On simultaneous data–based dimension reduction and hidden phase identification

Ilia Horenko
Institut für Mathematik,
Freie Universität Berlin,
Arnimallee 6, 14195 Berlin,
Germany
E-mail: horenko@math.fu-berlin.de

October 31, 2007
Abstract

We consider a problem of simultaneous dimension reduction and identification of hidden attractive manifolds in multidimensional data with noise. The problem is approached in two consecutive steps: (i) embedding of the original data in a sufficiently high-dimensional extended space in a way proposed by F. Takens in his embedding theorem (F. Takens, in: Dynamical Systems and Turbulence, D.A. Rand and L.S. Young eds., Springer, New York, 1981) followed by (ii) a minimization of the residual–functional. We construct the residual–functional to measure the distance between the original data in extended space and their reconstruction based on a low dimensional description. The reduced representation of the analyzed data results from projection onto a fixed number of unknown low–dimensional manifolds. Two specific forms of the residual–functional are proposed. They define two different types of essential coordinates: (i) localized essential orthogonal functions (EOFs) and (ii) localized functions which we call principal original components (POC). The application of the framework is exemplified on a Lorenz–attractor model with measurement noise and on historical air temperature data. It is demonstrated how the new method can be used for the elimination of noise and identification of the seasonal low-frequency components in meteorological data. We also present an application of the proposed POC-components in the context of the construction of low-dimensional predictive models.
Introduction

The aim of the present paper is to discuss some existing techniques and to demonstrate several new ideas of reduced stochastic modelling based on observed multidimensional meteorological data. Many dynamical systems in meteorology and climate research are characterized by the presence of low–dimensional attractors, i. e., the dynamics of those processes converges towards some low–dimensional manifolds of the high–dimensional phase space (Lorenz 1963; Dijkstra and Neelin 1995; Lorenz 2006; Guo and Huang 2006). An important question in the analysis of such systems is identification of those manifolds since they give a reduced representation of the original dynamics in few essential degrees of freedom. If knowledge about the system is present only in the form of observation or measurement data and there is no a priori knowledge about the equations governing the system’s dynamics, the challenging problem of identification of those attractive manifolds together with the construction of reduced low-dimensional models becomes a problem of time series analysis and pattern recognition in many dimensions. The choice of the appropriate data analysis strategies (implying a set of method-specific assumptions on the analyzed data) plays a crucial role in the correct interpretation of the available time series. Correct identification of attractors in the observed meteorological and climate data will allow for a reduced stochastic description of multidimensional processes and can help to improve the quality of resulting predictions.

In most of the cases the problem of the identification of attractors is hindered by the fact that they have very complicated, even fractal geometry and non–integer dimensionality, so the points close to each other in Euclidean space can be infinitely wide apart on the attractor. Therefore, in order to be able to apply the methods of topological data–analysis and clustering requiring the notion of a geometrical distance between the different points in the phase space, it is necessary to transform the data into appropriate form. First theoretical results about the existence of such transformations (or embeddings) for certain classes of dynamical systems were obtained by H. Whitney (Whitney 1936), followed by the method of delays proposed by N. Packard (Packard et al. 1980). Later, in the context of the turbulence dynamics, Takens embedding theorem provided a theoretical background to the method of delays (Takens 1981). However, none of these results gives a general strategy of how to calculate the dimensionality of the attractor if it is a priori unknown and how
to extract the relevant information from the measurements.

D. Broomhead and G. King have suggested to calculate the dimension \( m \) of the attractive manifold _aposteriori_ by taking the embedding dimension sufficiently high and then calculating \( m \) from the _numerical rank_ of the _covariance matrix_ of the embedded data (Broomhead and King 1986; Broomhead et al. 1988). From the mathematical point of view, this is equivalent to the application of the _principal component analysis_ (PCA) to the embedded data and approximation of the attractive manifold by an \( m \)-dimensional _linear manifold_ built by \( m \) dominant eigenvectors of the covariance. This ideas were further developed by R. Vautard, P. Yiou and M. Ghil in terms of the _singular-spectrum analysis_ (SSA) and were implemented for analysis of surface air temperature time series (Vautard et al. 1992). However, these approaches implicitly rely on the assumption that the attractor can be _globally_ approximated by a linear manifold. The _quality_ of this approximation (and therefore also the validity of the aforementioned assumption) remained beyond consideration. We understand the _approximation quality_ here in a sense of some functional describing the distance between the really existing attractive manifold and its approximation by a linear manifold.

In the present paper we show a way of constructing a _quality_ functional. In the framework given by Takens’ embedding theorem, this functional quantifies the _reconstruction error_, i.e., a distance between the original data _embedded_ into Euclidean space and their _reconstruction_ based on a low dimensional description. In contrast to traditional approaches like SSA, we assume that the attractor manifold can be represented as a combination of a fixed number \( K \) of \( m \)-dimensional _hyperplanes_ in an \( n \)-dimensional embedding (or _extended space_ as we will call it later). The reduced representation of the analyzed data results from a _projection_ on \( K \) unknown low–dimensional linear manifolds. We demonstrate how two special constraints imposed on the _projection operators_ result in two different forms of dimension reduction: (i) the first is an extension of the PCA towards multiple _hidden states_ (with _essential coordinates_ calculated as _linear combinations_ of all of the original degrees of freedom in extended space) and (ii) the second is what we call the _principal original components_ (POC) (with _essential coordinates_ that are _subsets_ of the original degrees of freedom in extended space).

We demonstrate the application of the proposed method to the Lorenz–oscillator system with measurement noise and to the analysis of histor-
ical air temperature data. It is exemplified how the new method can be used for the elimination of noise and identification of the seasonal low-frequent components in meteorological data and the results are compared with other existing methods of seasonal cycle elimination. We also present an application of the proposed POC-method in context of the construction of low-dimensional predictive models of temperature dynamics in the atmosphere.

**Extraction of topological information from time series**

Consider a dynamical system

\[
\frac{dx}{dt} = F(x),
\]

with vector \( x \in M \subset \mathbb{R}^n \) representing a state of the system and \( M \) being the space of all possible system configurations or phase space. If \( F(x) \) is locally Lipschitz, (1) defines an initial value problem (i. e., a unique solution curve passes through each point \( x \subset M \)). In this case we can formally write the solution at any time \( t \) given an initial value \( x_0 = \phi_t x_0 \), where \( \phi_t \) represents a so-called flow of dynamical system (1). In many dynamical systems it is the case that their flows evolve towards some low-dimensional objects called attractors. \( A \subset M \) is called attractor if the following conditions are satisfied: (i) \( A = \phi_t A, \forall t \), i. e. \( A \) is invariant under the flow, (ii) there is a neighborhood of \( A \), \( B(A) \) called the basin of attraction for \( A \), such that \( B(A) = \{s | \forall N \subset B(A), \exists T, \forall t > T, \phi_t s \in N \} \), and (iii) there is no subset of \( A \) with the first two properties.

The identification of attractors for dynamical systems is important because the projection of the original high-dimensional systems observation \( x \) on the attractor allows to find a reduced essential representation of the systems dynamics which is low-dimensional if \( \dim(A) \ll \dim(M) \). In the context of data-analysis this can help to reduce the dimensionality of the observation vector and to plan an optimal measurement process with a minimal number of measured system’s degrees of freedom involved. However, in context of time series analysis, this task can be hindered by the fact that the attractor may have very complicated, even fractal geometry and non-integer dimensionality, so the points close to each other in Euclidean space can be infinitely far apart on the attractor. This issue can restrict the application of topological data-analysis and
clustering methods that require the notion of a geometrical distance between the different points in the phase space. Therefore, in order to be able to extract the essential topological information out of the measured data, it is necessary to transform the data into an appropriate form.

As was first shown by H. Whitney, sufficiently smooth connected m-dimensional manifolds can be smoothly embedded in Euclidean \((2m + 1)\)-dimensional space, i. e., at least for certain classes of attractors such a transformation (having a form of embedding) exists (Whitney 1936). This result however doesn’t answer the question of how this embedding can be constructed. The answer to this question was given by F. Takens in his embedding theorem (Takens 1981). It states that such an embedding can be constructed in a form of a vector-function containing \((2m + 1)\) appropriately chosen consecutive discrete measurements of the attractor process. Takens embedding theorem gave a solid theoretical background to the method of delays first proposed by Packard et. al. (Packard et al. 1980). The basic idea of the method is to consider the frames or windows of a certain length \(q\) for a discrete observation series \(z_t \in \mathbb{R}^c\) resulting in a new extended observation series \(x_t \in \mathbb{R}^n, n = qc\).

The practical application of Takens embedding theorem is still limited since it gives no answer to the questions of how to determine the attractor dimensionality \(m\) and how to extract the attractor manifolds from the high-dimensional observation data. However, the fact that there exists a Euclidean embedding of some dimensionality \(n\) gave Broomhead et. al. the idea of combining the method of delays with techniques of statistical data analysis (Broomhead and King 1986; Broomhead et al. 1988).

Let us assume that we were able to estimate the upper bound \(n\) of the embedding dimension for the given time series \(\{z_t\}_{t=1,...,T}\). The idea of the method proposed in (Broomhead and King 1986) is to identify the attractor dimension \(m\) by finding the principal directions with the data variance exceeding a certain threshold (typically given by the machine tolerance (Broomhead et al. 1988)) in \(n\)-dimensional data \(x_t\) \((m \ll n)\). Therefore one can describe the resulting numerical strategy as a combination of Takens “method of delays” with the principal component analysis of the data in extended space. However, the application of PCA implicitly relies on the assumption that the attractor can be globally approximated by a linear manifold. The quality of this approximation (and therefore also the validity of the aforementioned assumption) remains beyond consideration. In the next chapter we will present a framework that allows to overcome this difficulty introducing the concept of states.
hidden in the data and replacing the global linear approximation of the attractor by local linear approximations specific for each of the hidden states.

Topological identification of hidden states

Residual functional with hidden states

The basic idea of state-specific topological dimension reduction consists of the assumption that there is a hidden process switching between attractor manifolds defined with the help of a sequence of $K$ linear transformation matrices $M_i \in \mathbb{R}^{n \times m}$, $T_i \in \mathbb{R}^{m \times n}$, $i = 1, \ldots, K$, where $T_i$ is understood to project onto the subspace spanned by the local principal directions and $M_i$ is a linear transformation casting the reduced vector back into the original space. Mathematically the problem of identifying $M_i$, $T_i$ can be stated as a minimization problem wrt. the residual–functional, describing the least–squares difference between the original observation and its reconstruction from the $m$–dimensional projection:

$$L(x_t, \theta, \gamma) = \sum_{i=1}^{K} \sum_{t=1}^{T} \gamma_i(t) \left\| (x_t - \mu_i) - M_i T_i (x_t - \mu_i) \right\|^2_2,$$

(2)

with $\theta = (T_1, M_1, \mu_1, \ldots, T_K, M_K, \mu_K)$, $\mu_i \in \mathbb{R}^n$ and $\gamma_i(t)$ (which we will call the hidden path) denotes the probability to optimally describe the $n$–dimensional vector $x_t$ at time $t$ with the local transformations $M_i$, $T_i$ ($\sum_{i=1}^{K} \gamma_i(t) = 1$ for all $t$). The quantity $\gamma_i(t)$ provides a relative weight to the statement that an observation $x_t$ belongs to the $i$th hidden state.

We will first consider the case that the hidden path $\gamma$ is known and fixed. In the case of the full optimization of (2) the problem can become ill-posed in general case since the number of unknown parameters can approach the number of data–points $x_t$ and optimized functional is not convex. To solve this problem we suggest two implicit regularization techniques imposed in the form of additional assumptions on the optimization strategy. As it was shown in (Horenko et al. 2006b; Horenko and Schuette 2007), the formulation of the full optimization strategy (this means w.r.t both $\theta$ and $\gamma$ ) can be achieved applying (i) the Hidden Markov Models (with few hidden states and Gaussian output in the
essential coordinates) and the Expectation-Maximization (EM) framework, or (ii) a wavelets representation of the hidden path $\gamma$ (with few wavelets coefficients involved). In both cases the essential step is the calculation of the re-estimation formulas resulting from setting the partial derivatives $\frac{\partial L}{\partial \theta}$ to zero for a fixed sequence of hidden probabilities $\gamma_i(t)$ and number $K$ of hidden states.

To guarantee the uniqueness of the resulting parameters $\theta = \text{argmin} L$ we have to impose certain constraints on the transformation matrices $M_i, T_i$. Depending on the requirements imposed on the resulting attractor dimensions there are various possible matrix structures to be imposed. In the following we consider two special forms of constraints on the projection operators and discuss the resulting numerical strategies in both cases.

**Case 1: State-specific Principal Component Analysis (PCA)** If we set $M_i = T_i^T$ and furthermore assume that the $T_i$ are orthonormal linear projectors, i.e.,

$$T_i T_i^T = \text{Id}_{m \times m},$$  \hspace{1cm} (3)

the Lagrange principle can be used and the explicit solution of the minimization problem can be found analytically if the sequence of $\gamma_i(t)$ is known and fixed (Horenko et al. 2006b; Horenko and Schuette 2007):

$$\left( \sum_{t=1}^{T} \gamma_i(t)(x_t - \mu_i)(x_t - \mu_i)^T \right) T_i = T_i \Lambda_i,$$  \hspace{1cm} (4)

$$\mu_i = \frac{\sum_{t=1}^{T} \gamma_i(t)x_t}{\sum_{t=1}^{T} \gamma_i(t)},$$  \hspace{1cm} (5)

where $\Lambda_i$ is a matrix with the $m$ dominant eigenvalues of the weighted covariance matrix $\sum_{t=1}^{T} \gamma_i(t)(x_t - \mu_i)(x_t - \mu_i)^T$ on the diagonal (non-diagonal elements are zero), i.e., each of the $K$ hidden states is characterized by a specific set of essential dimensions $T_i$ (which can be defined as corresponding dominant eigenvectors) and center vectors $\mu_i \in \mathbb{R}^n$ calculated from the conditional averaging of the time series wrt. corresponding occupation probabilities $\gamma_i(t)$ (Horenko et al. 2006b). Note that for $K = 1$ we automatically get a standard PCA procedure.

For the unknown sequence of $\gamma_i(t)$ two optimization strategies were proposed: (i) HMM-PCA (Horenko et al. 2006b) and (ii) Wavelets-PCA (Horenko and Schuette 2007).
A. Majda and co-workers have recently demonstrated the application of the HMM-Gaussian framework and EM-optimization in the atmospheric context (Majda et al. 2006; Franzke et al. 2007). They verified the power of the HMM–approach wrt. identification of the blocking events hidden in some output time series of several atmospheric models. They showed that the identification of the hidden states was possible even despite of the nearly Gaussian statistical distribution of the corresponding probability density function (because of this fact the traditional clustering methods usually fail to reveal the hidden states). In the following we will shortly describe a different HMM–technique originally developed in the context of molecular dynamics data analysis (Horenko et al. 2006b).

Let us assume that the knowledge about the analyzed data allows the two following assumptions: (1) the unknown vector of hidden probabilities $\gamma_i(t)$ can be assumed to be an output of the Markov process $X_t$ with $K$ states and (2) the probability distribution $P(T_i|x_t|X_t=i)$ (which is the conditional probability distribution of the projected data in the hidden state $i$) can be assumed to be Gaussian in each of the hidden states. If both of these assumptions hold then the HMM-framework can be used and one can construct a special form of EM-algorithm to find the minimum of the residual-functional (2) (for details of the derivation and the resulting algorithmic procedure we refer to our previous works (Horenko et al. 2006b; Horenko and Schuette 2007)). The resulting method is linear in $T$, scales as $O(mn^2)$ with the dimension of the problem and as $O(K^2)$ with the number $K$ of hidden states. However, as all of the likelihood-based methods in a HMM-setting, HMM-PCA does not guarantee the uniqueness of the optimum since the EM-algorithm converges towards a local optimum of the likelihood-function.

If the knowledge about the data is not sufficient to make the above assumptions, we can constrain the functions $\gamma_i(t)$ to the class of piecewise constant functions switching between the discrete values of 0 and 1 and represent them as a finite linear combination of $p$ Haar-Wavelets. As it was demonstrated in (Horenko and Schuette 2007), this allows to reduce the dimensionality of the parameter space necessary to minimize the residual–functional and to construct an iterative optimization scheme called Wavelets-PCA. The main advantage of this approach is that, in contrast to HMM-PCA, we do not need any a priori assumptions about the the data being Markov or Gaussian (since the resulting numerical scheme is a direct optimization of the functional (2); for de-
tails see (Horenko and Schuette 2007)). However, the major drawback of Wavelets-PCA is its quadratic scaling with the number \( p \) of Haar-wavelet functions. This constrains the applicability of the method to the cases where only few (in practice 10-20) transitions between essential manifolds are present in the data. Therefore in the case of a very long data series one can verify the assumptions needed to use the HMM-PCA via a comparison of both methods for short fragments of the series. If the resulting functions \( \gamma_i(t) \) and associated manifolds \( T_i \) are equal or very similar for both of the methods one can assume that the Markovianity and local Gaussianity assumptions are valid and the whole data series can be analyzed with the help of HMM-PCA.

Note that in contrast to the methodology presented in (Horenko et al. 2006b; Horenko and Schuette 2007) (where the observation data was directly analyzed with HMM-PCA and Wavelets-PCA approaches), the construction of the embedding and casting the data into extended space allows to combine both methods with Takens ”method of delays” (Takens 1981). The resulting approaches therefore circumvent the problem connected with the above mentioned global linearity assumption of the method proposed by Broomhead and King and allow for the adaptive construction of local linear attractors associated with each of the hidden states. Shortly speaking, the ”global attractor linearity” assumption of the method proposed in (Broomhead and King 1986) is replaced by a set of ”local linearity assumptions” in HMM-PCA and Wavelets-PCA methods. An additional advantage of the proposed strategy compared to the HMM-PCA-procedure as described in (Horenko et al. 2006b) consists in the fact that, as was mentioned above, casting the observed (possibly non-Markovian) data in an extended space of sufficiently high dimensionality makes the resulting data Markovian and thereby verifies one of the basic assumptions needed for HMM-PCA-method. This issue is closely related to the choice of the frame length \( q \) and will be discussed in detail at a later stage.

It is important to mention that the applicability of the above strategy is limited to cases were the essential attractor manifolds of different hidden states are unequal and can be identified as directions of dominant spatial variability in extended space. This problem is closely connected to the problem of the optimal choice of local reduced dimensionality \( m \). For observation data where the differences between the attractors are not significant for small \( m \) (corresponding to hidden states with almost identical \( m \) dominant local PCA modes) the value of \( m \) should be increased.
and the optimization procedure should be repeated. Besides the obvious numerical problems associated with the calculation of the $m$ dominant eigenvectors of the covariance matrix for large $m$, there is a natural upper limit of $m$ known from information theory which is related to the statistical uncertainty of identified eigenvectors of the covariance matrix and the signal/noise ratio of the data (Broomhead et al. 1988; Lisi 1996; Gilmore 1998). All of the above considerations restrict the applicability of the described framework for very high $n$ to the cases where a low $m$ is sufficient for the identification of the differences in local attractor manifolds.

The Gaussianity assumption for the observation process in the HMM–PCA–method gives an opportunity to estimate the confidence intervals of the manifold parameters $(\mu_i, T_i)$ straightforwardly. This can be done in a standard way of multivariate statistical analysis since the variability of the weighted covariance matrices (4) involved in the calculation of the optimal projectors $T_i$ is given by the Wishart distribution (Mardia et al. 1979). The confidence intervals of the $T_i$ can be estimated by sampling from this distribution and calculating the $m$ dominant eigenvectors of the sampled matrices.

Case 2: State-specific Principal Original Components (POC) In many cases, the knowledge about essential degrees of freedom of the observed system can help to plan the optimal measurement process, i.e., the measurement of the (few) essential degrees of freedom can be used to reconstruct the complete information about the system. However, the local PCA modes $T_i$ described above define the essential degrees of freedom as linear combinations of all observation dimensions with coefficients defined by the elements of the matrices $T_i$. It means that if all of these elements are significantly non-zero, one has to measure all of the original system’s dimensions in order to get a time series of essential coordinates. If the process of measurement is expensive one is typically interested in reducing the number of measurements. Because of that it is sometimes important to express the essential degrees of freedom in terms of as few observed system dimensions as possible. In order to be able to approach this problem mathematically, we can first fix the reduced dimensionality $m$ and hidden path $\gamma_i(t)$ and minimize the functional (2) subject to the special form of constraints imposed on the projector operator $T_i$. To satisfy the requirement that only $m$ different original dimensions of the observation process can be used in the
construction of the reduced dynamics, we have to demand that the respective elements of the \( m \times n \) matrices \( T_i \) can be either 0 or 1 and that there can not be more than one 1 per matrix column. We do not impose any explicit constraints on \( M_i \).

As a first approach to solve the optimization problem (2) subject to the aforementioned constraints, we suggest to go through all of the \( C_{n,m} \) possible projectors \( T_i \) (assuming that \( m \ll n \) so the number of possibilities is not too large). For each fixed parameter \( T_i \) we can calculate the explicit optimum of the functional (2) wrt. \( M_i \) by setting the respective partial derivative to 0 and solving the resulting matrix equation. It is easy to verify that this matrix equation can be solved analytically leading to the following expression for the optimal value of \( M_i \):

\[
M_i = \operatorname{Cov}_i T_i^T \left( T_i \operatorname{Cov}_i T_i^T \right)^{-1},
\]

where \( \operatorname{Cov}_i = \sum_{t=1}^T \gamma_i(t)(x_t - \mu_i)(x_t - \mu_i)^T \). The calculation of the optimal value for \( \mu_i \) is analogous to the PCA case and results in the same estimation formula (5). The set of parameters \( (T_i, M_i, \mu_i) \) calculated in such a way can then be substituted into the functional (2). It is obvious that one of the \( C_{n,m} \) parameter sets with the lowest value of \( L \) defines an optimal attractive manifold in POC-sense.

**Pipeline of data compression and reconstruction**

The solution of optimization problem (2) for a given time series provides the optimal hidden path and attractive manifold parameters \( (\mu_i, T_i, M_i) \). This information can be used to compress the original data \( z \), i.e., to find a low-dimensional description of the dynamics. The following diagram represents the pipeline of the data-compression process based on dimension reduction:

\[
z = \{ z_1, z_2, z_3, \ldots \} \quad \text{Embedding} \quad x = \begin{pmatrix} z_1 & z_2 & z_3 & \ldots \\ z_2 & z_3 & z_4 & \ldots \\ z_3 & z_4 & z_5 & \ldots \\ \vdots & \vdots & \vdots & \ddots \\ z_q & z_{q+1} & z_{q+2} & \ldots \end{pmatrix}
\]

\[
x_r(t) = \sum_{i=1}^K \gamma_i(t) \left( M_i x_p^i(t) + \mu_i \right) \quad \text{Projection} \quad x_p^i(t) = T_i x(t)
\]
As the first step, the frame length \( q \) is selected and the multidimensional process \( z \) is embedded in the extended space. As it follows from the Takens' theorem (Takens 1981), the frame length \( q \) should be chosen such that \( q \geq \frac{1}{c} (2m + 1) \), where \( m \) is estimated attractor dimension and \( c \) is the spatial dimension of the observation data \( z \). If the attractor dimension \( m \) is a priori unknown, the choice of \( q \) is bounded from above by the feasibility of computing the dominant eigenvectors of the \((n \times n)\) covariance matrix (where \( n = qc \)). Another limitation comes into play when analyzing the observation data with some memory \( \text{mem} > 1 \) (we define Markov-processes as those with \( \text{mem} = 1 \)). In this case, the frame length should be chosen such that \( q \geq \text{mem} \) in order to guarantee the Markov-property for the embedded data \( x \) and to fulfill the formal criteria of the Takens' theorem.

The embedding procedure results in a new time series \( x \), whose elements are the dynamical patterns of length \( q \) from the original time series \( z \). As the next step, the number of hidden states \( K \) and desired reduced dimensionality \( m \) should be selected (for example by inspection of the spectra of the resulting local covariance matrices) and functional (2) should be minimized numerically with one of the procedures described above. Afterwards, the \( m \)-dimensional projections \( x_p^i \) of the time series \( z \) on the respective manifolds can be calculated. The compression factor associated with this dimension reduction strategy is \( n(T - D)/(Kn(m + 1)) \) and converges to \( n/(Km) \) if \( T \to \infty \). The projected time series \( x_p^i \) can be projected back (or reconstructed) applying the projector operators \( M_i, i = 1, \ldots, K \). This operation delivers the reconstructed time series from the dynamical frames \( x_r \) in the extended space. Finally, the reconstruction \( z_r \) of the original time series \( z \) can be achieved via:

\[
x_r = \left\{ \begin{array}{cccc}
  z_1^{(1)} & z_2^{(2)} & z_3^{(3)} & \cdots \\
  z_2^{(1)} & z_3^{(2)} & z_4^{(3)} & \cdots \\
  z_3^{(1)} & z_4^{(2)} & z_5^{(3)} & \cdots \\
  \vdots & \vdots & \vdots & \ddots \\
  z_q^{(1)} & z_{q+1}^{(2)} & z_{q+2}^{(3)} & \cdots 
\end{array} \right\}
\]

\[
\text{ReconstructionII} \quad \to \quad z_r = \left\{ \begin{array}{c}
  z_1^{(1)}, \quad \frac{1}{2} (z_2^{(2)} + z_2^{(1)}), \quad \ldots 
\end{array} \right\}
\]

The last step can be formally written as

\[
z_r^{(t)} = \frac{1}{\min\{q,t\}} \sum_{j=0}^{\min\{q-1,t-1\}} z_t^{(t-j)},
\]
i. e., computing $z_r$ we calculate the expectation value over all of the re-
constructed dynamical patterns contained in $x_r$. As will be demonstrated
in numerical examples, this leads to the smoothing or filtration of the re-
constructed dynamics and will allow us to recover the attractor structures
hidden by the external noise.

Numerical examples

In the following we will illustrate the proposed strategy for dimension
reduction, dominant temporal pattern recognition, and decomposition
into metastable states by three examples: (1) a time series of the Lorenz
oscillator in 3 dimensions with the multivariate autoregressive process
(MVAR) added to it to mimic the noisy measurement process with mem-
ory, (2) a one–dimensional data set of historical daily air temperatures in
Berlin\(^1\) and (3) a set of historical averaged daily temperatures between
1976 and 2002 on a $20 \times 29$ spatial grid covering Europe and part of the
north Atlantic (for details concerning the data see (Horenko et al. Sept.
2007)).

Example (1) represents a toy model aiming to illustrate the proposed
framework on a simple physical system. Identified hidden states and
attractor manifolds can be straightforwardly interpreted according to our
knowledge about the process.

In the next Example (2) we demonstrate the power of topological pat-
ttern recognition and identification of hidden states associated with those
dynamic patterns. In contrast to all other methods for identification of
hidden states in one–dimensional data we are aware of (Hamilton 1989;
Horenko et al. 2006a, 2005), the proposed method allows assumption
free identification of the hidden states based on differences in topology
of corresponding dynamical patterns and simultaneous state–specific
filtering of the data. We will show how this property of the framework can
help to eliminate the low-frequent seasonal trend from the meteorologi-
cal data and compare the results with filtering obtained applying some
standard methods of seasonal cycle elimination.

Finally, in Example (3) we show the application of the $POC$–strategy
for dimension reduction of high–dimensional meteorological data. The
question to be answered in this context is about finding a (small) subset
of original measurement dimensions which is optimal wrt. the recon-

\(^1\)The data was kindly provided by H. Osterle and R. Klein from the Potsdam Institute of Climate Research
(PIK).
struction of the temporal data-patterns for the rest of the measurements. In the presented example this subset is represented by points on the geographic map where the temperature measurements are acquired. As is demonstrated in Example (3), the original measurement dimensions that are optimal wrt. the temporal pattern recognition can be used to construct global stochastic models for temperature prediction based on some additional assumptions about the dynamics in the extended space. We verify those assumptions for the analyzed time series and demonstrate the application of multivariate autoregressive models (MVAR) in this context. Finally, we analyze the performance of a resulting prediction strategy wrt. the number $m$ of original measurement dimensions involved in the construction of the predictive stochastic model.

(1) Lorenz system with measurement noise

As the first application example for the proposed framework we choose a well-known Lorenz model (Lorenz 1963):

\[
\begin{align*}
\frac{dx}{dt} &= -8/3x + yz, \\
\frac{dy}{dt} &= -10y + 10z, \\
\frac{dz}{dt} &= -xy + 28y - z.
\end{align*}
\]

(10)

We generate a trajectory starting with initial value $(x(0), y(0), z(0)) = (35, -10, -7)$ using the implicit Euler scheme with discretization time step $h = 0.01$ on a time interval $[0, 20]$. To the resulting time series we add an output of a three–dimensional MVAR(3) process to mimic a measurement noise with memory (Brockwell and Davis 2002). The resulting three–dimensional data series is shown in the upper panel of Fig. 1. Added noise is chosen to be quite intense so that the original smooth “butterfly” structure of the respective attractor is smeared out.

To analyze the resulting time series we first apply the standard PCA strategy to the extended space representation of the dynamics (with frame length $q = 12$) as described in (Broomhead and King 1986; Broomhead et al. 1988). In order to estimate the number $m$ of essential dimensions we first have a look at the logarithmic plot of the eigenvalues of the corresponding global covariance matrix. The upper panel of Fig. 2 shows that there is no clear spectral gap indicating the presence of dominant
Figure 1: Output of the Lorenz oscillator with measurement noise added (as MVAR(3)-process) (upper panel) and its reconstruction from the two–dimensional reduced HMM-PCA dynamics ($m = 2, K = 2, q = 12$) (lower panel).
Figure 2: Dominant part of a global covariance matrix spectrum (standard PCA) for the time series from the left panel of Fig. 1 (frame length $q = 12$) (upper panel) and dominant parts of local covariance matrix spectra in both of the identified hidden states (lower panel). The error bars denote the confidence intervals for the estimated eigenvalues. The spectral gap on the left panel indicates the presence of two–dimensional attractive manifolds for both of the hidden states (resulting in effective $m = 2$).
Figure 3: Comparison of the hidden paths calculated with Wavelet-PCA (dotted) and HMM-PCA (dashed) methods ($m = 12$, $K = 2$, $q = 12$).

...low–dimensional manifolds in the extended 36–dimensional space. This is a hint that the standard PCA dimension reduction strategy may not succeed in finding an adequate reduced representation of the dynamics (for $m < 3$). In fact, as it can be seen from the left panel of Fig. 4, there is a significant discrepancy between the original Lorenz trajectory and its reconstruction from a reduced representation of the noisy data ($m = 2$) based on the standard PCA approach.

Application of both HMM-PCA and Wavelets-PCA approaches to the same noisy data sequence from Fig. 1 (with $K = 2$, $m = 12$, $q = 12$) results in almost identical hidden paths (see Fig. 3). As can be seen from the lower panel of Fig. 2, the spectra of the local covariance matrices in each of the identified states show a presence of a spectral gap at $m = 2$. The lower panel of Fig. 1 shows the reconstruction of the three–dimensional dynamics from its reduced two–dimensional HMM-PCA-representation revealing the original "butterfly" structure of the Lorenz-attractor hidden in the noisy data. As can be seen from the comparison of both standard PCA and HMM-PCA reconstructions, the latter demonstrates a much higher resemblance to the original data. The lower panel of Fig. 1 also gives an explanation of the identified states: they correspond to the wings of the Lorenz-attractor "butterfly".
Figure 4: Comparison of the original Lorenz signal (without addition of noise, solid line) and noisy signal reconstructions from the reduced two-dimensional PCA-trajectory ($m = 2, q = 12$, crosses, upper panel) and from reduced HMM-PCA trajectory ($m = 2, K = 2, q = 12$, crosses for hidden state 1 and circles for hidden state 2, lower panel).
Figure 5: The Fourier-spectrum of the original temperature data (upper panel) and comparison of the Fourier-spectra for filtered data obtained applying different approaches of seasonal cycle elimination (lower panel): Fourier-filtering (dashed line, crosses), direct elimination of 1-year-period (dotted line, circles; explanation in text), singular-spectrum analysis (SSA) as described in (Vautard et al. 1992) (dash-dotted line) and Wavelets-PCA filtering for $m = 1$, $K = 2$, $q = 50$ (solid line).
Figure 6: Dominant part of the global covariance matrix spectrum (standard PCA) for the analyzed air temperatures in Berlin (frame length $q = 50$) (upper panel) and dominant parts of local covariance matrices spectra in both of the identified hidden states (lower panel). The error bars denote the confidence intervals for the estimated eigenvalues. The spectral gap on the left panel indicates the presence of one–dimensional attractive manifolds for both of the hidden states (resulting in effective $m = 1$).
Figure 7: Upper panel: Parameters $\mu_i, i = 1, 2$ (solid) together with intervals of standard deviation in extended space (dotted). Lower panel: the respective dominant dimensions (components of the optimal projector $T_i, m = 1$ as explained in text) for both identified states together with their confidence intervals. Identified by the Wavelets-PCA for $m = 1, K = 2, q = 50$. 
Figure 8: Upper panel: reduced one-dimensional Wavelets-PCA representation of the original time series for $m = 1, K = 2, q = 50$, the discontinuity of the representation is due to the discrete representation of the hidden path in the Wavelets-PCA approach. Lower panel: comparison of original time series with the reconstruction from the Wavelets-PCA reduced representation.
(2) One–dimensional example: historical temperatures in Berlin.
As a first realistic example for the proposed analysis strategy we take the historical daily air temperature measurements as averaged values from 00:00h, 06:00h, 12:00h, and 18:00h observations in Berlin between 1. Jan. 1976 and 26. Sep. 1978 (resulting in a time series of 1000 measurements).

In contrast to the previous example, the analyzed time series is one–dimensional and there is no obvious metastability in the data, i. e. the dynamics is not obviously switching between geometrically well–separated domains of the phase space as it was in the case of the Lorenz–oscillator. However, we can hope that the extended space representation of the dynamics will help us to reveal the hidden dynamical patterns in the data and to use them in identification of the hidden metastability.

Both the visual examination and the Fourier-analysis reveal a presence of the seasonal 1-year cycle in the considered data (see the top panel of Fig. 5). As the first step of the data-analysis, we apply and compare the three following strategies of seasonal cycle elimination (see the lower panel of Fig. 5): (i) direct elimination of the seasonal component with period $P = 1.0$ year from the original temperature series $\{z_t\}_{t=1,...,T}$ by a subtraction of the $P$-averaged temperature

$$\tilde{z}_t = z_t - \frac{1}{\| I(t) \|} \sum_{k \in I(t)} z_{t+kP}$$

(11)

where $I(t) = \{ k \in \mathbb{N} | 1 \leq (t + kP) \leq T \}$, (ii) Fourier-filtering of the signal $\{z_t\}_{t=1,...,T}$, i. e., elimination of the Fourier-modes exceeding beyond a threshold amplitude, and (iii) a singular-spectrum analysis (SSA) as described in (Vautard et al. 1992) for $m = 1, q = 50$. As it can be seen from the lower panel of Fig. 5, none of the above mentioned methods (i)-(iii) is able to eliminate the seasonal trend completely from the data, i. e., there remains a significant low-frequency periodicity in the data. This observation can be explained by the violation of the implicit assumptions of either approach: method (i) implies a presence of the exactly periodic component with $P = 1.0$ year in the analyzed data and stationarity of the discretized process $z_{t+kP}, \forall t \in [1, T], k \in I(t)$, method (ii) is based on application of periodic sine and cosine functions for the decomposition of the signal (which means that one should involve indefinitely many of them for certain classes of periodic signals), stationarity of the period is also implicitly assumed in (ii), and finally (iii) assumes existence of a single globally linear attractive manifold as
we have already discussed above. None of those assumptions is involved in the minimization of the functional (2) by means of Wavelets-PCA. As we will demonstrate later, this leads to a reliable extraction of the low-frequent modes from the analyzed data.

We start by applying the standard PCA-strategy and the Wavelets-PCA (with $K = 2, m = 49, q = 50$) to the extended space representation of the dynamics in the same way as it was described in the previous example. The upper panel of Fig. 6 demonstrates that there is no clear spectral gap indicating the presence of dominant low dimensional manifolds in extended space for the standard PCA dimension reduction strategy. In contrast, as it can be seen from the right panel of Fig. 6, Wavelets–PCA exhibits a clear spectral gap after the first eigenvalue for both of the hidden states and therefore indicates the presence of two one-dimensional attractive manifolds ($m = 1$). We repeat the Wavelets-PCA analysis for $K = 2, m = 1, q = 50$ and get the hidden sequence $\gamma_i(t)$ that is identical to the previous case with $m = 49$.

To check the applicability of the HMM-PCA-strategy we can test the conditional Gaussianity (for the 2 hidden states identified by Wavelets-PCA) of the embedded data $x_t$ for different values of the frame length $q \geq 4$ (since the memory of the analyzed data was found to be 4 according to the common test based on application of autoregressive models (Brockwell and Davis 2002)). Applying some standard statistical tests (like $\chi^2$–test, Kolmogorov–Smirnov–test and Shapiro–Wilks–test) to marginal statistical distributions of the extended data discards the Gaussian hypothesis (for the probability of error of type I being $\alpha = 0.05$). This feature indicates the data in the hidden states is non-Gaussian, therefore prohibiting application of HMM-PCA.

In order to interpret the results of the Wavelets-PCA-analysis and to understand the physical meaning of the identified states, we first look at the estimated manifold parameters $(\mu_i, T_i)$ for each of the states (see Fig. 7). The lower panel of Fig. 7 demonstrates that the identified optimal projectors $T_i, i = 1, 2$, are significantly different for both of the states. Parameters $\mu_i, i = 1, 2$ shown in the top panel of Fig. 7 can be understood as the mean dynamical patterns of length $q$ characteristic for the identified states. This means that the first hidden state is characterized by the almost linear mean dynamical pattern with upwards trend of the temperature dynamics (daily temperature difference of $0.12 \pm 0.018$), whereas the second hidden state is obviously characterized by the almost linear mean dynamical pattern with downwards trend of the temperature
dynamics (daily temperature difference of \(-0.12 \pm 0.02\)). It is important to mention that this obvious linearity of the hidden trends resulting from the Wavelets-PCA procedure is not a result of the assumptions done implicitly in the analysis procedure. The only assumption needed for the Wavelets-PCA dimension reduction strategy is discreteness of the underlying hidden path (Horenko and Schuette 2007).

In order to make the last point clear, we look at the one–dimensional reduced representation of the extended space dynamics based on the Wavelets-PCA strategy \((K = 2, m = 1, q = 50)\) (see the top panel of Fig. 8). Reduced dynamics exhibits the “sawtooth” form resulting from the consecutive linear upwards and downwards trends. The discontinuity of the reduced dynamics is explained by the jumps between the local attractive manifolds and is a direct consequence of the discreteness assumption involved in the Wavelets-PCA procedure. Finally, the bottom panel of Fig. 8 shows the comparison of the original time series with its reconstruction from the reduced Wavelets-PCA representation. Analogously to the previous example, the reconstructed series is smooth and clearly reveals the dynamical behavior associated with each of the hidden attractive manifolds. As we see on the bottom panel of Fig. 5, in contrast to all other standard filtering methods we have used, Wavelets-PCA-based identification of attractive seasonal trend components

After subtraction of the identified linear attractive seasonal trend components from the analyzed data we can check the applicability of local linear autoregressive models (AR) for the reduced data \(x_r(t) \in \mathbb{R}^1\) (local in this context means that for each of the hidden states we can test for applicability of a certain specific AR-model used to describe the filtered data locally in each state) (Brockwell and Davis 2002). In the next example, this analysis will help us to construct predictive models based on a Wavelets-PCA decomposition of data.

(3) Multidimensional example: historical temperatures on a grid.
In order to demonstrate the application of the presented framework on realistic multidimensional data with memory we choose daily mean values of the 2 meter surface air temperature from the ERA 40 reanalysis data (Simmons and Gibson 2000). We consider a region with the coordinates: 27.0 W – 45.5 E and 33.0 N – 73.5 N, which includes Europe and a part of the Eastern North Atlantic. The original spatial resolution of the data was reduced to approximately 2.0° × 2.5° latitude and longitude by spatial averaging. Thus we have temperature values on a grid of 20 × 29
points (resulting in $c = 580$). The time record is from 1976 till 2002 and it includes 9736 daily averaged values from 00:00h, 06:00h, 12:00h, and 18:00h observations.

Due to the fact that both HMM–PCA and Wavelets–PCA scale as $O(mn^2)$ wrt. the extended dimension $n = qc$ and reduced dimensionality $m$, our specific implementation of the presented dimension reduction strategy is restricted to the cases when $n$ is not too big (no more than $1000 - 1500$). This restriction is only due to the current implementation of the code and is not intrinsic to the presented strategy itself. Because of this restriction we cannot choose the frame length $q$ arbitrarily large since the extended space dimensionality $n$ is a product of the observation process dimension $c$ and the frame length $q$.

Application of the Wavelets-PCA strategy for $K = 2, q = 2, m = 1$ results in identification of the hidden path shown in Fig. 9. The figure shows that the hidden path is almost identical to the one calculated from the one-dimensional Berlin temperature data in the previous example. This indicates that the hidden process identified from the current multidimensional temperature data process is the same as in the example above and can be understood as a consequence of upwards and downwards trends in the dynamics. Inspection of the mean dynamical pat-
Figure 10: Dots on the Europe map represent the optimal subgrids for the 1 day temperature predictions on a uniform $20 \times 29$-grid based on MVAR-dynamics with mean average prediction error $< 2.5$ C. Lower panel: first hidden state ("winter-spring"). Upper panel: second hidden state ("summer-autumn").
terns associated with those trends reveals the linearity of the underlying dynamical trends. As in the example above, standard tests for linear autoregressive behavior of the filtered data can be applied (i.e., for the data with subtracted attractive trend components) and application of localized multivariate autoregressive processes (MVAR) for the construction of predictive models can be motivated (Brockwell and Davis 2002).

In contrast to the PCA-type of the dimension reduction (where the essential manifolds are constructed as linear combinations of all of the original degrees of freedom), the POC-approach identifies those manifolds from subsets of the original system dimensions. In the context of meteorological data analysis this approach has an important advantage compared to PCA since one needs only partial observation of the process to calculate the reduced representation of the overall system dynamics, i.e., it is enough to measure only few of the original process dimensions to reconstruct the dynamics in all of the remaining degrees of freedom and to make predictions.

To identify the essential original dimensions in the analyzed temperature time series in each of the hidden states, we solve the POC-optimization problems (2) for the increasing number $m$ of original dimensions involved and for the fixed hidden path $\gamma_i(t)$ as resulting from the Wavelets-PCA method (with $K = 2, q = 2, m = 1$). The time series

![Figure 11: Comparison of mean prediction errors as functions of the size $m$ of the optimal subgrids for both hidden states.](image-url)
data from the resulting POC subgrids are then used in identification of MVAR(2) model parameters in each of the hidden states separately. The identification of the MVAR(2) model parameters is done by applying the standard regression method as described in (Brockwell and Davis 2002). The one–day predictions based on those models are then used to reconstruct the full observation data on the $(20 \times 29)$ grid (applying the identified POC projectors $M_i, T_i$) and the mean prediction errors are calculated for different subgrids. As it can be seen from the Fig. 10, in order to achieve the mean one–day prediction error of 2.5C for the whole $(20 \times 29)$-grid it is enough to measure the temperature only on 4 grid points in winter and spring and on 9 grid points in summer and autumn. Fig. 11 shows the comparison of mean prediction errors based on hidden state-specific MVAR(2) models as functions of the number $m$ of original systems dimensions involved. As expected, the mean prediction error is monotonously decreasing for both states when the number $m$ is increasing.

Discussion

A numerical framework for the simultaneous identification of hidden states and essential attractor manifolds in high-dimensional data is presented. The idea of the method is a combination of the method of delays (based on Takens’ embedding theorem) with the minimization of a functional describing the Euclidean distance between original and reconstructed data (both in embedded representation). The extended space representation results in Euclidean embedding of the underlying dynamical process and verifies the application of the Euclidean residual-functional (2). Minimization of this functional results in approximation of the attractive manifolds by $K$ linear hyperplanes.

We have demonstrated how the solution of the general form of the optimization problem (2) can be approached numerically in two specific cases: (i) if the essential coordinates are linear combinations of all original degrees of freedom resulting in Wavelets–PCA and HMM–PCA approaches, (ii) if the essential coordinates are chosen as linear combinations of few optimally chosen original degrees of freedom resulting in a method of POC. The subtraction of the mean dynamical patterns associated with each of the hidden states allowed us to eliminate the low-frequent seasonal components much better then it was possible with other standard methods we applied (see right panel of Fig. 5). We
verified the application of a certain class of models for the filtered temperature series analysis and predictions, namely the linear multivariate autoregressive models (MVAR).

In the context of temperature data analysis, we have demonstrated that the hidden process switching between two attractive manifolds is almost exactly the same for one-dimensional temperature series from a single grid point and for the whole multidimensional data on a grid. It is important to mention that this finding was independent of the assumptions that are typically needed for the mean 1-year cycle subtraction (11) and therefore does not presuppose the existence and stationarity of the 1-year temperature cycle. It indicates a presence of the identical hidden low-frequent mode switching between the seasonal trends. This feature should be validated for the different available sets of meteorological data.

The future work will require the overcoming of two technical problems associated with the present version of the method’s realization on PC: (i) \( n \) can not be taken too high (1000-1500 in a present version of the code ) (ii) the number of hidden states \( K \) is chosen \textit{a priori}. While Problem (i) can be solved by the implementation of an efficient variant of the Lanczos eigenvalue-solver, the solution of Problem (ii) will require the incorporation of some theoretical results from the spectral theory of Markov-chains (Schütte and Huisinga 2003).

Another yet unresolved problem is connected with the effective implementation of the POC-optimization procedure, since the current version relies on the direct combinatorial search of optimal subgrids and is not applicable for high values of \( m \).

Acknowledgement

The author wishes to thank Christof Schuette (FU) and Rupert Klein (FU/PIK) for their friendship and constant support. I wish also to thank H. Osterle and S. Dolaptchiev from Potsdam Institute of Climate Research (PIK) for the possibility to use the historical temperature data. The author also thanks the anonymous referee for useful comments.
References


Broomhead, D., R. Jones, and G. King, 1988: Comment on "singular-value decomposition and embedding dimension". *Physical Review A*, 37, 5004.


