On clustering of non-stationary meteorological time series

Illia Horenko*
Institute of Mathematics
Free University of Berlin
Arnimallee 2-6, 14195 Berlin, Germany

Abstract

A method for clustering of multidimensional non-stationary meteorological time series is presented. The approach is based on optimization of the regularized averaged clustering functional describing the quality of data representation in terms of $\mathbf{K}$ regression models and a metastable hidden process switching between them. Proposed numerical clustering algorithm is based on application of the finite element method (FEM) to the problem of non-stationary time series analysis. The main advantage of the presented algorithm compared to HMM-based strategies and to finite mixture models is that no a priori assumptions about the probability model for hidden and observed processes are necessary for the proposed method. Another attractive numerical feature of the discussed algorithm is the possibility to choose the optimal number of metastable clusters and a natural opportunity to control the fuzziness of the resulting decomposition. The resulting FEM-K-Trends algorithm is compared with some standard fuzzy clustering methods on toy model examples and on analysis of multidimensional historical temperature data in Europe and a part of the North Atlantic.

Introduction

In the meteorology and climate research, recent years have seen a dramatic explosion in the amount and precision of raw data that is available in the form of time series. Due to the development of computational and measuring facilities in geo-sciences (e.g. reanalysis techniques in

*horenko@math.fu-berlin.de

meteorology) large amounts of measured and simulated information from all kinds of processes have been accumulated. Many of this processes are characterized by the presence of transitions between different local phases or regimes. Such phases can be found in meteorology (Tsonis and Elsner 1990; Kimoto and Ghil 1993a,b; Cheng and Wallace 1993; Efimov et al. 1995; Mokhov and Semenov 1997; Mokhov et al. 1998; Corti et al. 1999; Palmer 1999) and climatology (Benzi et al. 1982; Nicolis 1982; Paillard 1998).If knowledge about such systems is present only in the form of observation or measurement data, the challenging problem of identifying those persistent (or metastable) regimes together with the construction of reduced dynamical models of system dynamics becomes a problem of time series analysis and pattern recognition in high 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. The most popular methods for identification of multiple regimes in high-dimensional time series are: clustering methods (like K-means or fuzzy-c-means) (Höppner et al. 1999), methods based on hidden Markov models (HMMs) (Viterbi 1967; Bilmes 1998; Majda et al. 2006; Horenko et al. 2008b), finite mixture models (McLachlan and Peel 2000; Fruhwirth-Schnatter 2006), and neuronal networks (Monahan 2000).

All of the above methods share two basic problems: (i) number of clusters or phases present in the data is a priori unknown (Christiansen 2007), and (ii) each of the analysis methods implies some mathematical assumptions about the analyzed data. More specifically, most of the commonly used clustering methods imply the (local) stationarity of the analyzed data. This can lead to problems with identification of the optimal cluster partitioning in the case of the data with a time trend, i. e., it can happen by application of standard K-means and fuzzy-c-means algorithms to the analysis of historical temperature data. Presented paper aims at investigation of this problem, introduction of the methods of non-stationary data clustering in context of geophysical processes and comparison of different clustering approaches in context of historical temperature analysis.

A short overview of the most frequently used clustering methods is given, with a special emphasis on structural properties and implicit mathematical assumptions intrinsic for each of the methods. Fuzzy Clustering based on Regression Models (FCRM) algorithm for non-stationary data clustering is shortly explained (Hathaway and Bezdek 1993).The key part of the presented paper describes an extension of the standard K-means method in context of the finite element method (FEM)-based clustering methods to allow for analysis of non-stationary data. More specifically: we assume that the centers of the respective clusters evolve in time according to a linear combination of some predefined time-dependent basis functions with some (unknown) cluster-specific coefficients. Rewriting the problem in terms of the regularized averaged clustering functional allows us to apply the FEM-framework for simultaneous clustering of the data and identification of historical trends for each of the clusters. The main advantages of the presented method compared to the HMM-based methods are: (i) there is no need to assume

the Markovianity of the hidden process switching between the clusters, (ii) no explicit probabilistic model (like multivariate Gaussian in HMM-Gauss and HMM-PCA) for the observed data in the hidden states is needed, (iii) introduction of the regularization parameter allows to controls the metastability of the resulting cluster decomposition and helps to identify the number of persistent clusters.

We explain how the quality of the 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. The proposed framework is illustrated on some toy model systems and on analysis of historical 700 hPA geopotential height air temperature from the ERA 40 reanalysis data between 1958-2002.

1. Geometrical Clustering: K-Means, Fuzzy-c-Means and FCRM methods

a. *Cluster distance functional and K-Means clustering*

Let $x_t : [0,T] \to \Psi \subset \mathbf{R}^n$ be the observed n-dimensional time series. We look for $\mathbf{K}$ clusters characterized by $\mathbf{K}$ distinct sets of a priori unknown cluster parameters

$$\theta_1, \ldots, \theta_{\mathbf{K}} \in \Omega \subset \mathbf{R}^d, \tag{1}$$

(where d is the dimension of a cluster parameter space) for the description of the observed time series. Let

$$g\left(x_t, \theta_i\right) \quad : \quad \Psi \times \Omega \to [0, \infty)\,, \tag{2}$$

be a functional describing the distance from the observation x_t to the cluster i. For a given cluster distance functional (2), under data clustering we will understand the problem of a function $\Gamma(t) = (\gamma_1(t), \ldots, \gamma_{\mathbf{K}}(t))$ called the cluster affiliation (or the cluster weights) together with cluster parameters $\Theta = (\theta_1, \ldots, \theta_{\mathbf{K}})$ which minimize the averaged clustering functional

$$\mathbf{L}(\Theta, \Gamma) \quad = \quad \sum_{i=1}^{\mathbf{K}} \int_0^T \gamma_i(t) g\left(x_t, \theta_i\right) dt \to \min_{\Gamma(t), \Theta}, \tag{3}$$

subject to the constraints on $\Gamma(t)$:

$$\sum_{i=1}^{\mathbf{K}} \gamma_i(t) \quad = \quad 1, \quad \forall t \in [0,T] \tag{4}$$

$$\gamma_i(t) \quad \geq \quad 0, \quad \forall t \in [0,T]\,, \quad i = 1, \ldots, \mathbf{K}. \tag{5}$$

One of the most popular clustering methods in multivariate data-analysis is the so-called K-means algorithm (Bezdek 1981; Höppner et al. 1999). The affiliation to a certain cluster i is defined by the proximity of the observation $x_t \in \Psi$ to the cluster center $\theta_i \in \Psi$. In this case the cluster distance functional (2) takes the form of the square of the simple Euclidean distance between the points in n dimensions:

$$g\left(x_t, \theta_i\right) \;=\; \parallel x_t - \theta_i \parallel^2 . \tag{6}$$

If the analyzed data x_t is available only at some discrete observation times $t_j, j = 1, \ldots, n$, functional (3) gets the form

$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i(t_j) \parallel x_{t_j} - \theta_i \parallel^2 \;\rightarrow\; \min_{\Gamma(t),\Theta} . \tag{7}$$

K-means algorithm iteratively minimizes the functional (7) subject to constraints (4-5) assigning the new cluster affiliations $\gamma^{(l)}(t_j)$ and updating the cluster centers $\theta_i^{(l)}$ in iteration (l) according to the following formulas

$$\gamma_i^{(l)}(t_j) \;=\; \begin{cases} 1 & i = \arg\min \parallel x_{t_j} - \theta_i^{(l-1)} \parallel^2, \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

$$\theta_i^{(l)} \;=\; \frac{\sum_{j=1}^{n} \gamma_i^{(l)}(t_j) x_{t_j}}{\sum_{j=1}^{n} \gamma_i^{(l)}(t_j)}. \tag{9}$$

Iterations (8-9) are repeated until the change of the averaged clustering functional value does not exceed a certain predefined threshold value.

b. *Stationary data: Fuzzy c-Means Clustering*

As it can be seen from (8), the assignment of observed data point x_{t_j} to a certain cluster i is sharp, i. e., a single point can not be assigned simultaneously to different clusters. This can cause a problem in the case of geometrically overlapping clusters. To fix this problem, a following modification of the averaged clustering functional (7) was suggested (Bezdek 1981)

$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i^m(t_j) \parallel x_{t_j} - \theta_i \parallel^2 \;\rightarrow\; \min_{\Gamma(t),\Theta}, \tag{10}$$

where $m > 1$ is a fixed parameter called the fuzzyfier (Bezdek 1981; Bezdek et al. 1987). Analogously to the k-means, the fuzzy c-means algorithm is an iterative procedure for minimization of

(10)

$$\gamma_i^{(l)}(t_j) = \begin{cases} \frac{1}{\sum_{p=1}^{\mathbf{K}} \left(\frac{\| x_{t_j} - \theta_i^{(l-1)} \|^2}{\| x_{t_j} - \theta_p^{(l-1)} \|^2} \right)^{\frac{1}{m-1}}} & \text{if } \mathbf{I}_{x_{t_j}} \text{ is empty,} \\ \sum_{r \in \mathbf{I}_{x_{t_j}}} \gamma_r^{(l)}(t_j) = 1 & \text{if } \mathbf{I}_{x_{t_j}} \text{ is not empty}, i \in \mathbf{I}_{x_{t_j}}, \\ 0 & \text{if } \mathbf{I}_{x_{t_j}} \text{ is not empty}, i \notin \mathbf{I}_{x_{t_j}}, \end{cases} \quad (11)$$

$$\theta_i^{(l)} = \frac{\sum_{j=1}^{n} \gamma_i^{(l)}(t_j) x_{t_j}}{\sum_{j=1}^{n} \gamma_i^{(l)}(t_j)}. \quad (12)$$

where $\mathbf{I}_{x_{t_j}} = \{p \in \{1, \ldots, \mathbf{K}\} | \parallel x_{t_j} - \theta_p^{(l-1)} \parallel^2 = 0\}$ (Höppner et al. 1999). As it follows from (11), for any fixed fuzzifier m, cluster affiliations $\gamma_i^{(l)}(t_j)$ get values between 0 and 1, for $m \to \infty$ $\gamma_{i(t_j)}^{(l)} \to \frac{1}{\mathbf{K}}$. This feature allows clustering of overlapping data. However, the results are very much dependent on the choice of the fuzzifier m and there is no mathematically founded strategy of choosing this parameter dependent on the properties of the analyzed data. Moreover it is not a priori clear how many clusters are there in the data and which value should $\mathbf{K}$ take. Another problem is that the data is assumed being (locally) stationary, i.e., that the conditional expectation values θ_i calculated for the respective clusters i are assumed to be time independent. As we will see later, this can result in misinterpreting of the clustering results, if the data has a temporal trend.

c. *Non-stationary data: Fuzzy Clustering based on Regression Models (FCRM)*

To overcome the aforementioned stationarity restriction, R. Hathaway and J. Bezdek suggested the fuzzy c-regression models (FCRM) (also known in the literature as switching regression models) (Hathaway and Bezdek 1993). They suggested to describe each cluster-specific temporal trend as a certain (linear) regression model of a certain fixed order $\mathcal{R}$ given by some predefined basis functions $\phi_k(t), k = 0, \ldots, \mathcal{R}$ (e. g, time monomials t^k) and some a priori unknown regression coefficients θ_{ik} (lower index i denotes the number of the respective cluster). FCRM clustering algorithm yields simultaneous estimates of the regression parameters $\theta_{ik}, k = 0, \ldots, \mathbf{R}$ together with a fuzzy partitioning of the data based on the minimization of the modified form of the averaged clustering functional (10)

$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i^m(t_j) \parallel x_{t_j} - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \phi_k(t_j) \parallel^2 \to \min_{\Gamma(t), c}. \quad (13)$$

Comparison of (13) and (10) makes clear that the time-independent cluster centers θ_i in context of fuzzy c-means clustering are replaced by time-dependent functions

$$\theta_i(t) = \sum_{k=0}^{\mathcal{R}} \theta_{ik}\phi_k(t_j), \tag{14}$$

i. e., the cluster centers are assumed to be moving and the overall dynamics not assumed to be stationary. The overall algorithmic procedure can be efficiently implemented in context of Expectation-Maximization (EM) algorithms, if certain statistical assumptions about the underlying observation probability distribution can be made (Preminger et al. 2007). Analogously to fuzzy c-means clustering algorithm, the FCRM-algorithm is an iterative procedure with the same re-estimation formula for cluster weights (11) (except that $\theta_i^{(l)}$ has a form of (14)). However, it is not always clear whether the probabilistic assumptions (like Gaussianity of the regression residuals or their statistical independence (Preminger et al. 2007)) are fulfilled for the analyzed data. Moreover, similar to the fuzzy c-means algorithm, there is no practical and universal recipe for choosing the number of clusters $\mathbf{K}$ and fuzzifier m.

2. Regularized Averaged Clustering Functional: FEM-K-Trends algorithm

As it was emphasized above, the arbitrariness of parameter choice (especially for the number of clusters $\mathbf{K}$ and fuzzifier m) can make the application of the described clustering methods more problematic, especially in the case of the strongly overlapping data clusters. In the following, an extension of the recently proposed algorithm based on application of finite elements method (FEM) towards non-stationary data will be presented (Horenko 2008a). Dynamical approach to control the cluster-fuzzyness and the number of clusters will be introduced.

a. *Regularized Averaged Clustering Functional for Non-stationary Data*

Let us consider the clustering of non-stationary multidimensional data $x_t \in \mathbf{R}^d$ as a minimization problem (3) subject to constraints (4-5). The corresponding cluster distance functional (2) has the regression form as in the case above

$$g(x_t, \theta_i) = \| x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k(t) \|^2, \tag{15}$$

where $\theta_{ik} \in \mathbf{R}^d$ is a vector of regression coefficients, $\phi_k(t) \in \mathbf{R}^d$ is a vector of time-dependent regression functions and $\odot$ denotes a component-by-component product of 2 vectors. As it was

demonstrated in (Horenko 2008a), instead of the introduction of an artificial fuzzifier-parameter (as in case of c-means clustering and FCRM) and direct time discretization of (3), one incorporates some additional information into the optimization. One of the possibilities is to impose some smoothness assumptions in space of functions $\Gamma(\cdot)$) and then apply a finite Galerkin time-discretization of this infinite-dimensional Hilbert space. For example, one can impose the weak differentiability of functions γ_i, i. e.:

$$|\gamma_i|_{\mathcal{H}^1(0,T)} = \| \partial_t \gamma_i(\cdot) \|_{\mathcal{L}_2(0,T)} = \int_0^T (\partial_t \gamma_i(t))^2 dt \leq C_\epsilon^i < +\infty, \quad i = 1, \ldots, \mathbf{K}. \tag{16}$$

For a given observation time series, the above constraint limits the total number of transitions between the clusters and is connected to the metastability of the hidden process $\Gamma(t)$ (Horenko 2008a).

b. *Finite Element Approach: FEM-K-Trends algorithm*

To derive the algorithmic procedure for minimization of (3) subject to constraints (4), (5) and (16), one of the possibilities is to apply the Lagrange-formalism and incorporate the constraint (16) directly into the minimized functional with the help of the Lagrange-multiplier ϵ^2

$$\mathbf{L}^\epsilon(\Theta, \Gamma, \epsilon^2) = \mathbf{L}(\Theta, \Gamma) + \epsilon^2 \sum_{i=1}^{\mathbf{K}} \int_0^T (\partial_t \gamma_i(t))^2 dt \rightarrow \min_{\Gamma, \Theta}. \tag{17}$$

Let $\{0 = t_1, t_2, \ldots, t_{N-1}, t_N = T\}$ be a finite subdivision of the time interval $[0, T]$ with uniform time interval Δ_t. We can define a set of continuous functions $\{v_1(t), v_2(t), \ldots, v_N(t)\}$ called hat functions or linear finite elements (Braess 2007)

$$v_k(t) = \begin{cases} \frac{t - t_k}{\Delta_t} & 2 \leq k \leq N-1, t \in [t_{k-1}, t_k], \\ \frac{t_{k+1} - t}{\Delta_t} & 2 \leq k \leq N-1, t \in [t_k, t_{k+1}], \\ \frac{t_2 - t}{\Delta_t} & k = 1, t \in [t_1, t_2] \\ \frac{t - t_{N-1}}{\Delta_t} & k = N, t \in [t_{N-1}, t_N]. \end{cases} \tag{18}$$

Assuming that $\gamma_i \in \mathcal{H}^1(0, T)$ we can write

$$\begin{aligned} \gamma_i &= \tilde{\gamma}_i + \delta_N \\ &= \sum_{k=1}^{N} \tilde{\gamma}_{ik} v_k + \delta_N, \end{aligned} \tag{19}$$

where $\tilde{\gamma}_{ik} = \int_0^T \gamma_i(t)\, v_k(t)\, dt$ and δ_N is some discretization error. Inserting (19) into functional (17) and constraints (4,5) we get

$$\tilde{\mathbf{L}}^{\epsilon} = \sum_{i=1}^{\mathbf{K}} \left[a(\theta_i)^{\mathbf{T}} \tilde{\gamma}_i + \epsilon^2 \tilde{\gamma}_i^{\mathbf{T}} \mathbf{H} \tilde{\gamma}_i \right] \rightarrow \min_{\tilde{\gamma}_i, \Theta}, \tag{20}$$

$$\sum_{i=1}^{\mathbf{K}} \tilde{\gamma}_{ik} = 1, \quad \forall k = 1, \ldots, N, \tag{21}$$

$$\tilde{\gamma}_{ik} \geq 0, \quad \forall k = 1, \ldots, N, \quad i = 1, \ldots, \mathbf{K}, \tag{22}$$

where $\tilde{\gamma}_i = (\tilde{\gamma}_{11}, \ldots, \tilde{\gamma}_{iN})$ is the vector of discretized affiliations to cluster i,

$$a(\theta_i) = \left(\int_{t_1}^{t_2} v_1(t) g(x_t, \theta_i) dt, \ldots, \int_{t_{N-1}}^{t_N} v_N(t) g(x_t, \theta_i) dt \right), \tag{23}$$

is a vector of discretized model distances and $\mathbf{H}$ is the symmetric tridiagonal stiffness-matrix of the linear finite element set with $2/\Delta_t$ on the main diagonal, $-1/\Delta_t$ on both secondary diagonals and zero elsewhere. The only difference to the derivation presented in (Horenko 2008a) is the time-dependence of the cluster distance functional (15).

If $\epsilon^2 = 0$, then the above minimization problem (20-22), can be solved analytically wrt. $\tilde{\gamma}_i^{(l)}$ for a fixed set of cluster model parameters $\Theta^{(l)}$ (where l again denotes the index of current iteration) resulting in

$$\gamma_i^{(l)}(t_j) = \begin{cases} 1 & i = \arg\min \int_{t_j}^{t_{j+1}} v_j(s) \parallel x_s - \sum_{k=0}^{\mathcal{R}} \theta_{ik}^{(l)} \odot \phi_k(s) \parallel^2 ds, \\ 0 & \text{otherwise}, \end{cases} \tag{24}$$

If $\epsilon^2 > 0$, for a fixed set of cluster model parameters $\Theta^{(l)}$ the minimization problem (20-22), reduces to a sparse quadratic optimization problem with linear constraints which can be solved by standard tools of sparse quadratic programming (sQP) with computational cost scaling as $\mathcal{O}(N \log(N))$ (Gill et al. 1987; Arioli 2000). Therefore, from a computational point of view, the presented approach is more expensive (for $\epsilon^2 > 0$) then the traditional fuzzy c-means and FCRM algorithms (which both scale as $\mathcal{O}(N)$). However, as will be demonstrated by numerical examples, this drawback is compensated by nice properties of the presented method wrt. the choice of $\mathbf{K}$ and good performance in analysis of strongly overlapping data-clusters.

In addition, the minimization problem (20 -22) wrt. the parameters Θ for a fixed set of discretized cluster affiliations $\tilde{\gamma}_i$ is equivalent to the unconstrained minimization problem

$$\sum_{i=1}^{\mathbf{K}} a(\theta_i)^{\mathbf{T}} \tilde{\gamma}_i^{(l)} \rightarrow \min_{\Theta}. \tag{25}$$

Since $g\left(x_t, \theta_i\right)$ has a form of (15), this is a linear regression problem and can be solved explicitly using the least squares method.

Therefore, the clustering FEM-K-trends algorithm can be implemented as the following iterative numerical scheme:

FEM-K-Trends Algorithm.

Setting of optimization parameters and generation of initial values:

· Set the number of clusters $\mathbf{K}$, regularization factor ϵ^2, finite discretization of the time interval $[0, T]$, and the optimization tolerance TOL

· Set the iteration counter $l = 1$

· Choose random initial $\tilde{\gamma}_i^{(1)}, i = 1, \ldots, \mathbf{K}$ satisfying (21-22)

· Calculate $\Theta^{(1)} = \arg\min_{\Theta} \tilde{\mathbf{L}}^{\epsilon}\left(\Theta, \tilde{\gamma}_i^{(1)}\right)$ solving the linear regression problem (25)

Optimization loop:

do

· Compute $\tilde{\gamma}^{(l+1)} = \arg\min_{\tilde{\gamma}} \tilde{\mathbf{L}}^{\epsilon}\left(\Theta^{(l)}, \tilde{\gamma}\right)$ satisfying (21-22) applying QP (if $\epsilon^2 > 0$) or applying (24) (if $\epsilon^2 = 0$)

· Calculate $\Theta^{(l+1)} = \arg\min_{\Theta} \tilde{\mathbf{L}}^{\epsilon}\left(\Theta, \tilde{\gamma}_i^{(l+1)}\right)$ solving the linear regression problem (25)

· $l := l + 1$

while $\left|\tilde{\mathbf{L}}^{\epsilon}\left(\Theta^{(l)}, \tilde{\gamma}_i^{(l)}\right) - \tilde{\mathbf{L}}^{\epsilon}\left(\Theta^{(l-1)}, \tilde{\gamma}_i^{(l-1)}\right)\right| \geq$ TOL.

Major advantage of the presented algorithm compared to HMM-based strategies (Horenko et al. 2008b; Horenko 2008b; Horenko et al. 2008a) and to finite mixture models (McLachlan and Peel 2000; Fruhwirth-Schnatter 2006) is that no a priori assumptions about the probability model for hidden and observed processes are necessary in the context of the FEM-K-Trends algorithm.

3. Postprocessing of results

The quality of the clustering is very much dependent on the original data, especially on the length of the available time series. The shorter the observation sequence is, the bigger the uncertainty of the resulting estimates. The same is true, if the number K of the hidden states is increasing for the fixed length of the observed time series: the bigger K is, the higher will be the uncertainty for each of the resulting clusters. Therefore, in order to be able to statistically distinguish between different hidden states, we need to get some notion of the FEM-K-trends robustness. This can be achieved through the postprocessing of the clustering results and analysis of the

transition process and regression models estimated for the clusters. If there exist two states with overlapping confidence intervals for each of the respective model parameters, then those are statistically indistinguishable, $\mathbf{K}$ should be reduced and the optimization repeated. In other words, confidence intervals implicitly give a natural upper bound for the number of possible clusters. However, in many atmospherical applications, the question about the optimal number of clusters is highly non-trivial and is very difficult to answer without incorporation of some additional information (Christiansen 2007).

As was demonstrated in (Horenko 2008a), there is a connection between the regularization factor ϵ^2 and metastability of the resulting data decomposition. As it will be shown later in numerical examples, for fixed $\mathbf{K}$ the number of transitions between the identified clusters will decrease with growing ϵ^2. This means that respective mean exit times for the identified clusters get longer and the corresponding cluster decompositions become more and more metastable. Careful inspection of the transition process $\Gamma(t)$ identified for different values of ϵ^2 can help to find out the optimal number $\mathbf{K}$ of metastable cluster states.

Another possibility to estimate the optimal number of clusters can be used, if the identified transition process $\Gamma(t)$ is shown to be Markovian for given $\mathbf{K}, \epsilon^2$. Markovianity can be verified applying some standard tests, e. g., one can check the generator structure of the hidden process, see Metzner et al. (2007). In such a case the hidden transition matrix can be calculated and its spectrum can be examined for a presence of the spectral gap. If the spectral gap is present, then the number of the dominant eigenvalues (i. e., eigenvalues between the spectral gap and 1.0) gives the number of the metastable clusters in the system (Schütte and Huisinga 2003; Huisinga et al. 2004).

Positive verification of the hidden process' Markovianity has an additional advantage: it allows to construct a reduced dynamical model of the analyzed process and to estimate some dynamical characteristics of the analyzed process, e. g., one can calculate relative statistical weights, mean exit times and mean first passage times for the identified clusters (Gardiner 2004; Horenko et al. 2008a). Reduced Markovian description can also be helpful in construction of the operative weather predictions based on historical observation data.

Analysing the resulting regression coefficients for the identified clusters can help to reveal the temporal trends and the degree of non-stationarity of the analyzed data. Moreover, standard tools of regression analysis can be used to estimate the statistical significance of the identified trends, to calculate the confidence intervals of the identified parameters and to define the optimal regression order parameter $\mathcal{R}$ for each of the clusters (Kedem and Fokianos 2002).

4. Illustrative model examples

In the following we will illustrate the proposed strategy for clustering of non-stationary data with time trend and identification of metastable states on three examples: (a.) a model system build of two three-dimensional linear regressions and a predefined metastable process switching between them, (b.) a model system build of three three-dimensional linear regressions and a fixed transition process with two rapidly mixing states and two metastable states (c.) a set of historical averaged daily temperatures between 1958 and 2002 on a 31×18 spatial grid covering Europe and part of the north Atlantic.

Example (a.) represents a toy model aiming to illustrate the proposed framework on a simple and understandable system. The effects induced by the regularization parameter are explained and a comparison with the standard FCRM-algorithm for analysis of non-stationary data is performed.

In the next example (b.) we demonstrate two approaches to identifying the optimal number $\mathbf{K}$ of metastable clusters. In contrast to other methods based on HMMs and finite mixture models we are aware of (McLachlan and Peel 2000; Horenko 2008b; Horenko et al. 2008a), the proposed method allows assumption free identification of the hidden states for a given observation data.

Finally, in example (c.) the application of the FEM-K-trends–algorithm is demonstrated on clustering of historical temperature data. Markovianity of the identified transition process is verified and 3 metastable temperature clusters are identified. The identified Markovian transition process is investigated wrt. the inhomogeneity and a long-term variability of the transition matrix coefficients. Resulting regressions are compared with the trends calculated from standard stationary k-means clustering and the discrepancies are discussed.

a. *Two hidden states*

As the first application example for the proposed framework we consider a time series $x(t) \in \mathbf{R}^3$ generated as an output of two switching linear regressive models with Gaussian noise:

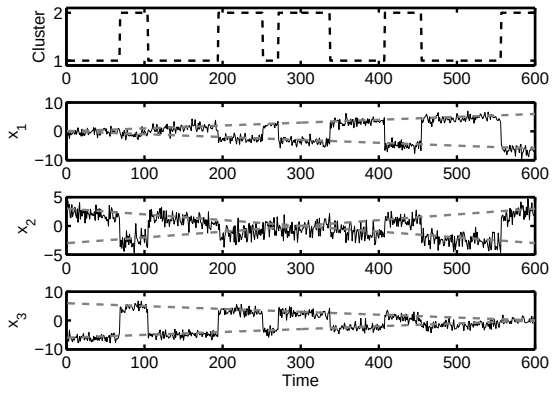

Figure 1: Upper panel: metastable transition process switching between two linear regressions (26). The other 3 panels demonstrate resulting three-dimensional time series for $\sigma = 1.0$ (solid). Dashed lines indicate the linear trends characteristic for both clusters in respective dimensions.

$$\begin{aligned} x_j(t) &= \theta_{i(t)j}(t - \bar{t}_j) + \sigma \mathbf{N}(0,1), \quad i = 1,2, \quad j = 1,2,3 \\ \theta_1 &= \begin{pmatrix} 0.01 & -0.01 & 0.01 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} -0.01 & 0.01 & -0.01 \end{pmatrix} \\ \bar{t} &= \begin{pmatrix} 0 & 300 & 600 \end{pmatrix} \end{aligned} \tag{26}$$

In the following numerical studies we will use the fixed transition process $i(t)$ that is shown in the upper panel of Fig. 1. The other panels of the Fig. 1 demonstrate a three-dimensional time series with 600 elements generated by the model (26) for the chosen $i(t)$ and noise intensity $\sigma = 1.0$.

The left panel of Fig. 2 shows the influence of the fuzzifier m on results of the FCRM-clustering. It demonstrates that the choice of the parameter has no significant impact on the clustering quality, it rather gets worse for increasing m and the identified clusters getting "blurred". The right panel of Fig. 2 illustrates the influence of a regularization factor ϵ^2 on assignment of data to respective clusters for FEM-K-trends algorithm. In contrast to FCRM-clustering, the regularization factor has a strong influence on the FEM-K-trends-clustering results. Increasing ϵ^2 results in a coarse graining of the identified affiliation functions, i. e., only "long living" structures in γ "survive" with increasing ϵ^2. It means that the regularization factor ϵ^2 has a direct connection to a dynamical behavior of the analyzed time series, i. e., it allows to control the metastability of underlying transition process.

Next, we compare the FEM-K-trends-method with FCRM-clustering algorithm wrt. the sensitivity to noise σ. Fig. 3 reveals that application of the FEM-K-trends-methods results in much

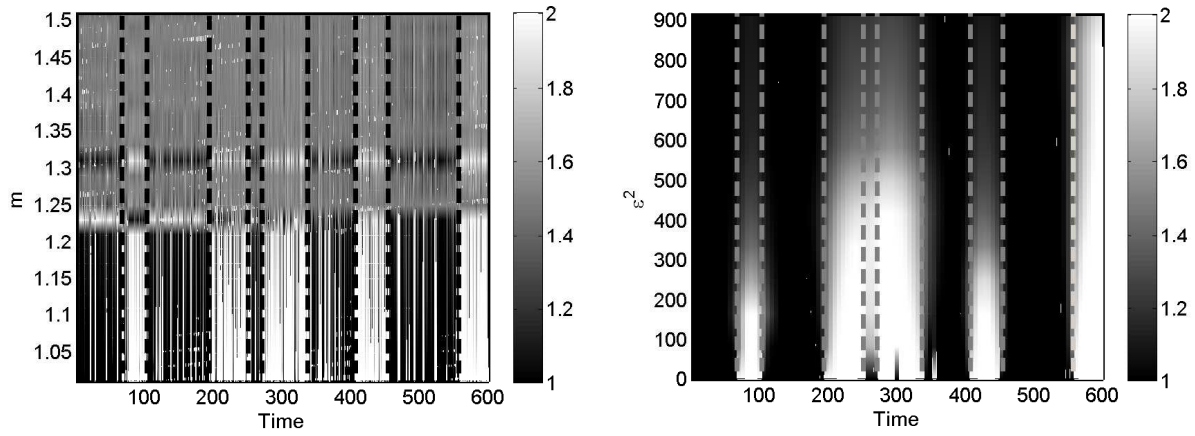

Figure 2: Identified transition path between the clusters 1 and 2 as a function of time (color denotes the affiliation to a corresponding cluster): (left panel) calculated for different values of fuzzifier m with FCRM-algorithm ($\mathbf{K} = 2$, optimization repeated 100 times with randomly generated initial values), and (right panel) for different values of regularization factor ϵ^2 (FEM-K-Trends-algorithm with $\mathbf{K} = 2$, optimization repeated 100 times with randomly generated initial values). The analyzed time series is in both cases the same, generated with model (26) with transition process from the upper panel of Fig. 1 and noise amplitude $\sigma = 7$. Dashed lines denote the moments when the original transition process from Fig. 1 was switching between the clusters.

more reliable cluster identification in the case of a noisy data. Fig. 3 also demonstrates that the introduction of the fuzzifier $m > 1$ in context of FCRM-method results in the worsening of cluster identification for well–separated clusters with relatively low noise intensity.

b. *Three hidden states*

In order to demonstrate the performance of the presented framework wrt. the identification of metastable cluster sets, we extend the previous example by adding a new linear regression cluster state and change the transition process in a way presented in Fig. 4. The hidden process switches frequently between the first and the second states and from time to time goes into the third state, i. e., the third state is metastable, as well as the combination of the first and the

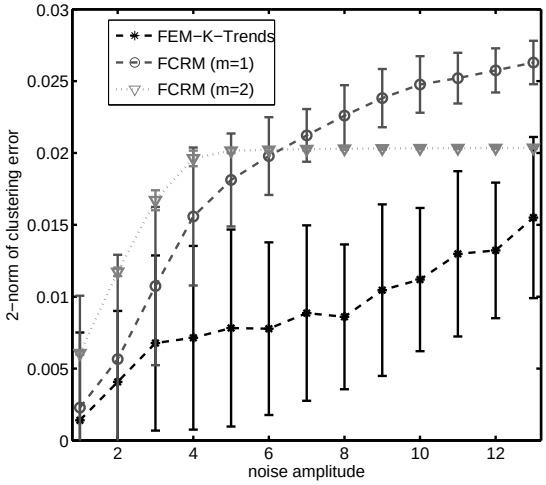

Figure 3: Comparison of the mean cluster assignment errors computed for 100 independent trajectories of model (26) with transition process from Fig. 1 for different values of the noise amplitude σ with the help of: FCRM-algorithm for $m = 1$ ($\mathbf{K} = 2$, circles), FCRM-algorithm for $m = 2$ ($\mathbf{K} = 2$, triangles) and FEM-K-trends-algorithm for $\epsilon^2 = 200$ ($\mathbf{K} = 2$, crosses). Error bars indicate the confidence intervals for estimated mean errors.

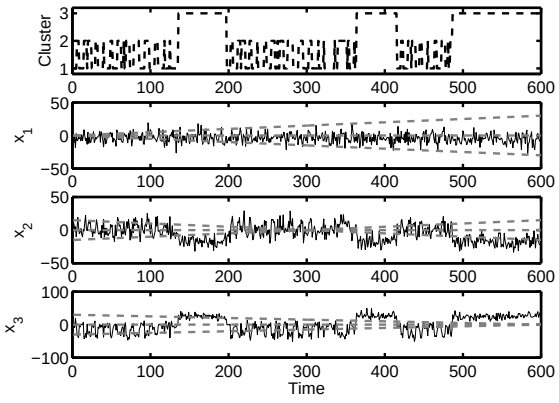

Figure 4: Upper panel: transition process with two metastable substates and two rapidly mixing states switching between three linear regressions (27). Other 3 panels demonstrate resulting three-dimensional time series for $\sigma = 1.0$ (solid). Dashed lines indicate the linear trends characteristic for the clusters in respective dimensions.

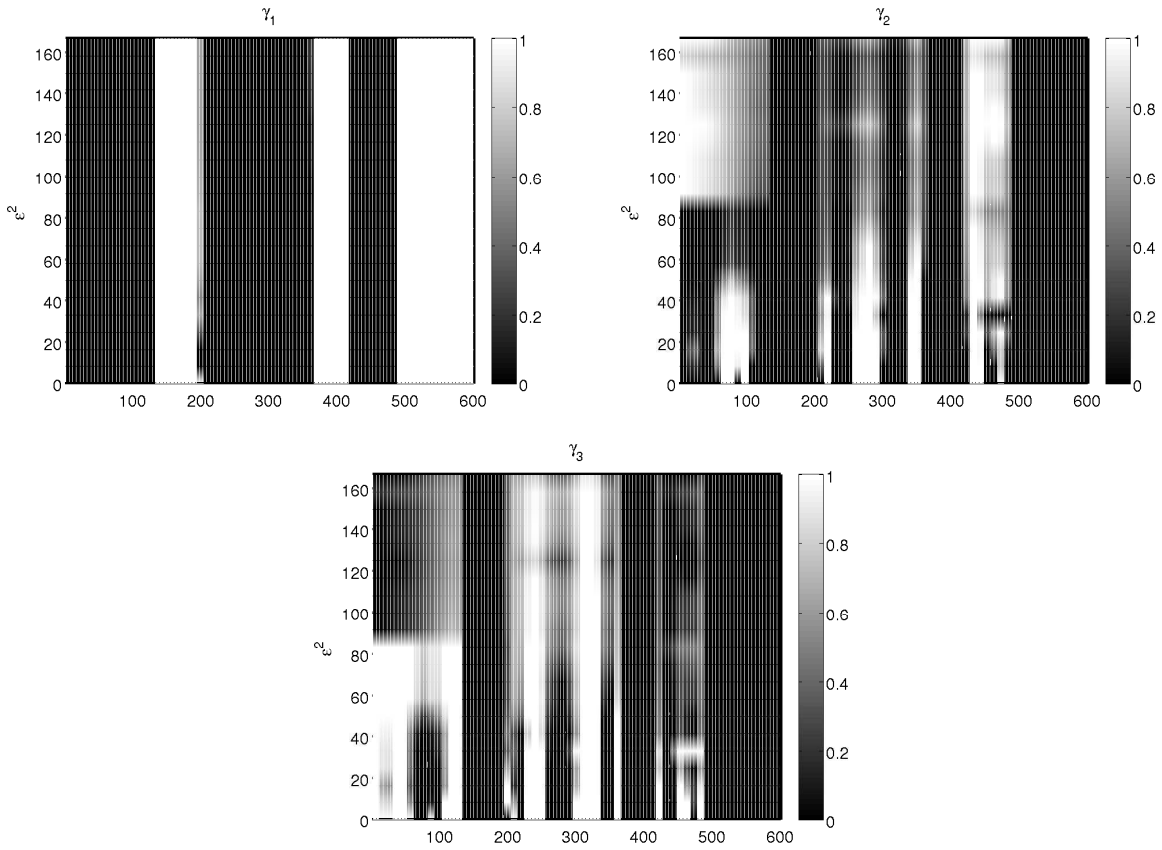

Figure 5: Cluster affiliation functions $\gamma_1(t)$, $\gamma_2(t)$ and $\gamma_3(t)$ (color indicates the value of the function between 0 and 1). The calculation is performed with the FEM-K-Trends-algorithm for different values of regularization factor ϵ^2 (with $\mathbf{K} = 3$, optimization repeated 100 times with randomly generated initial values). The analyzed time series is generated with model (27) with the transition process from the upper panel of Fig. 1 and noise amplitude $\sigma = 7$.

second states together builds the second metastable cluster set.

$$\begin{aligned} x_j(t) &= \theta_{i(t)j}(t - \bar{t}_j) + \sigma \mathbf{N}(0,1), \quad j = 1,2,3 \\ \theta_1 &= \begin{pmatrix} 0.0 & 0.0 & 0.0 \end{pmatrix} \\ \theta_2 &= \begin{pmatrix} 0.01 & -0.01 & 0.01 \end{pmatrix}, \quad \theta_3 = \begin{pmatrix} -0.01 & 0.01 & -0.01 \end{pmatrix} \\ \bar{t} &= \begin{pmatrix} 0.0 & 300 & 600 \end{pmatrix} \end{aligned} \tag{27}$$

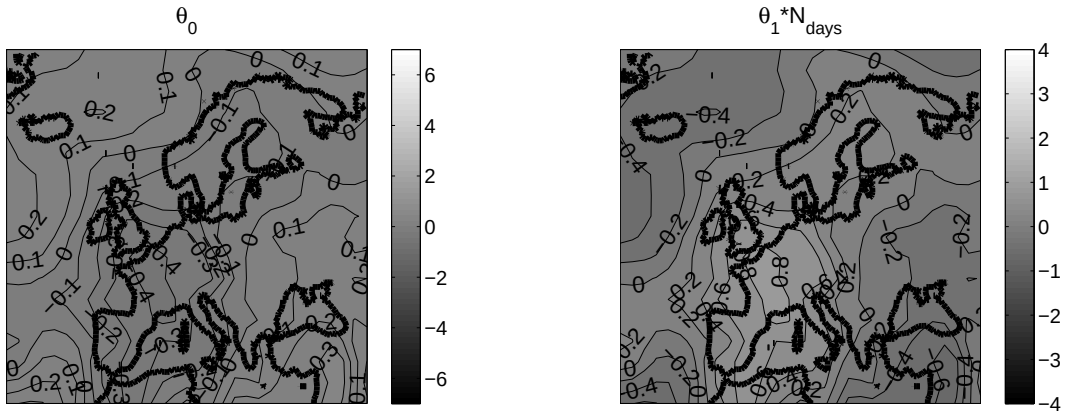

Figure 6: Results of the optimal linear regression fit $(\theta_0 + \theta_1 t)$ for the whole length of the analyzed temperature time series.

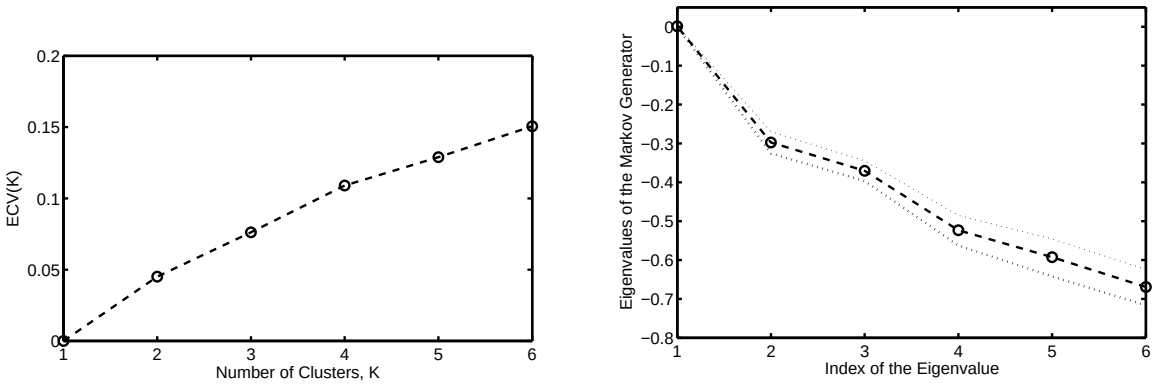

Figure 7: Comparison of two criteria for the choice of $\mathbf{K}$: Explained Cluster Variance (ECV) criterion (28) (left panel), and Markovian spectral gap criterion (right panel, dotted lines indicate the confidence intervals of the calculated Markov generator eigenvalues).

As it was already mentioned above, there are two basic possibilities to estimate the number of metastable sets in the analyzed data and thereby to choose the optimal $\mathbf{K}$: (i) spectral analysis of the Markov transition matrix and (ii) variation of the regularization parameter ϵ^2 and careful comparison of the respective cluster affiliations $\gamma_i(t), i = 1, \ldots, \mathbf{K}$. In the following, both approaches will be exemplified for the time series from Fig. 4.

(i) If the transition process resulting from an application of the clustering algorithm (with $\mathbf{K}$

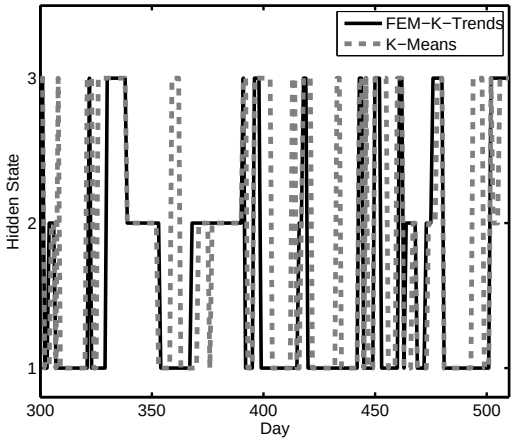

Figure 8: Comparison of transition pathes calculated with K-means algorithm for $\mathbf{K} = 3$ (dashed) and FEM-K-trends algorithm for $\mathbf{K} = 3, \mathcal{R} = 1, \epsilon^2 = 0$ (solid).

chosen a priori high enough) was found to be Markovian, one can investigate a spectrum of the correspondent transition matrix for a presence of the spectral gap that can help to identify the number of metastable Markovian sets in the data (Schütte and Huisinga 2003; Huisinga et al. 2004). Applying the FEM-K-trends algorithm with $\mathbf{K} = 4$ and $\epsilon^2 = 0$ we get the hidden transition process that can be shown to be Markovian (Metzner et al. 2007). Calculating a spectrum of the correspondent transition matrix we get the following eigenvalues $(1.0, 0.99, 0.57, 0.49)$. The spectral gap indicates the presence of two essential eigenvalues, 1.0 and 0.99, therefore the existence of $\mathbf{K} = 2$ metastable sets is shown.

(ii) Alternatively, if the Markovianity of the transition process is not fulfilled, one can choose some a priori value for $\mathbf{K}$ and repeat the FEM-K-trends clustering with increasing values of ϵ^2. For the time series from Fig. 4, respective results are summarized in Fig. 5 for $\mathbf{K} = 3$. Whereas the cluster affiliation $\gamma_1(t)$ indicates a sharp separation of two states (only taking values near 0 and 1 almost independently of the regularization factor ϵ^2), cluster affiliations $\gamma_2(t)$ and $\gamma_3(t)$ approach the value 0.5 very fast with growing ϵ^2. This means that both states become statistically indistinguishable and the number of cluster states $\mathbf{K}$ should be decreased. Analysis of the same data with $\mathbf{K} = 2$ results in a sharp separation of two metastable states (almost independently of ϵ^2). As in the case (i) before, this feature indicates $\mathbf{K} = 2$ as an optimal number of metastable sets in the data.

c. *Analysis of Historical Temperature Data in Europe*

Description of the data Using the method presented in the previous sections, we analyze daily mean values of the 700 hPa air temperature field from the ERA 40 reanalysis data (Simmons and Gibson 2000). We consider a region with the coordinates: 27.5°W – 47.5°E and 32.5°N – 75.0° N , which includes Europe and a part of the Eastern North Atlantic. The resolution of the data is 2.5°which implies a grid with 31 points in the zonal and 18 in the meridional direction. For the analysis we have considered temperature values only for the period 1958 till 2002, thus we end with a 558-dimensional time series of 16314 days.

In order 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 01.01.1959 we subtract the mean value over all days which are first of January and so on.

Discussion of the results We start the data analysis calculating the optimal linear regression fit ($\mathcal{R} = 1$) for the whole length of the analyzed time series. The correspondent expected mean temperature change during the whole observation period $N_{days} = 16314$ can then be calculated as $\theta_1 N_{days}$, where $\theta_1 \in \mathbf{R}^{558}$ is a vector of the first-order part coefficients for the analyzed data on the respective grid. As it can be seen from the right panel of the Fig. 6, mean overall temperature changes do not exceed 1.0°C.

Next, we cluster the data with FEM-K-trends(for $\mathbf{K} = 6, \mathcal{R} = 1, \epsilon^2 = 0$). In order to avoid the problem of trapping in local optima of the functional (20), we repeat the clustering procedure 100 times with different randomly initialized cluster parameters and keep the clustering results with the lowest value of the functional (20). Fig. 7 illustrates the comparison of two alternative criteria used to determine the number of clusters $\mathbf{K}$ in the data. First we apply the Explained Cluster Variance (ECV) criterion, defined as

$$\begin{aligned} \mathrm{ECV}\,(\mathbf{K}) &= 1 - \frac{\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_i^m(t_j) \parallel x_{t_j} - \mu_i \parallel^2}{\sum_{j=1}^{n} \parallel x_{t_j} - \mathbf{E}\,(x_t) \parallel^2}, \\ \mathbf{E}\,(x_t) &= \frac{1}{n} \sum_{j=1}^{n} x_{t_j}, \end{aligned} \tag{28}$$

where μ_i are the geometrical cluster centers. As it is demonstrated on the left panel of Fig. 7, the value of ECV $(\mathbf{K})$ increases uniformly with $\mathbf{K}$ and implicates no obvious choice of $\mathbf{K}$. The right panel of Fig. 7 shows the eigenvalues of the Markov-generator estimated from the transition process resulting from FEM-K-trends(for $\mathbf{K} = 6, \mathcal{R} = 1, \epsilon^2 = 0$). The presence of the spectral gap indicates existence of 3 metastable sets in the analyzed data.

Applying of the FEM-K-trends procedure (for $\mathbf{K} = 6, \mathcal{R} = 1, \epsilon^2 = 0$, clustering repeated 100 times with different randomly chosen initial cluster parameter values) results in the identification

of the transition process shown in Fig. 8 (solid line). Besides of some similarity, the identified path is quite different from the transition calculated with K-means-algorithm. This difference becomes more obvious if we compare the trends calculated for the K-means clusters as it was done in Philipp et al. (2007) (see Fig. 9) and the values resulting from the FEM-K-trends clustering (see Fig. 10). Standard methods of regression analysis were applied (Kedem and Fokianos 2002), statistical significance of the resulting linear trends was confirmed and confidence intervals for the regression coefficients were calculated demonstrating the credibility of the identified temperature trends.

As was already discussed above, K-means-framework uses an implicit assumption that the analyzed time series is (locally) stationary. In contrast, FEM-K-trends actually uses the non-stationarity of the data as an additional property which helps to cluster the data according to the differences in time trend. Figures 9-10 demonstrate how big the discrepancy between the clustering results obtained with different methods can be and how important it is to choose the right tool dependent on the analyzed data and the property of interest.

Finally, we analyze the identified transition process. As it was mentioned above, the major advantage of the presented FEM-K-trends approach compared to the HMM-based strategies (Horenko et al. 2008b; Horenko 2008b,b) is its independence on assumptions about the type of the probability model. Therefore one does not have to assume a priori that the hidden transition process is an output of the time-homogenous Markov chain. In context of FEM-K-trends this assumption can be checked a posteriori and can help to construct reduced predictive Markovian models based on the observation data. As it is shown in the Fig.11, the eigenvalues of the underlying generator can be assumed to be time-independent and the process switching between 3 regression models can be assumed to be Markovian. To investigate the time-dependence of the identified Markovian process switching between the linear regressive models from Fig. 10, we define a moving window of 1000 days, slide it along the time series of transition process $\Gamma(t)$ and calculate the transition matrix $P(t)$ of the underlying Markov chain for all t. The resulting transition probabilities as functions of time are shown in Fig. 12. It demonstrates that the Markovian transition process is not time-homogenous and considerable amount of long-term variability is present in it.

5. Conclusion

We have presented a numerical framework for clustering multidimensional non-stationary time series based on minimization of a regularized averaged clustering functional. Finite element discretization of the problem allowed us to suggest a numerical algorithm based on the iterative minimization of this functional. We have compared the resulting FEM-K-trends algorithm with standard clustering techniques and analyzed the connection between the regularization factor,

metastabilty and identification of optimal number of metastable substates in the analyzed data.

When working with multidimensional data, it is very important to be able to extract some reduced description out of it (e.g., in form of hidden transition pathes or reduced dynamical models). In order to control the reliability of the clustering, one has to analyze the sensitivity of obtained results wrt. the length of the time series and the number K of the identified clusters. We have given some hints for the selection of an optimal K and explained how the quality of the resulting reduced representation can be acquired.

As an application of the proposed method to analysis of historical temperature data, it has been demonstrated how the problem of temperature trend identification can be solved simultaneously with the clustering problem. Large discrepancies between the temperature trends identified for K-means and FEM-K-trends clusters were found. It reveals how big the impact of implicit method assumptions about the data (like local data stationarity in the case of the widely used K-means algorithm) on the analysis results and their interpretation is.

Acknowledgments

The author thanks R. Klein (FU-Berlin) for a helpful discussion as well as H. Oesterle (PIK) and S. Dolaptchiev (PIK/FU) who provided the ERA 40 reanalysis data from the European Center for Medium-Range Weather Forecasting. The work was supported by the DFG SPP METSTROEM "Meteorology and Turbulence Mechanics".

References

Arioli, M., 2000: The use of QR factorization in sparse quadratic programming and backward error issues. *SIAM J. on Matrix Analysis and Applications*, **21(3)**, 825 – 839.

Benzi, R., G. Parisi, A. Sutera, and A. Vulpiani, 1982: Stochastic resonance in climatic change. *Tellus*, **3**, 10–16.

Bezdek, J., 1981: *Pattern recognition with fuzzy objective function algorithms.*. Plenum Press, New York.

Bezdek, J., R. Hathaway, M. Sabin, and W. Tucker, 1987: Convergence theory for fuzzy c-mens: counterexamples and repairs. *IEEE Trans. Systems.*, **17**, 873–877.

Bilmes, J., 1998: *A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Thechnical Report*. International Computer Science Institute, Berkeley.

Braess, D., 2007: *Finite Elements: Theory, Fast Solvers and Applications to Solid Mechanics*. Cambridge University Press.

Cheng, X. and J. M. Wallace, 1993: Cluster analysis of the northern hemisphere wintertime 500-hpa height field: Spatial patterns. *Journal of Atmospheric Sciences*, **50**, 2674–2696.

Christiansen, B., 2007: Atmospheric circulation regimes: Can cluster analysis provide the number? *J. Climate*, **20(10)**, 2229–2250.

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.

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. *Quarterly Journal of the Royal Meteorological Society*, **121**, 1651–1679.

Fruhwirth-Schnatter, S., 2006: *Finite Mixture and Markov Switching Models*. Springer.

Gardiner, H., 2004: *Handbook of stochastical methods*. Springer, Berlin.

Gill, P., W. Murray, M. Saunders, and M. Wright, 1987: A Schur-complement method for sparse quadratic programming. Technical report, STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB.

Hathaway, R. and J. Bezdek, 1993: Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems.*, **1**, 195–204.

Höppner, F., F. Klawonn, R. Kruse, and T. Runkler, 1999: *Fuzzy cluster analysis.*. John Wiley and Sons, New York.

Horenko, I., 2008a: Finite element approach to clustering of multidimensional time series. *submitted to SIAM Journal of Sci. Comp.*, (available via biocomputing.mi.fu-berlin.de).

— 2008b: On simultaneous data–based dimension reduction and hidden phase identification. *J. Atmos. Sci*, **to appear**, (available via biocomputing.mi.fu-berlin.de).

Horenko, I., S. Dolaptchiev, A. Eliseev, I. Mokhov, and R. Klein, 2008a: Metastable decomposition of high-dimensional meteorological data with gaps. *J. Atmos. Sci*, **to appear**, (available via biocomputing.mi.fu-berlin.de).

Horenko, I., R. Klein, S. Dolaptchiev, and C. Schuette, 2008b: Automated generation of reduced stochastic weather models i: simultaneous dimension and model reduction for time series analysis. *SIAM Mult. Mod. Sim.*, **6(4)**, 1125–1145.

Huisinga, W., S. Meyn, and C. Schuette, 2004: Phase transitions and metastability in Markovian and molecular systems. *Ann. Appl. Prob*, **14(1)**, 419–458.

Kedem, B. and K. Fokianos, 2002: *Regression models for time series analysis*. Wiley Series in Probability and Statistics.

Kimoto, M. and M. Ghil, 1993a: Multiple flow regimes in the northern hemisphere winter. part i: Methodology and hemispheric regimes. *Journal of Atmospheric Sciences*, **50**, 2625–2644.

— 1993b: Multiple flow regimes in the northern hemisphere winter. part ii: Sectorial regimes and preferred transitions. *Journal of Atmospheric Sciences*, **50**, 2645–2673.

Majda, A., C. Franzke, A. Fischer, and D. Crommelin, 2006: Distinct metastable atmospheric regimes despite nearly gaussian statistics : A paradigm model. *PNAS*, **103**, 8309–8314.

McLachlan, G. and D. Peel, 2000: *Finite mixture models*. Wiley, New–York.

Metzner, P., I. Horenko, and C. Schuette, 2007: Generator estimation of Markov Jump processes based on incomplete observations nonequidistant in time. *Phys. Rev. E*, **76**, 0667021.

Mokhov, I., V. Petukhov, and V. Semenov, 1998: Multiple intraseasonal temperature regimes and their evolution in the iap ras climate model. *Izvestiya, Atmos. Ocean. Phys.*, **34**, 145–152.

Mokhov, I. and V. Semenov, 1997: Bimodality of the probability density functions of subseasonal variations in surface air temperature. *Izvestiya, Atmos. Ocean. Phys.*, **33**, 702–708.

Monahan, A., 2000: Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system. *J. Climate*, **13**, 821–835.

Nicolis, C., 1982: Stochastic aspects of climatic transitions-response to a periodic forcing. *Tellus*, **34**, 1–+.

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. *Journal of Climate*, **12**, 575–591.

Philipp, A., P. Della-Marta, J. Jacobeit, D. Fereday, P. Jones, A. Moberg, and H. Wanner, 2007: Long term variability of daily north Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering. *J. Climate*, **20(16)**, 4065–4095.

Preminger, A., U. Ben-Zion, and D. Wettstein, 2007: The extended switching regression model: allowing for multiple latent state variables. *J. of Forecasting*, **26**, 457–473.

Schütte, C. and W. Huisinga, 2003: Biomolecular conformations can be identified as metastable sets of molecular dynamics. *Handbook of Numerical Analysis*, P. G. Ciaret and J.-L. Lions, eds., Elsevier, volume X, 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 Forcasting, Reading.

Tsonis, A. and J. Elsner, 1990: Multiple attractors, fractal basins and longterm climate dynamics. *Beit. Phys. Atmos.*, **63**, 171–176.

Viterbi, A., 1967: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Informat. Theory*, **13**, 260–269.

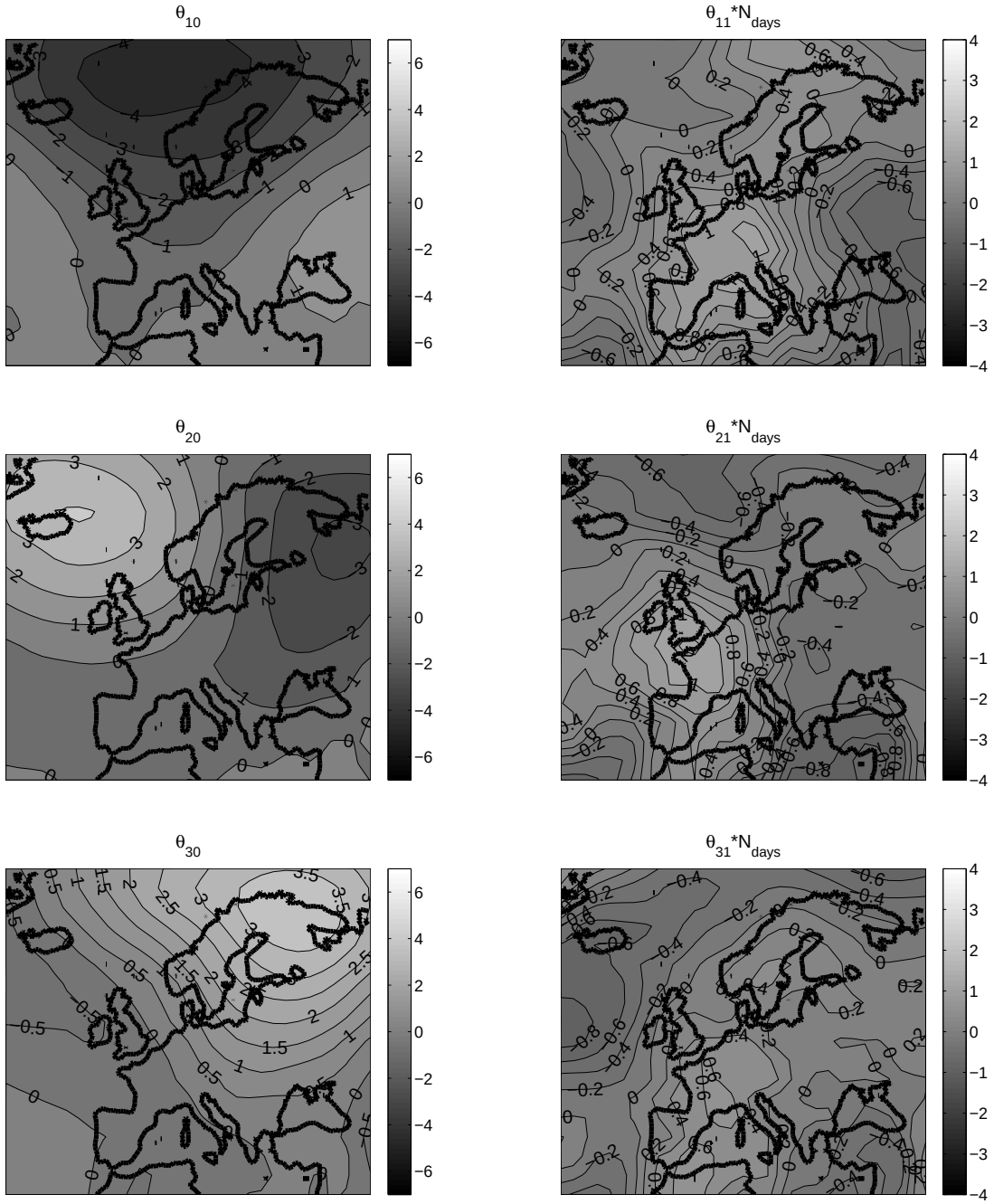

Figure 9: Results of the optimal linear regression fit $(\theta_{i0} + \theta_{i1}t)\,, i = 1, 2, 3$ (coloring in °C) calculated for each of the cluster states identified by the K-means algorithm for $\mathbf{K} = 3$.

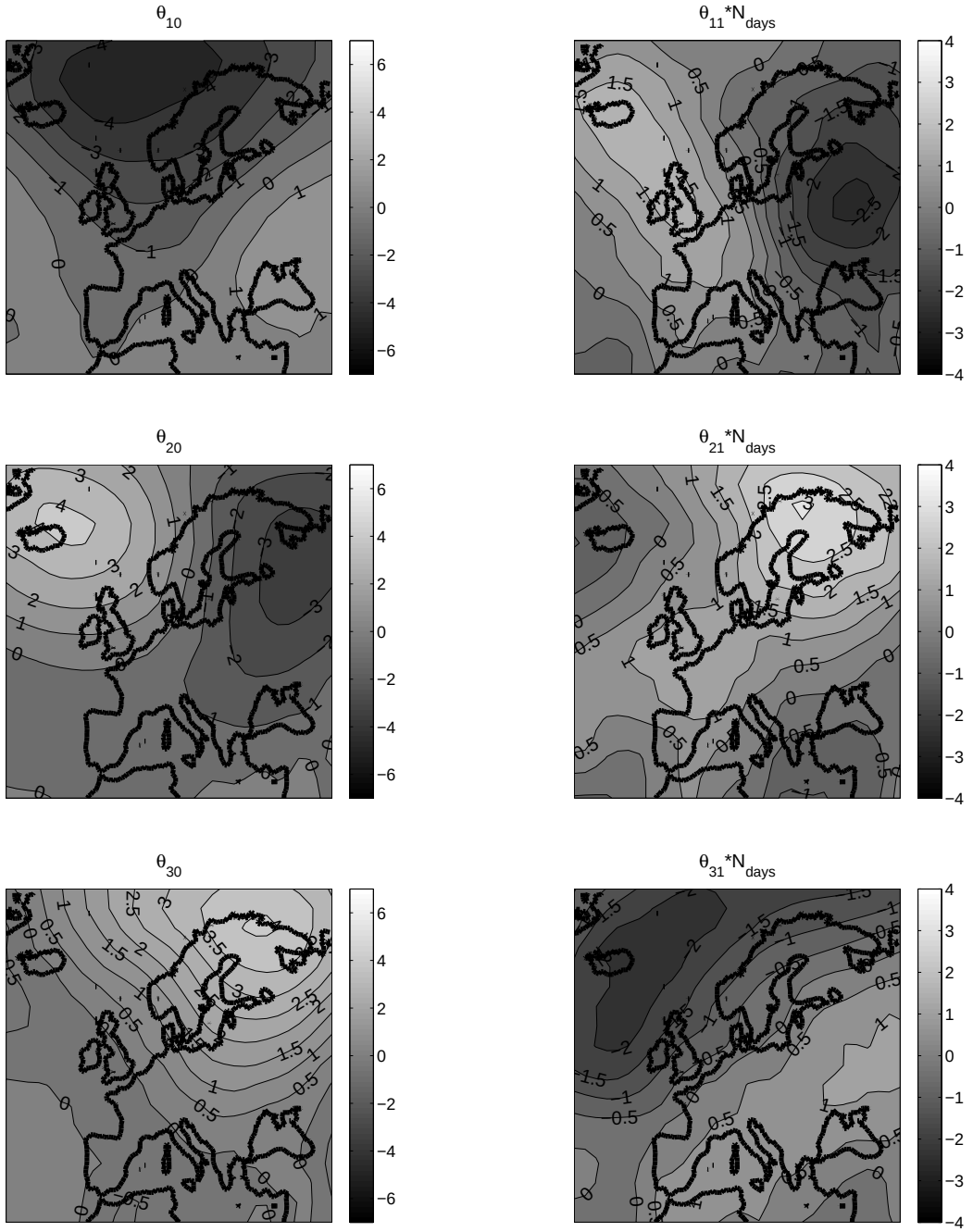

Figure 10: Regression coefficients (coloring in °C) $\theta_{i0}, i = 1, 2, 3$ and the mean temperature change $\theta_{i1} N_{days}, i = 1, 2, 3$ resulting from FEM-K-trends clustering for $\mathbf{K} = 3, \mathcal{R} = 1, \epsilon^2 = 0$. Confidence intervals for the estimated parameters do not exceed 0.4°C for θ_{i0} and 0.2°C for θ_{i1}

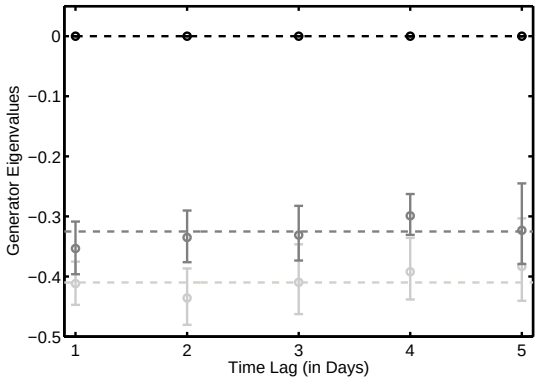

Figure 11: Markovianity test: generator eigenvalues estimated for different time lags are shown together with their confidence intervals (e. g.,time lag $\tau = 2$ means that only every second element of the transition path is taken for the estimation). Dashed lines show the mean estimates obtained for all of the shown time lags.

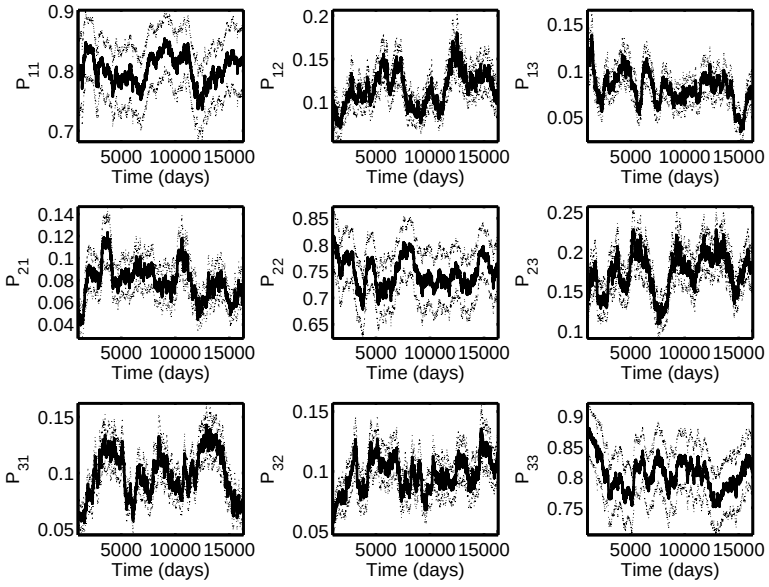

Figure 12: Inhomogeneity of the transition Markov process P: elements of the transition matrix P are calculated with the help of a moving frame of the length 1000 from the FEM-K-trends transition process. Dotted lines denote the confidence intervals for the calculated quantities.