On clustering of non-stationary meteorological time series^{*}

Illia Horenko**1

¹ Institut für Mathematik, Freie Universität Berlin Arnimallee 6, 14195 Berlin, Germany

Key words non-stationary time series analysis, clustering, regularization, finite element method Subject classification AMS: [62-07,62H30,62H25,65M60,60J10]

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 **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 Hidden Markov Models (HMM) and to finite mixture models is that no a priori assumptions about the probability model for the hidden and observed processes (e. g., Markovianity or stationarity) 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 persistent cluster states. 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 locally in Europe and on the global temperature data set.

This is a preliminary version. Do not circulate!

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 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 [1, 2, 3, 4, 5, 6, 7, 8, 9] and climatology [10, 11, 12]. 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 are: clustering methods (like *K-means* or *fuzzy-c-means*) [13], methods based on hidden Markov models (HMMs) [14, 15, 16, 17], finite mixture models [18, 19], and neuronal networks [20].

All of the above methods share two basic problems: (i) number of clusters or phases present in the data is a priori unknown [21], 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

 $^{^{\}ast}$ Supported by the DFG research center MATHEON "Mathematics for key technologies" in Berlin and SPP 1276 MetStroem.

^{**} E-mail: horenko@math.fu-berlin.de

data. Presented paper aims at investigation of this problem, introduction of the methods of nonstationary 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 [22]. 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, on analysis of historical 700 hPA geopotential height air temperature from the ERA 40 reanalysis data between 1958-2002 in Europe and worldwide.

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

1.1 Cluster distance functional and K-Means clustering

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

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

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

$$g(x_t, \theta_i) : \Psi \times \Omega \to [0, \infty),$$
 (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 functional¹

$$\mathbf{L}(\Theta, \Gamma) = \sum_{i=1}^{\mathbf{K}} \int_{0}^{T} \gamma_{i}(t) g\left(x_{t}, \theta_{i}\right) dt \to \min_{\Gamma(t), \Theta}, \qquad (3)$$

¹ Note that in context of many time series analysis methods, the model distance functional (2) is not the only quantity needed to formulate the numerical scheme. For example, in Bayesian framework (e. g., in the methods like Gaussian mixture models(GMMs) or Hidden Markov Models (HMMs) [15, 18]), the a priory chosen functional form of the probability density function for each of the clusters is also needed in each step of the numerical procedure. Therefore, in the following we will draw a distinction between methods where the probabilistic assumptions should be implied a priory (like GMM/HMM) and clustering methods that are based only on the notion of some (Euclidean, in many cases) model distance functional $g(x_t, \theta_i)$ (e. g., geometrical K-means clustering methods [23, 13]).

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

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

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

$$(5)$$

One of the most popular clustering methods in multivariate data-analysis is the so-called *K*-means algorithm [23, 13]. The affiliation to a certain cluster i is defined by the proximity of the observation $x_t \in \Psi$ to the cluster center $\theta_i \in \Psi$ (note that in this special case n = d, cf. (1)). 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:

$$g(x_t, \theta_i) = ||x_t - \theta_i||^2.$$
 (6)

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

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

$$\tag{7}$$

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

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

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

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

1.2 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 [23]

$$\sum_{i=1}^{\mathbf{K}} \sum_{j=1}^{n} \gamma_{i}^{m}(t_{j}) \parallel x_{t_{j}} - \theta_{i} \parallel^{2} \quad \to \quad \min_{\Gamma(t),\Theta},$$

$$(10)$$

where m > 1 is a fixed exponent called the *fuzzyfier* [23, 24]. Analogously to the *k*-means, the *fuzzy c*-means algorithm is an iterative procedure for minimization of (10)

$$\gamma_{i}^{(l)}(t_{j}) = \begin{cases} \frac{1}{\sum_{p=1}^{\kappa} \left(\frac{\|x_{t_{j}}-\theta_{i}^{(l-1)}\|^{2}}{\|x_{t_{j}}-\theta_{p}^{(l-1)}\|^{2}}\right)^{\frac{1}{m-1}}} & \text{if } \mathbf{I}_{x_{t_{j}}} \text{ is empty,} \\ \sum_{r \in \mathbf{I}_{x_{t_{j}}}} \gamma_{r}^{(l)}(t_{j}) = 1 & \text{if } \mathbf{I}_{x_{t_{j}}} \text{ is not empty,} i \in \mathbf{I}_{x_{t_{j}}}, \\ 0 & \text{if } \mathbf{I}_{x_{t_{j}}} \text{ is not empty,} i \notin \mathbf{I}_{x_{t_{j}}}, \end{cases} \end{cases}$$

$$\theta_{i}^{(l)} = \frac{\sum_{j=1}^{n} \gamma_{i}^{(l)}(t_{j}) x_{t_{j}}}{(\mathbf{I})}. \qquad (12)$$

$$P_{i}^{(l)} = \frac{\sum_{j=1}^{n} \gamma_{i}^{(l)}(t_{j}) x_{t_{j}}}{\sum_{j=1}^{n} \gamma_{i}^{(l)}(t_{j})}.$$
(12)

where $\mathbf{I}_{x_{t_j}} = \{p \in \{1, \dots, \mathbf{K}\} | \| x_{t_j} - \theta_p^{(l-1)} \|^2 = 0\}$ [13]. Note that the middle part of the assignment (12) means that the assignment is not unique and any assignment where $\sum_{r \in \mathbf{I}_{x_{t_j}}} \gamma_r^{(l)}(t_j) = 1$ can be used in this case. 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*.

1.3 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) [22]. 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)

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

$$\tag{13}$$

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

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

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 [25]. 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 [25]) 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 **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 [26]. Dynamical approach to control the cluster-fuzzyness and the number of clusters will be introduced.

2.1 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

$$g(x_t, \theta_i) = \| x_t - \sum_{k=0}^{\mathcal{R}} \theta_{ik} \odot \phi_k(t) \|^2,$$
(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 [26], 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.:

$$|\gamma_i|_{\mathcal{H}^1(0,T)} = \|\partial_t \gamma_i(\cdot)\|_{\mathcal{L}_2(0,T)} = \int_0^T \left(\partial_t \gamma_i(t)\right)^2 dt \le C_{\epsilon}^i < +\infty, \quad i = 1, \dots, \mathbf{K}.$$
 (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* $\Gamma(t)$ [26].

2.2 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

$$\mathbf{L}^{\epsilon}(\Theta,\Gamma,\epsilon^{2}) = \mathbf{L}(\Theta,\Gamma) + \epsilon^{2} \sum_{i=1}^{\mathbf{K}} \int_{0}^{T} \left(\partial_{t}\gamma_{i}\left(t\right)\right)^{2} dt \to \min_{\Gamma,\Theta}.$$
(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* [27]

$$v_{k}(t) = \begin{cases} \frac{t-t_{k}}{\Delta_{t}} & 2 \le k \le N-1, t \in [t_{k-1}, t_{k}], \\ \frac{t_{k+1}-t}{\Delta_{t}} & 2 \le k \le N-1, t \in [t_{k}, t_{k+1}], \\ \frac{t_{2}-t}{\Delta_{t}} & k = 1, t \in [t_{1}, t_{2}] \\ \frac{t-t_{N-1}}{\Delta_{t}} & k = N, t \in [t_{N-1}, t_{N}]. \end{cases}$$
(18)

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

$$\gamma_i = \tilde{\gamma}_i + \delta_N$$

= $\sum_{k=1}^N \tilde{\gamma}_{ik} v_k + \delta_N,$ (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

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

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

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

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

$$a(\theta_i) = \left(\int_{t_1}^{t_2} v_1(t)g(x_t, \theta_i)dt, \dots, \int_{t_{N-1}}^{t_N} v_N(t)g(x_t, \theta_i)dt \right),$$
(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 [26] 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

$$\gamma_{i}^{(l)}(t_{j}) = \begin{cases} 1 & i = \arg\min \int_{t_{j}}^{t_{j+1}} v_{j}(s) \parallel x_{s} - \sum_{k=0}^{\mathcal{R}} \theta_{ik}^{(l)} \odot \phi_{k}(s) \parallel^{2} ds, \\ 0 & \text{otherwise}, \end{cases}$$
(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 $\mathcal{O}(N \log(N))$ [28, 29]. 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 **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

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

Since $g(x_t, \theta_i)$ 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 **K**, regularization factor ϵ^2 , finite discretization
- of the time interval [0, T], and the optimization tolerance TOL
- \cdot Set the iteration counter l=1
- · Choose random initial $\tilde{\gamma}_i^{(1)}, i = 1, \dots, \mathbf{K}$ satisfying (21-22) · Calculate $\Theta^{(1)} = \underset{\Theta}{\operatorname{arg\,min}} \tilde{\mathbf{L}}^{\epsilon} \left(\Theta, \tilde{\gamma}_i^{(1)}\right)$ solving the linear regression problem (25)

Optimization loop:

do

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

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

while
$$\left| \tilde{\mathbf{L}}^{\epsilon} \left(\Theta^{(l)}, \tilde{\gamma}_{i}^{(l)} \right) - \tilde{\mathbf{L}}^{\epsilon} \left(\Theta^{(l-1)}, \tilde{\gamma}_{i}^{(l-1)} \right) \right| \ge \text{TOL}.$$

Major advantage of the presented algorithm compared to HMM-based strategies [17, 30, 31] and to finite mixture models [18, 19] 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 [21].

As was demonstrated in [26], 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 **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 **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}, ϵ^2 . Markovianity can be verified applying some standard tests, e. g., one can check the *generator structure* of the hidden process, see [32]. 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 [33, 34].

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 [35, 31]. 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 [36].

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 **K** of *metastable* clusters: (i) the one based on the estimation of the *number of statistically distinguishable* clusters, (ii) the other based on the *a posteriory* Markov assumption for the hidden process $\Gamma(t)$.

Finally, in example (c.) the application of the *FEM-K-trends*-algorithm is demonstrated on metastable clustering of historical temperature data, first on the <u>local European</u>, then on the <u>global worldwide time</u> <u>series data</u>. Two examples of a posteriory post-processing of the clustering results is presented: (i) construction of the reduced non-stationary Markovian model for the European data and (ii) Fourier-analysis for the global data.



Fig. 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.



Fig. 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.

4.1 Two hidden states

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



Fig. 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.

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

4.2 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



Fig. 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.



Fig. 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 $\sigma = 7$.

is *metastable*, as well as the combination of the first and the second states together builds the second *metastable cluster set*.

$$\begin{aligned} x_{j}(t) &= \theta_{i(t)j}(t - \bar{t}_{j}) + \sigma \mathbf{N}(0, 1), \quad j = 1, 2, 3\\ \theta_{1} &= \left(\begin{array}{ccc} 0.0 & 0.0 & 0.0 \end{array}\right)\\ \theta_{2} &= \left(\begin{array}{cccc} 0.01 & -0.01 & 0.01 \end{array}\right), \quad \theta_{3} = \left(\begin{array}{cccc} -0.01 & 0.01 & -0.01 \end{array}\right)\\ \bar{t} &= \left(\begin{array}{cccc} 0.0 & 300 & 600 \end{array}\right) \end{aligned}$$
(27)



Fig. 6 Maximal number of statistically distinguishable cluster states for different values of ϵ^2 as calculated for the time series from Fig. 4.



Fig. 7 Results of the optimal linear regression fit $(\theta_0 + \theta_1 t)$ for the whole length of the analyzed temperature time series.



Fig. 8 Comparison of two criteria for the choice of **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 K: (i) spectral analysis of the Markov transition matrix and (ii) variation of the regularization parameter ϵ^2 and careful comparison of the



Fig. 9 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).

respective cluster affiliations $\gamma_i(t), i = 1, ..., \mathbf{K}$. 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 **K** 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 [33, 34]. 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 [32]. 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 **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 **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 **K** should be decreased. Analysis of the same data with **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 **K** = 2 as an optimal number of metastable sets in the data. Results of the consequent application of the described procedure for different values of ϵ^2 are summarized in the left panel of Fig. 12.

4.3 Analysis of Historical Temperature Data in Europe (1959-2002)

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 $[40]^{12}$. 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 1959 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.

 $^{^{12}}$ The author thanks the *Potsdam Institute for Climate Impact Research (PIK)* for the possibility to use the ERA 40 data.



Fig. 10 Mean temperature deviation patterns $\Delta T(t)$ at the beginning (left column) and at the end of the analyzed period (right column). Patterns are calculated from the optimal linear regression fit ($\Delta T(t) = \theta_{i0} + \theta_{i1}t$, i = 1, 2, 3 (coloring in °C) for each of the cluster states identified by the *K*-means algorithm for $\mathbf{K} = 3$.

In the following we will demonstrate the application of different clustering strategies to this time series of *deviations from the mean temperatures* and will compare the resulting *mean temperature deviation patterns* wrt. their long-term behavior.

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. 7, 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. 8 illustrates the comparison of two alternative criteria used to



Fig. 11 Mean temperature deviation patterns $\Delta T(t)$ at the beginning (left column) and at the end of the analyzed period (right column). Patterns are calculated from the optimal linear regression fit ($\Delta T(t) = \theta_{i0} + \theta_{i1}t, i = 1, 2, 3$ (coloring in °C) for each of the cluster states identified by the *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} .

determine the number of clusters \mathbf{K} in the data. First we apply the *Explained Cluster Variance (ECV)* criterion, defined as

$$ECV(\mathbf{K}) = 1 - \frac{\sum_{i=1}^{K} \sum_{j=1}^{n} \gamma_{i}^{m}(t_{j}) || x_{t_{j}} - \mu_{i} ||^{2}}{\sum_{j=1}^{n} || x_{t_{j}} - \mathbf{E}(x_{t}) ||^{2}},$$

$$\mathbf{E}(x_{t}) = \frac{1}{n} \sum_{j=1}^{n} x_{t_{j}}, \quad \mu_{i} = \sum_{j=1}^{n} \gamma_{i}^{m}(t_{j}) x_{t_{j}} / \sum_{j=1}^{n} \gamma_{i}^{m}(t_{j})$$
(28)

where μ_i are the geometrical cluster centers. As it is demonstrated on the left panel of Fig. 8, the value of ECV (**K**) increases uniformly with **K** and implicates no obvious choice of **K**. The right panel of Fig. 8 shows the eigenvalues of the Markov-generator estimated from the transition process resulting



Fig. 12 Left panel: Maximal number of statistically distinguishable cluster states for different values of ϵ^2 as calculated for the European temperature time series with FEM-K-trends algorithm. Right panel (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.



Fig. 13 Time dependence of the transition Markov process P: elements of the transition matrix P are calculated with the help of a local Gaussian kernel smoothing algorithm [37, 38] (width of the Gaussian moving window is taken to be 7 years) from the *FEM-K-trends* transition process (see Fig. 9). Dotted lines denote the confidence intervals calculated with the bootstrap-algorithm [39].

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. 9 (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 a posteriory trends calculated for the *K-means* clusters as it was done in [41] (see Fig. 10) and the values resulting from the *FEM-K-trends* clustering (see Fig. 11). Confidence intervals for the regression coefficients were calculated using the standard bootstrap algorithm in both cases [39] (respective mean error of mean temperature change $\theta_{i1}N_{days}$ was estimated to be ± 0.4 °C). As



Fig. 14 Time dependence of the instantaneous Markovian statistical weights (30). Metastable state 1 corresponds to the negative temperature anomaly over Scandinavia (upper row of Fig. 9), state 2 stands for the positive temperature anomaly over northern Atlantic (middle row of Fig. 9) and state 3 denotes the positive temperature anomaly over Scandinavia (lower row of Fig. 9). Dotted lines denote the confidence intervals calculated with the bootstrap-algorithm [39].



Fig. 15 The ERA 40 temperature series with eliminated seasonal component on one of the grid points (left panel) and its Fourier spectrum (right panel).

it can be seen from comparison of the left and the right columns of Fig. 10, in the case of the K-Means clustering, the overall change of the mean temperature deviation pattern $\Delta(t)$ is small and is comparable with the size of the respective confidence interval. This means that the identified trends are hardly distinguishable from stationary regimes with no trend. It demonstrates the influence of the implicit mathematical assumption beyond the applied clustering method: data is supposed to be locally stationary in each of the regimes for K-Means algorithm. Fig. 11 on the other hand illustrates that application of the FEM-K-Trends strategy (where the data is assumed having temporal trend a priory) results in much more significant changes of mean temperature deviation patterns for the cluster states, with different statistically significant temporal trends. Fig. 11 also demonstrates that temporal trends



Fig. 16 Maximal number of statistically distinguishable cluster states for different values of ϵ^2 as calculated for the worldwide temperature time series with FEM-K-trends algorithm.



Fig. 17 Cluster affiliation for the global temperature data as identified with FEM-K-trends algorithm ($\mathbf{K} = 2, \epsilon^2 = 400$).

identified for the analyzed data in 558-dimensional vector space of temperature deviations $\Delta T(t)$ can be interpreted as *spatial shifts* of the mean patterns over the European continent. For example, the positive temperature anomaly over northern Scandinavia characteristic for the third cluster state at the beginning of the analyzed period (left panel in the last row of the Fig. 11) is moving southwards and finally has a maximum over southern Finnland and the European part of northern Russia, simultaneously an additional negative anomaly is originating over the northern part of Atlantic at the end of the analyzed period (right panel in the last row of the Fig. 11).

Finally, we analyze the process of transitions between the identified cluster states and try to find an appropriate *dynamical* description for it. As it was mentioned above, the major advantage of the presented *FEM-K-trends* approach compared to the HMM-based strategies [17, 30, 30] is its independence on assumptions about the type of the probability model. Therefore one does not have to assume a priori that the transition process is an output of the time-homogenous Markov chain. In context of



Fig. 18 Fourier spectra of the hidden process $\Gamma(t)$ from Fig. 17 (left panel) and of the time-dependent statistical weights (see upper panel of Fig. 19) of two cluster states obtained by a local Gaussian kernel smoothing algorithm [37, 38] of the process $\Gamma(t)$ (Gaussian window is taken to have the width of two years to filter out the seasonal signal).



Fig. 19 Upper panel: Time-dependent statistical weight (see upper panel of Fig. 19) of the cluster state 2 obtained by a local Gaussian kernel smoothing algorithm [37, 38] of the process $\Gamma(t)$ from Fig. 17 (Gaussian window is taken to have the width of two years to filter out the seasonal signal). Middle panel: The dominant Fourier component of the above signal with 11.1 years/cycle periodicity. Lower panel: number of sunspots as the function of time in the same time period (from http://solarscience.msfc.nasa.gov/SunspotCycle.html)

FEM-K-trends this assumption can be checked a posteriori and can help to construct reduced dynamical Markovian models based on the observation data. We can apply standard methodology based on generator estimation [32] to check the Markov property for the transition process $\gamma(t)$ (identified with the FEM-K-Trends algorithm). As it is shown in the Fig.12, the eigenvalues of the underlying generator can be assumed to be time-independent, therefore the process switching between 3 regression models can be assumed to be Markovian (at least in homogenous approximation). To investigate a time-dependence of the identified Markovian process switching between the linear regressive models from Fig. 11, we



Fig. 20 Mean temperature deviation patterns $\Delta T(t)$ at the beginning (left column) and at the end of the analyzed period (right column) for the global temperature data (1959-2008). Patterns are calculated from the optimal linear regression fit ($\Delta T(t) = \theta_{i0} + \theta_{i1}t$, i = 1, 2 (coloring in °C) for each of the cluster states identified by the *FEM-K-trends clustering* for $\mathbf{K} = 2, \mathcal{R} = 1, \epsilon^2 = 400$. Confidence intervals for the estimated parameters do not exceed 0.5°C for θ_{i0} and 0.3°C for θ_{i1} .

apply the standard *local Gaussian kernel smoothing* algorithm [37, 38]. A moving Gaussian window $\gamma(t, t_0) = \frac{1}{c} \exp(-\frac{(t-t_0)^2}{\sigma^2})$ (where c is the normalization constant) of 7 years width is defined, we slide it along the time series of transition process $\Gamma(t)$ and weight the observed transitions with the value of the current Gaussian function. Then

$$P_{ij}(t_0) = \frac{\sum_{t \in \{t_{ij}\}} \gamma(t, t_0)}{\sum_{t \in \{t_i\}} \gamma(t, t_0)},\tag{29}$$

where $\{t_i\}$ is the set of all time instances when the state *i* was visited and $\{t_{ij}\}$ is the set of all time instances when the transitions between the clusters *i* and *j* are observed.

The resulting transition probabilities as functions of time are shown in Fig. 13. It demonstrates that the Markovian transition process is not time-homogenous and that there is a significant increase of the probability to stay in the cluster 1 (describing the negative temperature anomaly over northern Scandinavia) and increasing probabilities to make a transition to the cluster 3 (correspondent to the positive temperature anomaly over Scandinavia).

In order to get another interpretation of the obtained results, instantaneous Markovian statistical weights $\pi_i(t)$, i = 1, 2, 3

$$\pi(t) P(t) = \pi(t), \quad \pi(t) = (\pi_1(t), \dots, \pi_m(t))$$
(30)

can be calculated. This quantities characterize the equilibrium probabilities to find the analyzed system in one of the clusters in different moments of time. Fig. 14 shows that the probability of the cluster state 1 decreases and probability of cluster state 3 increases, i. e. the negative temperature anomalies

over Scandinavia became less probable, whereas the probability of the positive anomalies over the Scandinavian region was increasing.

Together with the spatial shift of the respective anomaly positions in Fig. 11, this finding gives an interesting insight into the character of the warming process in the northern Europe in last years.

4.4 Analysis of Global Temperature Data (1959-2002)

Description of the data Finally, we analyze the worldwide daily mean values of the 700 hPa air temperature field from the ERA 40 reanalysis data $[40]^{21}$. The analyzed time series are given for every point on the 146×82 grid covering the whole Earth surface. For the analysis, in the same way is it was done in the above example of European data analysis, we have considered temperature values only for the period 1959 till 2002, thus we end with a 11972-dimensional time series of 16314 days.

In order to remove the seasonal trend, we apply a standard seasonal trend elimination procedure as described in above example. In the following we will demonstrate an application of the described FEM-K-trends algorithm to analysis of this data, we will exemplify the estimation of optimal number of persistent cluster states and postprocessing of the results by means of the Fourier-analysis.

Discussion of the results Before applying the FEM-K-trends procedure to the considered time series, we can have a look at one of the time series at some arbitrarily chosen grid point. As it can be seen from the Fig. 15, although the seasonal component was eliminated from the data, there is still a strong 1-year/cycle period in the data, present mainly in the data variance. However, just looking at the Fourier spectrum of one single data dimension makes not possible to say something about how general this feature is for the whole 11972-dimensional data and what kind of other periods are present.

We apply the procedure described in Sec. 3 to determine the optimal number of statistically distinguishable metastable states in the analyzed time series. Fig. 16 demonstrates that for a very wide range of regularization parameters ϵ^2 there are basically only two distinguishable metastable clusters in the data (parameter \mathbf{R} was chosen to be 1, i. e., the linear regressive models for the cluster states were considered). Careful inspection of the corresponding hidden process $\Gamma(t)$ shown in Fig. 17 suggests that both identified clusters have something to do with the seasonal states of the data, similar to the results obtained with the stationary HMM-PCA-SDE procedure [17]. This suggestion is verified when looking at the Fourier spectrum of the $\Gamma(t)$ (see the left panel of Fig. 18). In contrast to the Fourier-spectrum of the single one-dimensional grid point series from Fig. 15, the variable $\Gamma(t)$ characterizes the global dynamics of the whole 11972-dimensional data. It has two clearly pronounced Fourier components, the one correspondent to the one-year period and the second correspondent to a period of approximately 11.1 years. Applying the local Gaussian kernel smoothing algorithm [37, 38] to $\Gamma(t)$ (Gaussian window is taken to have the width of two years to filter out the seasonal signal) (see the upper panel of Fig. 19) reveals a signal with dominant 11.1 year period (see the right panel of Fig. 18). Comparison of this dominant Fourier component (see the middle pannel of Fig. 19) with the time series of sun spots for the same period reveals that the temporal behavior of this component coincides (up to a certain phase-shift) with the 11-year solar activity cycle.

Finally, respective overall temperature deviation trends are considered, in the same manner as it was done for the previous example (see Fig. 14). As can be seen from the Fig. 20, the first cluster states is characterized by more pronounced positive temperature deviations in the northern hemisphere, whereas the second state is characterized by more positive anomalies in the southern hemisphere. The identified 11972-dimensional trends shown in the Fig. 20 are spatially inhomogeneous and are mostly pronounced in the polar regions, with the overall tendency to warming.

5 Conclusion

Short overview of the existing clustering methods was given, implicit mathematical assumptions imposed on the data by different analysis methods were explained and compared. It was shown how big is the

 $^{^{21}}$ The author thanks the *Potsdam Institute for Climate Impact Research (PIK)* for the possibility to use the ERA 40 data.

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.

New numerical framework for clustering multidimensional meteorological time series was presented, specially designed to identify the metastable (or persistent) cluster regimes in the multidimensional nonstationary data driven by long term trends in the expectation values. Approach is based on minimization of a specially constructed *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 states* in the analyzed data. It was investigated, how the optimal number of the persistent atmospheric states and their long-term variability can be obtained, applying the new methodology without assuming the stationarity of the analyzed regimes.

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 identification of persistent atmospherical states. Significant discrepancies between the temperature trends identified for *K*-means and *FEM*-*K*-trends clusters were found.

Markovianity of the identified transition process for the European data was a posteriori verified and 3 metastable temperature clusters were identified. The same number of metastable cluster states was identified with the help of the strategy based on the estimation of the maximal number of distinguishable clusters, without a posteriory Markov-assumption. Resulting cluster regressions are compared with the trends calculated from standard stationary K-means clustering and the discrepancies were discussed.

The identified *persistent* cluster states in both cases were investigated wrt. the long-term variability of their relative statistical weights. It was demonstrated how the a posteriori post-processing of the clustering results can be used to get an additional insight into the multidimensional non-stationary data. In the case of the European temperature data, how the Markov assumption can be verified and how the reduced non-homogenous Markov model can be constructed. In the case of the worldwide data it was shown how the Fourier analysis can be used to extract some *global* periodicity patterns in the data.

Presented methodology has a potential to become helpful for the analysis and assimilation of very large data sets in different areas of geophysical research.

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" and the DFG Research Center MATHEON "Mathematics for Key Technologies".

References

- A.A. Tsonis and J.B. Elsner. Multiple attractors, fractal basins and longterm climate dynamics. <u>Beit. Phys.</u> Atmos., 63:171–176, 1990.
- [2] M. Kimoto and M. Ghil. Multiple flow regimes in the northern hemisphere winter. part i: Methodology and hemispheric regimes. Journal of Atmospheric Sciences, 50:2625–2644, August 1993.
- [3] M. Kimoto and M. Ghil. Multiple flow regimes in the northern hemisphere winter. part ii: Sectorial regimes and preferred transitions. Journal of Atmospheric Sciences, 50:2645–2673, August 1993.
- [4] X. Cheng and J. M. Wallace. Cluster analysis of the northern hemisphere wintertime 500-hpa height field: Spatial patterns. Journal of Atmospheric Sciences, 50:2674–2696, August 1993.
- [5] V. V. Efimov, A. V. Prusov, and M. V. Shokurov. Patterns of interannual variability defined by a cluster analysis and their relation with enso. <u>Quarterly Journal of the Royal Meteorological Society</u>, 121:1651–1679, October 1995.
- [6] I.I. Mokhov and V.A. Semenov. Bimodality of the probability density functions of subseasonal variations in surface air temperature. Izvestiya, Atmos. Ocean. Phys., 33:702–708, 1997.
- [7] I.I. Mokhov, V.K. Petukhov, and V.A. Semenov. Multiple intraseasonal temperature regimes and their evolution in the iap ras climate model. Izvestiya, Atmos. Ocean. Phys., 34:145–152, 1998.
- [8] S. Corti, F. Molteni, and T. N. Palmer. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. Nature, 398:799–802, April 1999.
- [9] T. N. Palmer. A Nonlinear Dynamical Perspective on Climate Prediction. <u>Journal of Climate</u>, 12:575–591, February 1999.
- [10] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. Stochastic resonance in climatic change. <u>Tellus</u>, 3:10–16, 1982.
- [11] C. Nicolis. Stochastic aspects of climatic transitions-response to a periodic forcing. Tellus, 34:1-+, 1982.
- [12] D. Paillard. The timing of pleistocene glaciations from a simple multiple-state climate model. <u>Nature</u>, 391:378–381, January 1998.
- [13] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. <u>Fuzzy cluster analysis</u>. John Wiley and Sons, New York, 1999.
- [14] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans. Informat. Theory, 13:260–269, 1967.
- [15] J.A. Bilmes. <u>A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for</u> <u>Gaussian Mixture and Hidden Markov Models. Thechnical Report</u>. International Computer Science Institute, Berkeley, 1998.
- [16] A. Majda, C. Franzke, A. Fischer, and D. Crommelin. Distinct metastable atmospheric regimes despite nearly gaussian statistics : A paradigm model. PNAS, 103(22):8309–8314, 2006.
- [17] I. Horenko, R. Klein, S. Dolaptchiev, and Ch. Schuette. Automated generation of reduced stochastic weather models i: simultaneous dimension and model reduction for time series analysis. <u>SIAM Mult. Mod. Sim.</u>, 6(4):1125–1145, 2008.
- [18] G. McLachlan and D. Peel. Finite mixture models. Wiley, New-York, 2000.
- [19] S. Fruhwirth-Schnatter. Finite Mixture and Markov Switching Models. Springer, 2006.
- [20] A.H. Monahan. Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system. J. Climate, 13:821–835, 2000.
- B. Christiansen. Atmospheric circulation regimes: Can cluster analysis provide the number? J. Climate, 20(10):2229-2250, 2007.
- [22] R.H. Hathaway and J.C. Bezdek. Switching regression models and fuzzy clustering. <u>IEEE Transactions on</u> Fuzzy Systems., 1:195–204, 1993.
- [23] J.C. Bezdek. Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York, 1981.
 [24] J.C. Bezdek, R.H. Hathaway, M.J. Sabin, and W.T. Tucker. Convergence theory for fuzzy c-mens: counterexamples and repairs. IEEE Trans. Systems., 17:873–877, 1987.
- [25] A. Preminger, U. Ben-Zion, and D. Wettstein. The extended switching regression model: allowing for multiple latent state variables. J. of Forecasting, 26:457–473, 2007.
- [26] I. Horenko. Finite element approach to clustering of multidimensional time series. <u>submitted to SIAM</u> Journal of Sci. Comp., 2008. (available via biocomputing.mi.fu-berlin.de).
- [27] D. Braess. Finite Elements: Theory, Fast Solvers and Applications to Solid Mechanics. Cambridge University Press, 2007.
- [28] P.E. Gill, W. Murray, M.A. Saunders, and M.H. Wright. A Schur-complement method for sparse quadratic programming. Technical report, STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB, 1987.
- [29] M. Arioli. The use of QR factorization in sparse quadratic programming and backward error issues. <u>SIAM</u> J. on Matrix Analysis and Applications, 21(3):825 – 839, 2000.

- [30] I. Horenko. On simultaneous data-based dimension reduction and hidden phase identification. J. Atmos. Sci, to appear, 2008. (available via biocomputing.mi.fu-berlin.de).
- [31] I. Horenko, S. Dolaptchiev, A.V. Eliseev, I.I. Mokhov, and R. Klein. Metastable decomposition of high-dimensional meteorological data with gaps. J. Atmos. Sci., to appear, 2008. (available via biocomputing.mi.fu-berlin.de).
- [32] P. Metzner, I. Horenko, and Ch. Schuette. Generator estimation of Markov Jump processes based on incomplete observations nonequidistant in time. Phys. Rev. E, 76:0667021, 2007.
- [33] Ch. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In P. G. Ciaret and J.-L. Lions, editors, <u>Handbook of Numerical Analysis</u>, volume X, pages 699– 744. Elsevier, 2003.
- [34] W. Huisinga, S. Meyn, and Ch. Schuette. Phase transitions and metastability in Markovian and molecular systems. Ann. Appl. Prob, 14(1):419–458, 2004.
- [35] H. Gardiner. <u>Handbook of stochastical methods</u>. Springer, Berlin, 2004.
- [36] B. Kedem and K. Fokianos. <u>Regression models for time series analysis</u>. Wiley Series in Probability and Statistics, 2002.
- [37] C. Loader. Local likelihood density estimation. The Annals of Statistics, 24, 1996.
- [38] C. Loader. Local Regressions and Likelihood. Springer, New Yorck, 1999.
- [39] A. Davison and D. Hinkley. <u>Bootstrap methods and their applications</u>. Cambridge Series in Statistical and Probability Mathematics, 2006.
- [40] A. Simmons and J. Gibson. The ERA 40 project plan. In <u>ERA 40 Project Rep. Ser. 1</u>, 2000. European Center for Medium-Range Weather Forcasting, Reading.
- [41] A. Philipp, P.M. Della-Marta, J. Jacobeit, D.R. Fereday, P.D. Jones, A. Moberg, and H. Wanner. Long term variability of daily north Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering. J. Climate, 20(16):4065–4095, 2007.