represented as a finite linear combination of (few) discrete Haar-wavelet functions $\phi(x)$:

$$\phi(x) = \xi_{(0,1)} = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \text{any other.} \end{cases}$$
(7)

With their help any hidden occupation probability function $\gamma_i(t) \in \mathbf{L}_2(\mathcal{R})$ for a fixed scaling factor $J \in \mathbb{Z}$ can be represented by a linear combination of the scaled wavelet-functions:

$$\gamma_i(t) = \sum_{r \in \mathbb{Z}} c_r^i \phi(2^J t - r)$$
$$c_r^i = \int_0^1 \gamma_i [2^{-J}(r+s)] \, ds. \tag{8}$$

If the number of ansatz functions involved in expansion (8) can be assumed to be small, it allows us to project the original high-dimensional optimization problem to the low-dimensional space of the wavelet coefficients c_r^i . The integral transformation between the wavelet representation and the occupation probabilities $\gamma_i(t)$ can be efficiently implemented using the fast Haar-wavelet transformation (FWT; Strang and Nguyen 1997).

In our specific implementation of the wavelet-based optimization procedure (HoSc), we made two simplifying assumptions: (i) we assumed that the occupation probability functions $\gamma_i(t)$ can take only discrete values 0 and 1 (i.e., the occupation probabilities are assumed to be discrete step functions) and (ii) we fixed the upper limit of the Galerkin subspace dimension for each of the optimization runs (i.e., together with the first assumption, it means that we set the upper limit of transitions between *K* hidden states).

The main advantage of the resulting Wavelet–PCA approach is that it is independent of the model assumptions (Markovianity and Gaussianity) of the HMM– PCA method. However, our specific implementation of the method scales quadratically with the number of involved Haar-wavelet functions; that is, the method is not applicable for very long time series with large numbers of transitions between the hidden states. But it can be used for validation of the model assumptions of the HMM–PCA by comparison of the $\gamma_i(t)$ values identified by both methods for relatively short segments of the analyzed time series.

3. Hierarchical approach

As demonstrated above, the application of the hidden Markov framework to the HMM-PCA approach results in a specific assumption about causal dependence inside of the data series. It means that the construction of the likelihood function implies that (i) the data sequence being subjected to the HMM-PCA analysis has to be contiguous and (ii) the time intervals between the consecutive observations should be equal (Horenko et al. 2006). Whereas assumption (ii) is usually satisfied for most of the available datasets, assumption (i) is much more restrictive because there are a lot of processes which cannot be permanently observed (e.g., financial data are available only during trading sessions on the stock market and are not available on weekends and holidays). Assumption (i) will also prohibit the application of the HMM-PCA in cases for which one is interested in analyzing only specific segments of available data (e.g., the meteorological data restricted to certain seasons) or in which the time series is subjected to hierarchical decomposition into metastable substates. It is worth mentioning that one can still apply the Wavelets-PCA method in all of these cases but, as was already mentioned above, the applicability of Wavelets-PCA is restricted to the cases where there are only a few transitions between the hidden states.

Therefore, we are interested in extension of the HMM-PCA framework toward the cases where there are gaps in observation sequence where the causal dependence implied by a Markovian model is broken. To cast the description of the data into the HMM framework, we first define the complete observation set $X_t = (X_t, x_t)$, where X_t is an output of some unobserved data (or hidden Markov chain) and x_t is observed data. We will further assume that (i) the observation data $\{x_t\}_{t=1}, \ldots,$ T consists of a sequence of N_{traj} contiguous observation sequences x^{i} , that is, $\{x_{t}\}_{t=1,...,T} = \{x^{1}, x^{2}, ..., x^{N_{\text{traj}}}\};$ (ii) the time intervals between subsequent observations in each of the contiguous data sequences are equal, and (iii) the gaps between the neighboring data sequences are so big that each consecutive data sequence can be assumed to be statistically independent from the predecessor sequence. We will refer to the original time series as a data sequence, whereas the contiguous segments of it with time-equidistant observations will be called subsequences. The last assumption means that the following relation is valid for a joint conditional probability distribution function $P(\chi_t|\lambda)$ (also called the likelihood):

$$P(X_1, X_2, \dots, X_T | \lambda) = \prod_{l=1}^{N_{\text{traj}}} P(X_{l_1}^l, \dots, X_T^l | \lambda), \quad (9)$$

where $\lambda = (\pi, \mathbf{A}, \pi_1, \mathbf{T}_1, \dots, \pi_K, \mathbf{T}_K)$, **A** is the transition matrix of the hidden Markov process X_t , π is the invariant distribution of initial states of the hidden process, and (μ_i, \mathbf{T}_i) are parameters of essential linear manifolds characteristic for each of the hidden states $(t_1^l \text{ and } T^l \text{ define the start and the end of the contiguous subsequence <math>l$ inside of the observation data).

We define the log-likelihood functional of the process as

$$\mathbf{L}^{\log}(\lambda|\mathbf{X}) = \log P(X_1, X_2, \dots, X_T|\lambda)$$
$$= \sum_{l=1}^{N_{\text{traj}}} \mathcal{L}(\lambda|X_{t_1}^l, \dots, X_T^l), \quad (10)$$

where \mathcal{L} are standard HMM–PCA log-likelihood functions for contiguous time series with observations equidistant in time (Horenko et al. 2006).

We employ the EM algorithm to maximize both likelihood and log-likelihood functions simultaneously. Starting with some initial model λ_0 , we iteratively refine the model within two steps: the expectation step and the maximization step.

a. The expectation step

In this step, the state occupation probability $\gamma_t^l(i) = P(X_t = i | x_t, \lambda)$ and the transition probability $\eta_t^l(i, j) = P(X_t = i, X_{t+1} = j | x_t, \lambda)$ are calculated for each time $t \in [t_1^1, \ldots, T^l]$, given the observation x_t and the current model λ . To calculate the two conditional probabilities of the expectation step (or E-step), we first define two additional variables:

$$\alpha_t^l(i) = P(x_{t_1}^l \dots x_t, X_t = i | \lambda) \quad \text{and} \tag{11}$$

$$\beta_t^l(i) = P(x_{t^{1}+1}, x_{t^{l}+2} \dots x_{T^l} | X_t = i, \lambda), \qquad (12)$$

where $\alpha_l^l(i)$ and $\beta_l^l(i)$ are forward and backward variables, respectively. The interpretation of $\alpha_l^l(i)$ is as follows: it denotes the probability of the observation subsequence l up to time t together with the information that the system is in hidden state i at time t conditioned w.r.t. the given model parameters λ . The following formulas show that the computation of the sequence $\alpha_l^l(i)$ for the whole sequence is possible with $K^2 T$ operations:

$$\alpha_{t_1^l}^l(i) = \pi_i \rho_i(x_{t_1^l}), 1 \le i \le K,$$
(13)

$$\alpha_{t+1}^{l}(j) = \left[\sum_{i=1}^{K} \alpha_{t}^{l}(i)A_{ij}\right] \rho_{j}(x_{t+1}),$$
(14)

$$1 \le t \le T - 1, 1 \le j \le K,$$
 (15)

where $\rho_j(x_{t+1}) = \rho(x_{t+1}|X_{t+1} = j)$ defines the conditional observation probability of the data at time t + 1 in the hidden state *j*. The backward variable $\beta_t^l(i)$ can be computed with analogous formulas:

$$\beta_T^l(i) = 1, 1 \le i \le K \tag{16}$$

$$\beta_{t}^{l}(i) = \sum_{j=1}^{K} A_{ij} \rho_{j}(x_{t+1}) \beta_{t+1}^{l}(j), \qquad (17)$$

$$t = T^{l} - 1, T^{l} - 2, \dots, 1, 1 \le i \le K.$$
 (18)

Using the definitions (11) and (12), one can finally write for the probability $P(x_{t'1}, \ldots, x_{T'}|\lambda)$ (Bilmes 1998)

$$P(x_{t_1'}, \dots, x_{T_l} | \lambda) = \sum_{i=1}^K \beta_1^l(i) \pi_i \rho_i(x_{t_1'})$$
(19)

or equivalently

$$P(x_{t_1}^{\prime},\ldots,x_T^{\prime}|\lambda) = \sum_{i=1}^K \alpha_t^{\prime}(i)\beta_t^{\prime}(i), \qquad (20)$$

where *l* is chosen in such a way that $t \in [t_1^l, \ldots, T^l]$. The two conditional probabilities of the E-step can be calculated efficiently by using the forward–backward variables:

$$\eta_{t}^{l}(i,j) = \frac{\alpha_{t}^{l}(i)A_{ij}\rho_{j}(x_{t+1})\beta_{t+1}^{l}(j)}{P(x_{t}^{l},\ldots,x_{T}^{l}|\lambda)}.$$
(21)

With these values, the probability of being in state *i* at time *t* can be expressed as

$$\gamma_{t}^{l}(i) = \sum_{j=1}^{K} \eta_{t}^{l}(i, j).$$
 (22)

Note that the expected number of transitions from *i* to any other state (including itself) within the whole observation is $\sum_{l=1}^{N_{\text{traj}}} \sum_{t=t_1^{T^{l-1}}}^{T^{l-1}} \gamma_t^l(i)$, and the expected number of transitions from *i* to *j* is $\sum_{l=1}^{N_{\text{traj}}} \sum_{t=t_1^{t}}^{T^{l-1}} \gamma_t^l(i, j)$.

b. The maximization step

This step finds a new model $\hat{\lambda}$ via a set of reestimation formulas. The maximization guarantees that the likelihood does not decrease in each iteration.

To apply the EM algorithm, we need to re-estimate parameters λ describing the hidden Markov model and

essential linear manifolds via the maximum likelihood estimator. Thus, the observation x_t at time $t \in [t_1^l, \ldots, T^l]$ has to be weighted with the probability for the hidden state $i \gamma_t^l(i)$ for the respective subsequence l. To calculate this re-estimation, we fix the sequence X_t of hidden states [this means also keeping the sequence of $\gamma_t^l(i)$ fixed] and calculate the derivatives of the functional (10) w.r.t. the parameter set λ . By setting all of the partial derivatives to zero for some fixed reduced dimensionality m we get a coupled system of nonlinear algebraic equations for the parameters that can be solved analytically [analogous to the derivation shown in Horenko et al. (2006) and HoSc]. We will skip the derivation here and just present the final re-estimation formulas:

$$\mu_{i} = \frac{1}{\sum_{l=1}^{N_{\text{traj}}} \sum_{t=t_{1}^{l}}^{T^{l}-1} \gamma_{t}^{l}(i)} \sum_{l=1}^{N_{\text{traj}}} \sum_{t=t_{1}^{l}}^{T^{l}-1} \gamma_{t}^{l}(i) x_{t}, \qquad (23)$$

 $\operatorname{Cov}_{i}\mathbf{T}_{i} = \mathbf{T}_{i} \max_{m}[\operatorname{spec}(\operatorname{Cov}_{i})], \qquad (24)$

where $\max_m[\operatorname{spec}(\operatorname{Cov}_i)]$ denotes *m* dominant eigenvalues of the covariance matrix Cov_i ;

$$\operatorname{Cov}_{i} = \frac{1}{\sum_{l=1}^{N_{\operatorname{traj}}} \sum_{t=t_{1}^{l}}^{T^{l}-1} \gamma_{i}^{l}(i)} \sum_{l=1}^{N_{\operatorname{traj}}} \sum_{t=t_{1}^{l}}^{T^{l}-1} \gamma_{i}^{l}(i)(z_{t} - \mu_{i})(z_{t} - \mu_{i})^{T}}.$$
(25)

The E and M steps are iteratively repeated until a predetermined maximal number of iterations is reached or the improvement of the likelihood becomes smaller than a given limit. The entire EM algorithm has the nice property that the likelihood function is nondecreasing in each step (i.e., we iteratively approximate local maxima). We will call the presented method ensemble HMM–PCA to refer to the ability of the new method to deal with an ensemble of statistically independent subsequences and to stress the difference from the standard HMM-PCA. As for the scaling of numerical effort, the resulting ensemble HMM-PCA method is linear in the length of the observation series x_t , quadratic in the number K of hidden Markov states (essentially because the transition matrix elements of the hidden Markov chain should be estimated), and scales as $O(mn^2)$ in the reduced dimensionality m (because only m dominant eigenvectors of Cov, matrix are required, they can be obtained with numerically efficient subspace methods such as the Raley-Ritz iteration or Lanczos method). Therefore the ensemble HMM-PCA

approach is applicable to systems with very high dimensionality and very long observation data sequences. This feature is demonstrated in section 5 where the method is used for analysis of a multidimensional meteorological dataset.

4. Estimation of confidence intervals and choice of K

It is intuitively clear that the quality of the resulting reduced model is very much dependent on the original data, and especially on the length of the available time series. The shorter the observation sequence, the bigger the uncertainty of the resulting parameters. The same is true if the number K of hidden states is increasing for the fixed length of the observed time series: the bigger K is, the higher the uncertainty will be for each of the states. Therefore, to statistically distinguish between different hidden states we need to get some notion of the HMM–PCA robustness. This can be achieved through the estimation of confidence intervals for both parts of the model: the hidden Markov process and the extended EOFs.

a. Hidden Markov process

To estimate the confidence intervals of the hidden transition probabilities A_{ij} we first make use of the second derivatives $(\partial^2 \mathbf{L}/\partial A_{ij}^2)(\overline{A})$ (also called Fisher information) of the log-likelihood function (10) subject to the constraint

$$\sum_{j=1}^{K} A_{ij} = 1, \forall i = 1, \dots, K$$
(26)

(**A** is the hidden transition matrix of the Markov chain estimated by the HMM–PCA algorithm). We denote the number of the transitions in the identified Markovian sequence X_i between the states *i* and *j* as N_{ij} . The most probable sequence X_i of the hidden states can be directly computed from the hidden probabilities $\gamma_i^l(i)$ by applying, for example, the Viterbi algorithm (Viterbi 1967). Then it is easy to verify that the explicit expression for the Fisher information of the identified Markov chain X_i is

$$\frac{\partial^2 \mathbf{L}^{\log}}{\partial A_{ij}^2}(\overline{A}) = -\frac{\left(\sum_{k=1}^K N_{ij}\right)^2}{N_{ij}}.$$
 (27)

The confidence intervals of the hidden Markov process are then given by $[\overline{A}_{ij} - \delta(\overline{A}_{ij}), \overline{A}_{ij} + \delta(\overline{A}_{ij})]$, where

$$\delta(\overline{A}_{ij}) = 1.96 \left[-\frac{\partial^2 \mathbf{L}^{\log}}{\partial A_{ij}^2} (\overline{A}) \right]^{-0.5};$$
(28)

the multiplier 1.96 comes from the definition of 95% confidence interval in Gaussian statistics.

b. Extended EOFs

The Gaussianity assumption for the observation process in the HMM–PCA method gives an opportunity to estimate the confidence intervals of the manifold parameters (μ_i , \mathbf{T}_i) straightforwardly. This can be done in a standard way of multivariate statistical analysis because the variability of the weighted covariance matrices (25) involved in the calculation of the optimal projectors \mathbf{T}_i is given by the Wishart distribution (Mardia et al. 1979). The confidence intervals of \mathbf{T}_i can be estimated by sampling from this distribution and calculating the *m* dominant eigenvectors of the sampled matrices, whereas the confidence intervals of μ_i can be acquired from the respective standard deviations (Mardia et al. 1979).

c. Optimal choice of K

If there exist two states with confidence intervals overlapping for each of the respective reduced model parameters, then those are statistically indistinguishable and K should be reduced and the HMM–PCA calculation repeated. In other words, confidence intervals implicitly give a natural upper bound for the number of hidden states. On the other hand, the spectral theory of the Markov processes connects the number Kof metastable states with the number of the dominant eigenvalues in the so-called Perron cluster (Schütte and Huisinga 2003). This allows us to apply the Perron cluster–cluster analysis (PCCA; Deuflhard and Weber 2005) to find the lower bound of K. Both these criteria in combination can help to find the optimal number Kof the hidden states in each specific application.

5. Analysis of the hidden transition matrix

Application of the HMM–PCA algorithm to the analyzed multidimensional data results in a twofold dimension reduction: in addition to the identification of dominant local extended orthogonal functions describing the directions of maximal data variability, HMM–PCA reveals a hidden discrete Markov process switching between different sets of those extended EOFs. Analysis of the corresponding hidden transition matrix A can help us to understand the global properties of the underlying multidimensional dynamics, which is now given by the series of one-dimensional discrete hidden variable X_t . We will now briefly sketch some of these properties and explain how to calculate them. For more details, we refer the reader to the standard literature on Markov chains (e.g., Gardiner 2004).

a. Relative statistical weights

The vector π of relative statistical weights of the hidden states can be calculated as the fixed point of the Markovian transition operator; that is,

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{A}.\tag{29}$$

Note that we use multiplication from the left because **A** is a stochastic matrix with row sums all equal to 1.0.

b. Mean exit times

The mean exit time τ_i^{ex} is the expected time for the process X_i to stay in the hidden state *i* until it switches to any other state. Thus, it is one of the basic quantities and can be used to compare different hidden states w.r.t. their metastability. It can be directly computed from the diagonal elements of the transition matrix **A**:

$$\tau_i^{\rm ex} = \frac{\delta t}{1 - A_{ii}},\tag{30}$$

where δt is the time step between the observations.

c. Mean first passage times

For any pair of two different hidden states *i* and *j*, the mean first passage time τ_{ij}^{pas} represents the expected time for the process X_t to start in the state *i* and to reach the state *j* for the first time. It can be calculated from the solution of the following linear system of equations:

$$\tau_{ij}^{\text{pas}} = \begin{cases} \delta t + \sum_{k=1}^{K} \tau_{kj}^{\text{pas}} A_{ik}, & i \neq j, \\ 0, & i = j. \end{cases}$$
(31)

This quantity describes the dynamical properties of the process X_t and can be used to analyze and compare different transition pathways between metastable states.

6. Analysis of historical geopotential height data

a. Description of the data

Using the method presented in the previous sections, we analyze daily mean values of the 500-hPa geopotential height field from the ERA-40 data (Simmons and Gibson 2000). We consider a region with the coordinates 32.5°–75.0°N, 27.5°W–47.5°E, which includes Europe and a part of the eastern North Atlantic. The combination of land and sea makes the selected region pref-





FIG. 1. (left) Dominant eigenvalues of the autocorrelation matrix and (right) AR coefficients, both computed for different time lags.

erable for the appearance of dynamically relevant phenomena; it also captures the area of maximum Atlantic block formation (Wiedenmann et al. 2002). The resolution of the data is 2.5° , which implies a grid with 31 points in the zonal and 18 in the meridional direction. We have also tested the sensitivity of the results presented here by reducing the resolution by a factor of 2, taking only 16×9 grid points.

For the analysis we have considered geopotential height values only for winter and for the period 1958/59 to 2001/02, where a winter includes the months December to February; thus, we end with a nonequidistant time series of 3960 days. The reasons for considering winter months only were (i) because of the increased equator-to-pole temperature gradient, the synoptic eddies and the quasi-stationary Rossby waves in the atmosphere are much more intense during winter, which suggests much more pronounced regime behavior, and (ii) if we focus on blocking events only, representing a kind of metastability in the circulation, there is a pronounced maximum in the block formation for the considered region during winter (Lupo et al. 1997).

We have mentioned already in the introduction the problem with the seasonal cycle when analyzing atmospheric data w.r.t. metastable behavior. To remove the seasonal trend we apply a standard procedure, where from each value in the time series we subtract a mean build over all values corresponding to the same day and month (e.g., from the data on 1 January 1959 we subtract the mean value over all days which are the first of January, and so on).

b. The blocking index

For the purpose of interpreting the results of the presented method w.r.t. metastability of blocking events, we compute the Lejenas–Okland index from the data. It indicates the appearance of a blocking anticyclone and the duration of the event. We have a blocking if the geopotential height difference at 500 hP between 40°N and 60°N is negative over a region with 20° zonal extent. The exact formula is given in Lupo et al. (1997); for the purpose of representation we have computed a zonally averaged value of the index, rescaled it, and reversed its sign. (A part of the time series of the index is shown in Fig. 7.)

c. Discussion of the results

To choose the lower bound of the frame length in the algorithm, the memory depth of the data was estimated from the autocorrelation and partial autocorrelation function. The dominant eigenvalues of the autocorrelation matrix and of the autoregressive (AR) coefficients computed at different time lags are presented in Fig. 1. From the spectrum of the AR coefficients one can see that the data has an internal memory of about 5 days and it can be approximately modeled by an autoregressive process of the order 5; the oscillations after the fifth day are interpreted as noise. We conclude that a frame length of 5 days will be sufficient to make the data Markovian.

To choose the optimal number of hidden states K, we first start the HMM–PCA algorithm with K = 8 for different values of d = 1, 5, 10, 20, and 40 and m = 1.



FIG. 2. Spectrum of the hidden transition matrix A for K = 8. Only the first four dominant eigenvalues are statistically significant because the parameter confidence intervals of the hidden states corresponding to the lower part of the spectrum (eigenvalues 5–8) are overlapping. This indicates K = 4 as the upper bound for the number of statistically distinguishable hidden states.

As mentioned above, because only a relatively short time series is available we need first to estimate the upper bound for K by comparing the confidence intervals of HMM-PCA parameters. To avoid the inherent problem of the EM algorithm-namely, that it only converges to the local maximum of the likelihood functional (dependent on the initial parameter values)-we perform the optimization with different randomly chosen sets of initial parameters 100 times and take the result with maximal likelihood. One of the transition matrix spectra is shown in Fig. 2. If the confidence intervals for a pair of states are overlapping it means that the corresponding states are statistically undistinguishable and the whole optimization procedure should be repeated for K = K - 1. It comes out that only for K =4 are all of the hidden states statistically distinguishable; therefore, we proceed further with 4 hidden states.

Next, we have to verify the assumptions needed to apply the HMM–PCA method. The first possibility is to a posteriori check the Gaussianity of the data in the hidden states and the Markovianity of the hidden process. However, it will not guarantee that these assumptions will also be fulfilled in any of the EM iterations. Another possibility is to compare the results of the HMM–PCA optimization with, for example, some fragment of Wavelets–PCA results (because Wavelets– PCA is much slower but does not imply any assumptions about the analyzed data). This will give us a possibility to estimate the robustness of optimization w.r.t. the model assumptions. As we see from Fig. 3, the re-



FIG. 3. Comparison of hidden Viterbi paths identified with the Wavelets–PCA (dashed) and HMM–PCA (solid) algorithms (both for K = 4, d = 5).

spective Viterbi paths are almost identical for both methods; therefore, it verifies the usage of the HMM– PCA analysis.

Next, we have studied the sensitivity of the results w.r.t. different frame lengths. The calculated Viterbi paths, showing the most probable sequence of hidden states, are displayed in Fig. 4. When the frame length increases, the transitions between the hidden states reduce and the occupation duration increases. The discrepancy of the Viterbi paths for different frame lengths can happen because data with smaller frame lengths are non-Markovian but the algorithm can still find some metastable regime behavior, which is filtered out if the larger frame length is applied.

We have tested the dependence of the results on the resolution, using data on a 16×9 and on a 32×18 grid for the analysis. The Viterbi paths for both grids are shown in Fig. 4; they are nearly identical. Figures 5 and 6 display the center vectors μ_i for the two different resolutions and d = 1. In both cases, the large-scale structure of the pattern is captured by the algorithm.

From Figs. 5 and 6, we see that the hidden states describe two different regimes: μ_1 and μ_3 are characterized by a negative geopotential anomaly at higher latitudes and a positive anomaly at lower latitudes, whereas the other two states, μ_2 and μ_4 , have anomalies reversed in sign. Thus, the states in the first regime are associated with an intensification of the zonal flow and those in the second regime with a weakening of it. Each regime can be then subdivided into states with stronger anomalies (μ_3 and μ_4) and weaker anomalies (μ_1 and μ_2).

We expect that blocking events will be captured mostly by hidden state 4; this is confirmed if we plot the



FIG. 4. The Viterbi path of the hidden Markov chain for different resolutions of the data and different frame lengths d: results from data on a 32 × 18 grid (solid) and on a 16 × 9 grid (dotted–dashed; only for d = 1, 5).

probability γ_4 and the blocking index (see Fig. 7). From the Viterbi paths and the blocking index, we calculated that state 4 and state 2 respectively capture 46% and 36% of all blocking events. If we consider as blocking situations cases in which the blocking index is negative over a period larger than 6 days (filtered index), the numbers above change to 58% and 29%, respectively. Looking at individual events, we found that the two states also represent other weather patterns with an anomalous geopotential gradient (e.g., cut-off lows). Nevertheless, about 73% of all days in state 4 are associated with blockings; for state 2 this number is 47%. If we consider the filtered blocking index, the numbers change to 52% and 21%, respectively.

Calculating the projection matrix \mathbf{T}_i , we find the leading *m* EOFs within each of the hidden states and compare the variance patterns computed in this way with those from a standard EOF analysis of the dataset. A particular difference is the absence of the first EOF pattern from the standard method in the local EOFs \mathbf{T}_i from the HMM–PCA algorithm. This mode describes the variability of the meridional geopotential gradient and, as discussed above, such a dynamics is already captured by the time evolution of the functions γ_i , i = 1, ..., 4. The leading three variance patterns produced by the HMM–PCA algorithm looked very similar for the four hidden states; other EOFs differed if the corresponding states had positive/negative or weak/strong geopotential anomalies.

But how do the results change when we make the data Markovian, considering an extended space with the dimension n = d * c? We can split the center vector μ_i into d parts with the original dimension c, representing the mean state of the system at different time lags. The resulting sequence can be interpreted as the "mean time evolution" of the mean state in *i*. Fig. 8 displays such a sequence for μ_4 , showing the growth in time of the meridional geopotential gradient anomaly.

To represent the results for larger frame lengths and different states, we have computed the geopotential height difference between 40°N and 60°N from the vector μ_i at different time lags, using exactly the same criteria as for the calculation of the blocking index (see section 2b), but now we consider all values, not only the negative one. The results are displayed in Fig. 9. We see that the overall time evolution is characterized by a growth or a decay of the meridional geopotential gradient, which for q = 5 reaches at the end its values from the analysis with q = 1. For larger frame lengths, the amplitude of the gradient is strongly reduced but the time evolution shows more complex character with changing phases of decay and growth (e.g., state 4 in the



FIG. 5. The vector μ_i for the four different hidden states computed using a frame length of 1 day and data on the 31 \times 18 grid.

case of q = 40). This can probably be explained by the fact that because in those cases the duration of the blocking is smaller than the dynamical frame length d, many creations and/or destructions of the blocking situations are getting averaged out.

Because the proposed technique for frame lengths ≥ 2 is a special type of time-lagged statistics, it can be used to study onsets and withdrawal of diagnosed features. In this, given a time lag q, we have computed conditional composites for diagnosed events. For onsets, we have selected time slices t_j , $j = j_1, j_2, \ldots, j_{Ne}$ (N_e is the number of diagnosed events) when the occupation probability for the state $4(\gamma_4^l)$ reaches unity. This state was selected because it corresponds most closely to blockings as diagnosed by the employed blocking index. An additional condition is imposed that γ_4^l remains unity at least for five consequent days (a condition of persistence). For these time slices, a conditional average is computed:

$$x^{o}(q) = \frac{\sum_{t_{j=1}}^{N_{e}} \gamma_{4,t_{j-q}}^{l} x_{t_{j-q}}}{\sum_{t_{j=1}}^{N_{e}} \gamma_{4,t_{j-q}}^{l}}.$$
(32)

An analogous conditional average is computed for withdrawals by selecting the time slices t_k , $k = k_1$, k_2, \ldots, k_{N_e} as the last days of the diagnosed events when $\gamma_4^{t}=1$. After that, a conditional average for withdrawals is computed:

$$x^{w}(q) = \frac{\sum_{t_{k}=1}^{N_{e}} \gamma_{4,t_{k-q}}^{l} x_{t_{k-q}}}{\sum_{t_{k}=1}^{N_{e}} \gamma_{4,t_{k-q}}^{l}}.$$
(33)

In both cases q = 0, ..., d - 1, where d is the frame length.

We note different interpretations of $x^o(q)$ and $x^w(q)$. In the former case, q covers time intervals before the block onset. As a result, the composite $x^o(q)$ corresponds to typical synoptic conditions before the block onset. In contrast, for $x^w(q)$, q covers time moments when a block exists and, generally, is well developed. As a result, $x^w(q)$ has to be interpreted as a typical pattern of mature blocking state.

For onsets, the composite pattern exhibits developing meridional wavy structure (Fig. 10). This feature first appears in the southwestern part of the studied



FIG. 6. As in Fig. 5, but on the 16×9 grid.



FIG. 7. The time evolution of the zonally averaged blocking index (dashed) and the probability for hidden state 4 γ_4 (solid line).



FIG. 8. The vector μ_4 at different time lags *t* computed using a frame length of 5 days and the 31 \times 18 resolution.



FIG. 9. Zonally averaged geopotential height difference between 40°N and 60°N for the μ_i vectors at different time lags. The plots represent the results for different frame lengths *d*. The values for d = 1 are indicated by lines with markers in the first plot, in which squares, stars, circles, and diamonds correspond to states 3, 1, 2 and 4, respectively.



FIG. 10. The computed onset composite $x^{o}(q)$ at different time lags q. For x_{t_j} in (32) we have used data with an annual cycle; units are gpdm.

domain as a positive anomaly of geopotential height (q = 4 - 2). Afterward, at q = 1 - 0 this anomaly spreads to the east and becomes more pronounced, forming a ridge (a trough) in the southern (northern) part of the domain. Eventually, this trough-ridge system evolves to the blocked state. These features are common for the development of typical Atlantic blocking (Berggren et al. 1949; Rex 1950a; Diao et al. 2006).

For withdrawals (Fig. 11), we see a very marked positive anomaly of geopotential height in the southern part of the domain and a negative one in the northern part. Neither anomaly moves for different values of q within this composite. This emphasizes a stationarity of blockings within their life cycles. However, it becomes more marked if one travels from q = 4 to q = 0. The reason for this is the chosen length of frame, 5 days, which is comparable to the typical duration of blockings (e.g., Rex 1950b; Wiedenmann et al. 2002; Lupo et al. 1997; Diao et al. 2006; Croci-Maspoli et al. 2007). The fully developed anomaly spreads above the greater part of the northern Atlantic and attains a large magnitude.

Next we analyze the hidden transition matrix identified by the HMM–PCA in the Markovian case (K = 4, m = 1, d = 5). The transition graph correspondent to the identified matrix A is shown in Fig. 12. Each of the hidden states corresponds to a dynamical pattern of 5 days. As we have seen above in Fig. 9, each of the patterns is associated with specific blocking formation or destruction events. Therefore, by analyzing the transition graph from Fig. 12 we can gain some insight into the kinetics of such events. We start with the calculation of relative statistic weights of the respective hidden states. The solution of (29) yields $\pi_1 = 0.2363$, $\pi_2 =$ 0.1836, $\pi_3 = 0.4234$, and $\pi_4 = 0.1567$; that is, the dynamical pattern corresponding to the blocking formation in hidden state 4 is the most infrequent one. To compare the metastability of the hidden states, we can calculate the mean exit times τ_i^{ex} from (30). We get the following values: $\tau_1^{\text{ex}} = 4.3$, $\tau_2^{\text{ex}} = 5.3$, $\tau_3^{\text{ex}} = 14$, and $\tau_4^{\text{ex}} = 16$ days. Together with Fig. 12, these can be interpreted to mean that both 3 and 4 are metastable states, whereas 1 and 2 correspond to a transition pathway between them. Blocking events associated with the hidden state 4 represent a metastable event in the Markovian model: its typical duration is 16 days and two typical transition pathways in the system are $3 \rightarrow$ $1 \rightarrow 2 \rightarrow 4$ and $4 \rightarrow 2 \rightarrow 1 \rightarrow 3$. To characterize and compare these two pathways, we calculate the mean first passage times. As results from (31), $\tau_{34}^{\text{pas}} = 131$ and $\tau_{43}^{\text{pas}} = 49$ days; that is, it takes much longer to "create" a blocking situation then to "destroy" it. This is also in a good agreement with the respective statistical weights



FIG. 11. The computed withdraw composite $x^w(q)$ at different time lags q. For x_{t_j} in (33) we have used data with an annual cycle; units are gpdm.

 π of the corresponding states; the "unblocked" metastable state 3 is visited almost 3 times more frequently then the "blocked" state 4.

7. Conclusions

We have presented a numerical framework for the simultaneous identification of hidden states and respective essential orthogonal functions (EOFs) in highdimensional data with gaps. It allows us to construct reduced representation of analyzed data in the form of a discrete Markov jump process switching between different sets of EOFs. We discussed the model assumptions and explained the necessity of combining different methods relying on separate sets of model assumptions for data analysis.

We have also demonstrated what kind of additional insight into underlying dynamics can be gained from a



FIG. 12. Graphic representation of the hidden Markov process responsible for the metastable behavior of the analyzed time series (identified with the HMM–PCA algorithm for K = 4, m = 1, d = 5). Circles denote the hidden states and arrows show the connections between them; the numbers represent respective probabilities of transitions.

reduced Markovian representation (e.g., in the form of transition probabilities, statistical weights, mean first exit times, and mean first passage times). The proposed pipeline of data analysis based on HMM-PCA was exemplified in an analysis of 500-hPa geopotential height fields in winter. Correspondence between the hidden probability in one of the metastable states and the zonally averaged blocking index was found, and the respective mean dynamical patterns in the hidden states were found to be describing the creation and destruction of the blocking situations. We estimated a transition matrix (Fig. 12) of the hidden Markov process describing the transition probabilities between different atmospheric regimes. Respective Markovian processes give a reduced model for the dynamics of the 500-hPa geopotential field and can be used for predicting the blocking or strengthening of the zonal flow in operative weather forecast.

One of the basic problems of the multivariate meteorological data is that only relatively short fragments of the observation process are available for the analysis. Therefore it is very important to be able to extract the reduced description out of the data and to control the sensitivity of the analysis w.r.t. the length of the time series and the number K of the hidden states. We gave some hints for the selection of optimal K and explained how the quality of the resulting reduced representation can be acquired.

Acknowledgments. We are thankful to H. Oesterle who provided us with the ERA-40 reanalysis data from the European Center for Medium-Range Weather Forecasting. Illia Horenko's contribution was supported by the DFG research center Matheon "Mathematics for key technologies" in Berlin, and Stamen Dolaptchiev and Rupert Klein's contributions were partially supported by Deutsche Forschungsgemeinschaft, Grant KL 611/14.

REFERENCES

- Baum, L., 1972: An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, **3**, 1–8.
- Benzi, R., G. Parisi, A. Sutera, and A. Vulpiani, 1982: Stochastic resonance in climatic change. *Tellus*, 34, 10–16.
- Berggren, R., B. Bolin, and C. G. Rossby, 1949: An aerological study of zonal motion, its perturbations and break-down. *Tellus*, 1, 14–37.
- Bilmes, J., 1998: A gentle tutorial of the EM algorithm and its applications to parameter estimation for Gaussian mixture and hidden Markov models. International Computer Science Institute Tech. Rep., 13 pp.
- Brockwell, P., and R. Davis, 2002: Introduction to Time Series and Forecasting. 2nd ed. Springer, 434 pp.

- Charney, J. G., and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. J. Atmos. Sci., **36**, 1205–1216.
- Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. J. Atmos. Sci., 50, 2674–2696.
- Corti, S., F. Molteni, and T. N. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799–802, doi:10.1038/19745.
- Croci-Maspoli, M., C. Schwierz, and H. Davies, 2007: A multifaceted climatology of atmospheric blocking and its recent linear trend. J. Climate, 20, 633–649.
- Deuflhard, P., and M. Weber, 2005: Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.*, 398, 161–184.
- Diao, Y., J. Li, and D. Luo, 2006: A new blocking index and its application: Blocking action in the Northern Hemisphere. J. *Climate*, **19**, 4819–4839.
- Efimov, V. V., A. V. Prusov, and M. V. Shokurov, 1995: Patterns of interannual variability defined by a cluster analysis and their relation with ENSO. *Quart. J. Roy. Meteor. Soc.*, **121**, 1651–1679.
- Franzke, C., D. Crommelin, A. Fischer, and A. Majda, 2008: A hidden Markov model perspective on regimes and metastability in atmospheric flows. J. Climate, 21, 1740–1757.
- Gardiner, C. W., 2004: Handbook of Stochastical Methods for Physics, Cmeistry, and the Natural Sciences. 3rd ed. Springer-Verlag, 415 pp.
- Horenko, I., J. Schmidt-Ehrenberg, and C. Schütte, 2006: Setoriented dimension reduction: Localizing principal component analysis via hidden Markov models. *Computational Life Sciences II*, R. G. M. R. Berthold and I. Fischer, Eds., Lecture Notes in Bioinformatics, Vol. 4216, Springer, 98–115.
- —, R. Klein, S. Dolaptchiev, and C. Schuette, 2008: Automated generation of reduced stochastic weather models. I: Simultaneous dimension and model reduction for time series analysis. *Multiscale Model. Simul.*, 6, 1125–1145.
- Kimoto, M., and M. Ghil, 1993a: Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. J. Atmos. Sci., 50, 2625–2644.
- —, and —, 1993b: Multiple flow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions. J. Atmos. Sci., 50, 2645–2673.
- Lupo, A. R., R. J. Oglesby, and I. I. Mokhov, 1997: Climatological features of blocking anticyclones: A study of Northern Hemisphere CCM1 model blocking events in present-day and double CO₂ concentration atmospheres. *Climate Dyn.*, **13**, 181–195.
- Majda, A., C. Franzke, A. Fischer, and D. Crommelin, 2006: Distinct metastable atmospheric regimes despite nearly Gaussian statistics: A paradigm model. *Proc. Natl. Acad. Sci. USA*, **103** (22), 8309–8314.
- Mardia, K., J. Kent, and J. Bibby, 1979: *Multivariate Analysis*. Academic Press, 521 pp.
- Mokhov, I., and V. Semenov, 1997: Bimodality of the probability density functions of subseasonal variations in surface air temperature. *Izv. Atmos. Ocean. Phys.*, **33**, 702–708.
- —, V. Petukhov, and V. Semenov, 1998: Multiple intraseasonal temperature regimes and their evolution in the IAP RAS climate model. *Izv. Atmos. Ocean. Phys.*, 34, 145–152.
- Nicolis, C., 1982: Stochastic aspects of climatic transitions— Response to a periodic forcing. *Tellus*, **34**, 1–9.
- Paillard, D., 1998: The timing of Pleistocene glaciations from a

simple multiple-state climate model. *Nature*, **391**, 378–381, doi:10.1038/34891.

- Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. J. Climate, 12, 575–591.
- Rex, D. F., 1950a: Blocking action in the middle troposphere and its effects upon regional climate. I: An aerological study of blocking action. *Tellus*, 2, 196–211.
- —, 1950b: Blocking action in the middle troposphere and its effects upon regional climate. II: The climatology of blocking action. *Tellus*, 2, 275–301.
- Schütte, C., and W. Huisinga, 2003: Biomolecular conformations can be identified as metastable sets of molecular dynamics. *Handbook of Numerical Analysis*, Vol. X, P. G. Ciaret and J.-L. Lions, Eds., Elsevier, 699–744.

Simmons, A., and J. Gibson, 2000: The ERA-40 project plan.

ERA-40 Project Rep. Ser. 1, European Center for Medium-Range Weather Forecasting, 63 pp.

- Strang, G., and T. Nguyen, 1997: Wavelets and Filter Banks. Wellesley-Cambridge Press, 490 pp.
- Tsay, R., 2005: Analysis of Financial Time Series. 2nd ed. Wiley, 605 pp.
- Tsonis, A., and J. Elsner, 1990: Multiple attractors, fractal basins and longterm climate dynamics. *Beitr. Phys. Atmos.*, 63, 171– 176.
- Viterbi, A., 1967: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory*, **13**, 260–269.
- Wiedenmann, J. M., A. R. Lupo, I. I. Mokhov, and E. A. Tikhonova, 2002: The climatology of blocking anticyclones for the Northern and Southern Hemispheres: Block intensity as a diagnostic. J. Climate, 15, 3459–3473.