Reconstruction of the parameter spaces of dynamical systems

Stefan Güttler,[1] Holger Kantz,[2] and Eckehard Olbrich[3]
[1]Fraunhofer-Institut für Produktionstechnik und Automatisierung, Nobelstrasse 12, 70569 Stuttgart, Germany
[2]Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, D 01187 Dresden, Germany
[3]Physik Institut, Universität Zürich-Irchel, Winterthurerstrasse 190, CH 8057 Zürich, Switzerland
(Received 11 August 2000; published 24 April 2001)

Parameter variations in the equations of motion of dynamical systems are identified by time series analysis. The information contained in time series data is transformed and compressed to feature vectors. The space of feature vectors is an embedding for the unobserved parameters of the system. We show that the smooth variation of d system parameters can lead to paths of feature vectors on smooth d-dimensional manifolds in feature space, provided the latter is high-dimensional enough. The number of varying parameters and the nature of their variation can thus be identified. The method is illustrated using numerically generated data and experimental data from electromotors. Complications arising from bifurcations in deterministic dynamical systems are shown to disappear for slightly noisy systems.

DOI: 10.1103/PhysRevE.63.056215 PACS number(s): 05.45.Tp, 02.50.Tt

I. INTRODUCTION

Time series analysis has become a popular approach to the investigation of dynamical behavior of systems in experiments and field measurements. Methods which are usually called *nonlinear* refer to the reconstruction and exploitation of structure in phase space [1,2], and are very powerful if a time series is generated by an almost deterministic low-dimensional dynamical system. These, like also almost all other time series analysis methods, require a strict stationarity of a time series.

However, in a huge number of dynamical phenomena the variability of the dynamics is much more relevant or interesting than the dynamics itself. One example is the activity of the human heart reflected by electrocardiagram recordings: The electrophysiological mechanism which creates the signature of an individual heartbeat is rather well understood and rather robust, but the heart rate has a large variability, and this variability may tell a lot about the physical condition of the heart as an organ.

Although there are already different well established approaches to dynamical pattern recognition, data classification, and the extraction of dominant modes, another point of view will be taken here. In this paper we explicitly assume that the variability of a given system's dynamics originates from the change of a small set of system parameters. It will be the goal of this paper to gain access to these unobserved and typically unknown parameters through the analysis of a set of different time series reflecting this dynamical variability. We propose to reconstruct the parameter space from the time series data, where *features* of the underlying time series form the elements of the reconstructed parameter vectors. Their changes will reflect the variation of system parameters, e.g., when modifying experimental conditions or in a nonstationary setting. This idea implies that the fast time scale associated with the dynamics of the system itself is eliminated by a kind of projection, and changes of system parameters on larger time scales thus gain enhanced observability. We will comment on the *a posteriori* distinction between dynamical variables and parameters later, in particular in the case of a nonstationary setting, where the parameters themselves are also time dependent.

We will consider two related settings for the reconstruction of parameter spaces. One is the situation of nonstationarity, where we assume that parameters vary as a function of time. Provided that this time dependence is slow enough compared to the time scales of the system's dynamics, we can map this situation by a segmentation of the time series onto the second setting, where we assume to possess a sample of time series from a given system with different constant parameter settings. We will show that in an idealized setting for both situations, the following holds: If d parameters are varied, the set of all conceivable feature vectors is confined to a d-dimensional manifold in the reconstructed parameter space. If the number of variable parameters is unknown, it can be identified as the maximal dimensionality of the set of feature vectors, and, if the variation is time dependent, we can follow the path in the reconstructed parameter space in order to identify the nature of the nonstationarity. Examples will demonstrate that we can also obtain meaningful results in realistic situations.

Potential future applications of this idea range from a better understanding of laboratory experiments, where the constancy of control parameters can thus be checked, over dynamical variability in field measurements, and the understanding of driving forces causing nonstationarity for tasks such as data classification and failure detection. Data classification of a similar type was suggested in Ref. [3], and nonstationarity was traced in a kind of feature space in Ref. [4]. Feature extraction is a well known problem in statistics which can be studied under several aspects such as pattern recognition or signal representation.

The concept of parameter reconstruction will be introduced and the main claim will be formulated in an abstract way in Sec. II. There, relations to established methods of data classification will also be discussed. For the class of linear stochastic models our claim can be proven by referring to well known properties of these systems. In a more general setting, deterministic and stochastic dynamical systems are discussed in Sec. IV, where certain complications can arise, in particular through bifurcations. We illustrate theoretical considerations by numerical examples. In Sec. VI we apply the method to time series data from electromotors, and show that essentially two parameters drive the large dynamic variability contained in 84 different time series. In Sec. VII we compare our approach to a method of parameter identification suggested by Parlitz *et al.* [5].

II. RECONSTRUCTION OF UNOBSERVED SYSTEM PARAMETERS

We shall establish the concept of parameter reconstruction for random dynamical systems with the general form

$$\frac{d\vec{x}}{dt}=\vec{f}(\vec{x}(t),\vec{\xi},\vec{p}) \tag{1}$$

where $\vec{x}\in\Gamma\subset\mathbf{R}^l$ is the state vector, $\vec{p}\in\mathbf{R}^d$ is a parameter vector, and $\vec{\xi}$ represents a noise term. If $\vec{\xi}(t)\equiv 0$ and $\vec{p}=\text{const}$, Eq. (1) defines a deterministic dynamical system. The time series is then a sequence of (most often scalar) observations $\{s_n\}, n=1,\ldots,N$, where $s_n=h(\vec{x}(t=n\Delta t))$, with a measurement function h and a sampling interval Δt. If we allow the parameters $\vec{p}$ to be time dependent, the function $\vec{p}(t)$ is assumed to vary on much larger time scales than $\vec{x}(t)$ or even $\vec{\xi}$ do. However, when the starting points are time series data of an unknown system, there may be some ambiguity which will be discussed in Sec. V. The concept of parameter reconstruction will include two limiting cases. (a) An ensemble of different time series obtained from a solution of the system in Eq. (1) with the same measurement function h and same sampling interval Δt is given, but where each single series corresponds to some individual fixed $\vec{p}$. This situation is the typical starting point for the classification task.

(b) In a single long time series, the parameters $\vec{p}$ are slowly varying in time. This corresponds to the problem of nonstationarity. However, the nonstationary situation can be approximately mapped to (a) by cutting the time series into (potentially overlapping) segments, if the change of the parameters is sufficiently slow. Since in this case the time series segments are still nonstationary, time averages needed to compute features do not represent averages according to an invariant distribution. Hence, for the formal derivation of the method we assume an (unrealistic) case in which, in the nonstationary situation, the parameters vary in a steplike manner with small absolute changes from one time series segment to the successive one, with no changes inside each segment. In this case, (b) is exactly mapped to (a).

Hence the formal derivation and the essential part of the numerical illustrations will focus on the classification problem for generality, and only in Sec. V will we come back to the non-stationarity. The main idea of this paper is then the following: Whitney's embedding theorem [6] states that any given d-dimensional smooth manifold can be embedded in an $\mathbf{R}^k$ where $k\geqslant 2d+1$. Now let this d-dimensional manifold be the parameter space of a system, i.e., equations of motion like Eq. (1) depend on d parameters as stated above. Following Whitney's theorem, we thus need $k\geqslant 2d+1$ independent smooth representations of these d parameters in order to reconstruct the parameter space in an $\mathbf{R}^k$. A scalar time series obtained from a trajectory of the system should thus be converted into k different and mutually independent values: the features. In order to represent the parameters and not the details of the time series, these features should be independent of the initial condition or the particular realization of the time series (invariant under shift in time), but they have to depend on the underlying parameters.

Some definitions are necessary in order to formalize this idea. As stated in Eq. (1), we do not want to distinguish between purely deterministic data and time series from random processes. To be general, we will thus speak about the source of the time series and on how many parameters the source depends. The dependence of the source on the parameters can be such that it is essentially unobservable through the time series (which might depend on the measurement function). We naturally want to exclude these parameters from our discussion, and we thus define the following two influences, which will be the only ones covered by this paper.

Distribution influence: At least one of the moments of the probability distribution of the observable depends on the parameter.

Dynamical influence: At least one of the nth order temporal correlations depends on the parameter.

The *standardizing sample* $\mathcal{S}=\{S_j\}$, $j=1,\ldots,D$, $S_j=\{s_1^j,\ldots,s_{N_j}^j\}$ is an ensemble of D time series which represents different settings of those parameters which are actually varied. Each single time series S_j is assumed to be stationary with constant parameters, and is obtained with the same measurement function and sampling interval.

The *number of effective parameters* is d_e. Since different parameters might have identical effects, or some parameters might be kept constant in the standardizing sample, the number of effective parameters d_e might be smaller than d. This might be trivial (if a parameter a is just split into $a=a_1+a_2$) or nontrivial, and might, e.g., be shown by nontrivial transformations of the equations defining the source.

Finally, we define a *feature f*, a statistical quantity which does not (at least in the limit of infinite sample size) depend on the realization of the process (e.g. initial condition, individual piece of trajectory) but only on the parameters of the source is defined to be a feature. A set of features $\{f_i\}$ is called independent, if there is no function g such that $f_i=g(\{f_j\})$ $i\neq j$ *for all arbitrary time series*.

The standardizing sample will then be compressed into a set of identical cardinality of feature vectors $\{\vec{f}_j\}$, where each feature vector is composed of the values of a set of independent features f_i on the corresponding time series S_j: $(\vec{f}_j)_i=f_i(S_j)$. Based on these definitions we make the following three *propositions*.

(a) When the standardizing sample represents the variation of d_e effective parameters which possess either distribution influence or dynamical influence or both, then there exists a $(k=2d_e+1)$-dimensional embedding space constructed by the direct product of $k=2d_e+1$ suitable independent features, in which the parameter space is uniquely immersed.

(b) The set of feature vectors, representing a standardizing sample of an unknown source and an unknown effective number of parameters d_e, is confined to a $\leqslant d_e$-dimensional manifold in embedding space formed by the direct product of $\geqslant 2d_e$ independent features. The extrapolated dimensionality of this set of feature vectors saturates at a value $\leqslant d_e$, where the identity is found for suitable features and for $k > d_e$.

(c) If the standardizing sample is derived by segmentation from a single long time series, the time ordering of the feature vectors defines a path in feature space. If $k > 2d_e$, this path is topologically equivalent to the path $\vec{p}(t)$ of the time dependent parameters in their parameter space.

A. Remarks

To exploit proposition (b), one has to consider that the finite set of feature vectors obtained from the standardizing sample has a dimension zero. It is assumed to be a finite random sample of a set with an unknown dimension d_e. The dimension d_e has to be extrapolated numerically, e.g., as in numerical dimension estimates of strange attractors, and thus the standardizing sample has to be sufficiently large. An unsuitable choice of features can only hide certain parameters but cannot enhance their number.

The distribution influence can show up in static characteristics such as a mean, standard deviation, or scaling properties such as dimensions, i.e., in all quantities which involve the invariant measure of the deterministic dynamical system or the invariant probability density function (PDF) of the stochastic system. The dynamical influence can express itself by quantities involving time lags, such as power spectrum, entropies, or Lyapunov exponents. The distinction between distribution and dynamical influence is not a sharp one. Distributions in the time delay embedding space also contain essential dynamical information.

For Gaussian distributions with zero mean, the variance and the fourth moment are related by $3\langle x^2\rangle^2 - \langle x^4\rangle = 0$. Nonetheless, second and fourth moments are two independent features since no general relationship holds for samples (time series) with arbitrary PDF's.

There exist parameters which have dynamical influence, without distribution influence and vice versa. A simple example is a series of Gaussian random variables, where the variance of the distribution and the correlations between successive values can be tuned independently through the model [an AR(1) model; see below]

$$x_{n+1} = \frac{c_1}{\sigma^2} x_n + \left(\sigma^2 - \frac{c_1^2}{\sigma^2} \right) \xi_n, \qquad (2)$$

where ξ_n is Gaussian white noise with unit variance, $\sigma^2 = \langle x_n^2 \rangle$ is the variance, and $c_1 = \langle x_n x_{n+1} \rangle$ is the autocorrelation function at lag 1.

The reason why we need the Whitney $2d_e+1$ ''overembedding'' lies in the nonlinear way in which the parameters might show up in the features. By the choice of features one introduces a metric in feature space, and thus determines the nature of the manifold which represents the parameter space. Optimal choices of the features might impose a Euclidean nature on it, so that $k=d_e$ might be sufficient.

The features are quantities (such as the mean value) which are well defined for a given process, but whose values on a finite sample are computed by a statistical estimator. Since the reconstruction task is a comparative task on samples of equal size, we do not need an estimate that is as accurate as possible, so that a potential bias or inconsistency does not matter too much. More relevant for a sharp distinction is the variance of an estimator: Given different finite time series as realizations of the same process with identical parameters, how much does the outcome of our estimate vary? In order to be able to identify parameter changes, these statistical fluctuations should be as small as possible, and we need to employ estimators which have a minimal variance.

We will present a proof of the propositions only for the class of linear stochastic models. For more general dynamical systems, certain complications arise which typically are not destructive from a practical point of view, but might be prohibitive for a more general proof. Since Whitney's theorem requires a smooth dependence of the coordinates of the embedding space on the position in the original manifold, bifurcations in deterministic dynamical systems will turn out to pose problems, which can be circumvented by the inclusion of the dynamical noise $\vec{\xi}$ in Eq. (1). It is thus plausible that a general proof is possible only for noise driven systems.

B. Relation to statistical approaches

Our approach uses ideas of feature selection and extraction similar to those in other statistical methods. We can employ the same procedures, e.g., principal components, Fourier or wavelet analysis. In this paper we restrict ourselves to a small set of different features, and it seems that there is no general guideline for an optimal choice of features. The main difference of our approach compared to more traditional methods of classification and characterization lies in our hypothesis about the source of the data and hence in the information which we want to extract.

In statistical pattern recognition [7] one begins with a set of random vectors of length *N*, where each vector is a set of random variables which can be, but are not required to be, a time series, so that in general there is no dynamical information contained in them. The idea behind classification is then that there exist different classes of vectors which are characterized and distinguished by different probability distributions in this *N*-dimensional space. The minimal classification error is given essentially by the overlap of these distributions. Since in the typical classification task these distributions are unknown *a priori*, they also have to be estimated from a sample of vectors. The extraction of features, and hence the projection from this usually high-dimensional space onto a low-dimensional feature space, has two objectives: more reliable estimators of the probability distributions and the goal of simplifying the construction of the classifier. Good features are those which increase the classification error by this projection least.

In our approach we start with a distinction between a state space spanned by state variables $\vec{x}$, and a parameter space. This distinction is not present in the traditional framework. Thus, in the language of classification, we have a continuum of classes, since every set of fixed d system parameters, i.e., every point in the parameter space, defines a new class. On the other hand, since all classes can be distinguished in theory by a d-tuple of parameter values, we know that a projection onto a $(2d+1)$-dimensional feature space will suffice for our classification problem. In the ideal case (infinite data sets) there was no overlap of distributions. In realistic situations with finite time series, our feature vectors suffer from statistical inaccuracy; hence there is some overlap, but it is of a different origin than in statistical pattern recognition.

III. RECONSTRUCTION OF PARAMETERS IN LINEAR STOCHASTIC MODELS

A frequently used method for (also time dependent) feature extraction consists in fitting AR or ARMA models to data (see, e.g., Ref. [8] for two examples of electroencephalogram data). In our setting, these are models where the relation between features and parameters can be fully understood. The class of linear stochastic models is completely covered by the well studied autoregressive moving average ARMA(M,N) models,

$$x_{n+1}=\sum_{j=0}^{M-1} a_j x_{n-j}+\sum_{j=0}^{N} b_j \eta_{n-j}, \tag{3}$$

where η_j are independently Gaussian distributed random variables, and the set of fixed coefficients a_j and b_j determines the properties of the model [9]. These are the parameters of the system. The outcome x_n is again a Gaussian random variable with zero mean. Its distribution is fully characterized by its variance $s^2=\langle x_n^2\rangle$, i.e., all static features (higher moments of the probability distribution) are functions of s^2. Hence in such a feature space, an arbitrary ensemble of ARMA time series can at most yield $d_e=1$.

Consequently, one also has to test for dynamical influence in order to exclude (or verify) that more than one parameter is varied. For ARMA(M,N) models the set of coefficients a_j and b_j can be directly mapped onto the properties of the power spectrum $s^2(k)$,

$$s^2(k)=\left|\frac{\sum_{j=0}^{N} b_j e^{i2\pi kj/L}}{1-\sum_{j=1}^{M} a_j e^{i2\pi kj/L}}\right|, \tag{4}$$

for a time series of length L, which can be mapped to the autocorrelation function by another Fourier transform. In addition, it is well known that all higher order statistics can be derived from this second order statistics uniquely [9]. Evidently, an arbitrary feature vector is thus confined to an $(N+M)$-dimensional manifold in feature space, proving our proposition. Moreover, (selected frequency bands of) the power spectrum or, equivalently (selected lags of), the autocorrelation function, provide a very simple set of features which represent the full d_e parameters in feature space.

By the methods discussed in Sec. IV A, one can directly determine the coefficients a_j of the AR part of the model, but this way one does not have a handle on the MA part. This nonetheless makes sense if the systems under study are driven by external colored noise represented by the MA part, whereas only the AR part reflects the system's own dynamics which is suspected to vary.

IV. DETERMINISTIC SYSTEMS WITHOUT AND WITH WEAK NOISE

A. Equations of motion from data

The equations of motion of deterministic dynamical systems [i.e., $\vec{\xi}\equiv 0$ in Eq. (1)] can in principle be fully reconstructed from observed data [2]. In stochastic models (such as the ARMA) or weakly noisy deterministic models, only the deterministic feedback part can be reconstructed; however, this is often the relevant part. For simplicity, we restrict our discussion to discrete time [10], but there are no difficulties in extending the arguments to continuous time systems (differential equations) [11]. Let $\vec{x}_{n+1}=\vec{F}_p(\vec{x}_n)$ be the iteration of a map in the $\mathbf{R}^l$, and $s_n=g(\vec{x}_n)$ be a scalar measurement. Moreover, we assume that there are d parameters in $\vec{F}_p$ which will be varied. If instead of $\{s_n\}$ the series of vectors $\{\vec{x}_n\}$ were measured, the most direct and obvious way to reconstruct the d parameters would be to estimate $\vec{F}_p$ from the data. This is conveniently done by choosing a suitable functional form $\vec{G}_q$ for the unknown $\vec{F}$, depending on a huge set of free coefficients, and to minimize the one-step prediction error e where $e^2=(1/N)\Sigma[\vec{x}_{n+1}-\vec{G}_q(\vec{x}_n)]^2$, with respect to the coefficients q. If the true function $\vec{F}_p$ can be approximated by $\vec{G}_q$ with sufficient precision, a variation of the parameters p will induce a change of the coefficients q [3]. Our claim from above now means that the variation of the (in principle arbitrarily large) set of coefficients q takes place on a d-dimensional manifold. This statement becomes trivially true if $\vec{F}$ can be obtained from $\vec{G}$ by setting all but d coefficients identical to zero.

When only a scalar time series $\{s_n\}$ is recorded, due to the theorems of Sauer *et al.* [12], an equivalent of $\vec{F}$ exists in the delay embedding space: $s_{n+1}=H(\vec{s}_n)=H(s_n,s_{n-1},\ldots,s_{n-m+1})$. Again, H can be estimated from the data by choosing a suitable functional form and minimizing the one-step-prediction error e, where $e^2=(1/N)\Sigma[s_{n+1}-H_q(\vec{s}_n)]^2$ with respect to the free coefficients q. Due to the fact that there is a one-to-one relation between the dynamics in the delay embedding space and in the original phase space, H_q depends on the parameters p in $\vec{F}_p$. Under variation of p the set of arbitrarily many q's is confined to a d-dimensional manifold. The details of how the set of q's depends on the set of p's depend on the details of $\vec{H}$ and the measurement function g by which the scalar ob-

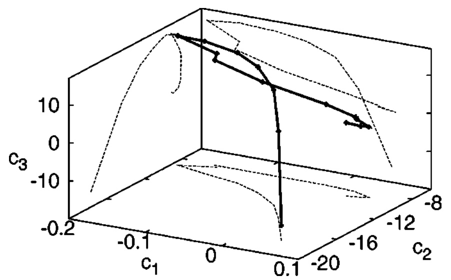

FIG. 1. The coefficients c_1, c_2, and c_3 of the fits to data generated by the tent map [Eq. (5)], (connected by lines for increasing a) and the projections onto each of the three planes.

servables are obtained. In general, the dependence is nonlinear, such that the manifold to which the q's are confined is curved and thus can be embedded in an $\mathbf{R}^k$ only for sufficiently large k, where Whitney's theorem states that $k > 2d$ is sufficient.

As an illustration, we study time series generated by tent maps of the type

$$x_{n+1} = \tfrac{1}{2} - a|x_n| \tag{5}$$

for $a \in [1,1.6]$. The chaotic motion is confined to two subintervals of the interval $[-1/2,1/2]$. We assume that we do not know the source of the data, and perform a fit with a sixth order polynomial, $G_c(x) = \sum_{i=0}^{6} c_i x^i$. Resulting fit coefficients obtained from 20 time series of length 1000, each for different values of a, are shown in Fig. 1: The variations of c_1, c_2, and c_3 (the other coefficients yield equivalent figures) clearly confirm that all time series are related to the variation of a single parameter. This rather simple example demonstrates that we do not have to require the (unrealistic) case that the ''true'' equation of motion can be represented exactly by our fit.

The reconstruction of the equations of motion is the most obvious way to get a handle on the unobserved parameters. However, in principle, every feature which depends smoothly on the parameters should be suited as one coordinate in the reconstruction space. In the tent-map example, the one-step prediction error e obtained by the fits with sixth order polynomials, together with, e.g., the variance of the distribution of the data, and its mean, again yields feature vectors which nicely align on a one-dimensional curve.

B. Bifurcations

Complications for deterministic dynamical systems arise from the possibility of bifurcations. Under smooth changes of parameters, the dynamical behavior can change drastically at a bifurcation point in the parameter space, such as the birth of stable periodic orbits from a chaotic attractor. Most features will thus strongly change at a bifurcation point. From the theoretical point of view most bifurcation types create a continuous, albeit fast, change of properties such as the invariant measure and the Lyapunov exponents. As a prototypical example, let us consider a part of the bifurcation diagram of the Hénon map, $x_{n+1} = 1 - a x_n^2 + b x_{n-1}$, for a path

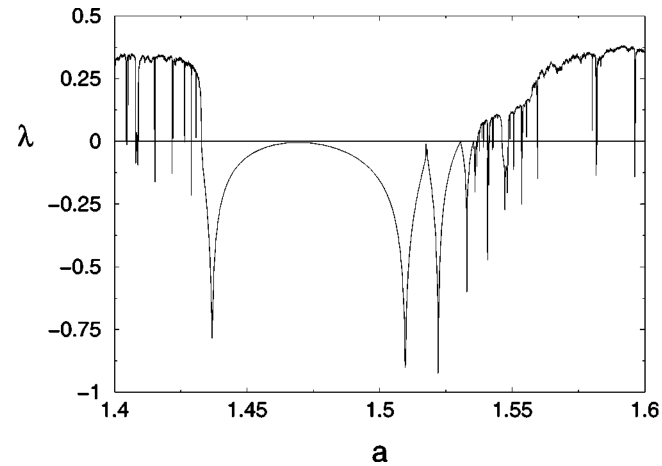

FIG. 2. The maximal Lyapunov exponent λ of the Hénon map [Eq. (6), without the noise term] plotted vs the parameter a. The corresponding path in the parameter space (a,b) is indicated in Fig. 3 by a black line. Although λ varies smoothly at the bifurcation points (zero crossings), and continuously elsewhere, large fluctuations due to bifurcations occur on arbitrary small scales on the parameter axis.

in the (a,b) plane shown in Fig. 3. The variation of the maximal Lyapunov exponent of typical orbits is shown in Fig. 2.

Subcritical bifurcations create a discontinuous change of all features. But even if bifurcations are continuous as in the above example of the Hénon map, the changes can be so rapid as a function of the control parameter that they will, in practice, appear to be discontinuous. Features will thus effectively jump at bifurcations, and it will be hard to identify whether time series before and after a bifurcation are related to each other by the variation of a single parameter. The typical dynamical system is not uniformly hyperbolic, and thus can have bifurcations everywhere in parameter space. From a theoretical point of view this is destructive of our concept, since embedding in the spirit of Whitney is only possible if the features are continuous functions of the parameters if we want to conserve topology, and even smooth functions of the parameters if we want to conserve a metric structure.

In practice, however, purely deterministic low-dimensional dynamical systems are rare, and as soon as randomness comes into play all changes at bifurcations can be expected to be not only continuous but even smooth. Introducing a few percent of interactive noise into the Hénon map, i.e., replacing the Hénon map by a stochastic process of the form

$$x_{n+1} = 1 - a x_n^2 + b x_{n-1} + \xi_n(x_n), \tag{6}$$

where $\xi(x_n)$ is white noise (for technical reasons correlated with the data such that trajectories do not leave the basin of attraction and escape to infinity), smoothes the bifurcations sufficiently. The smoothness is related to the fact that interactive noise can be reinterpreted as some fast stochastic fluctuation of the parameters a or b:

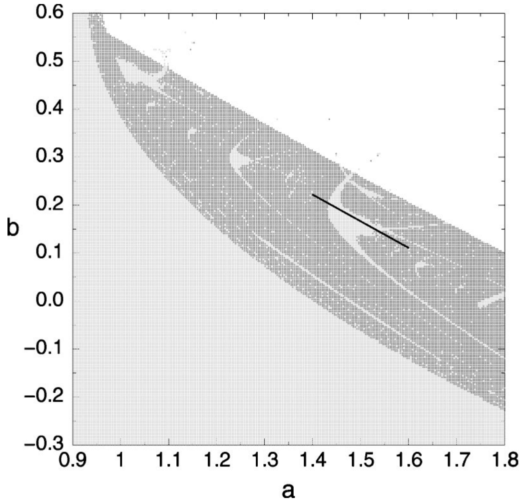

FIG. 3. The parameter space (a,b) of the Hénon map [Eq. (6), without the noise term]. Only the largest stable islands are resolved here. A negative maximal Lyapunov exponent λ of the trajectories is represented by light gray, while the darkly shaded region corresponds to positive λ. For parameter values in the white region, no bounded solution exists. The black line indicates the path in parameter space for which λ in Fig. 2 is calculated.

$$x_{n+1}=1-a_n x_n^2+b x_{n-1}, \quad a_n=a(1+\tilde{\xi}_n(x_n)). \tag{7}$$

Hence adding noise to the equations of motion corresponds to the elimination of all structure in parameter space on scales smaller than some cutoff related to the noise amplitude.

Numerical simulations of the noisy Hénon map illustrate this nicely. In Figs. 3 and 4 the parameter plane of Eq. (6) without (with) the noise term is shown. Different gray scales represent the three types of asymptotic dynamical behaviors, namely, stable periodic orbits (light gray), chaotic trajectories in a bounded region of phase space (dark gray), and escape to infinity (white). In addition to this qualitative characterization, we computed the maximal Lyapunov exponent of every trajectory through expansion in tangent space (eigenvalues of products of Jacobians along the trajectory), which is defined for both purely deterministic and noisy trajectories. Depending on the choice of parameters (a,b) (and hence the variance of the signal), the standard deviation of the noise amplitude is $\approx 2-4$ % of the standard deviation of the time series. Evidently, the small amount of noise wipes out all structures on small length scales in the parameter space. This leads to the vanishing of all stable islands in the darkly shaded region in Fig. 3; however, even if some larger one remained, the maximal Lyapunov exponent and the invariant measure would change smoothly when the parameters approach the stability regime of a periodic orbit.

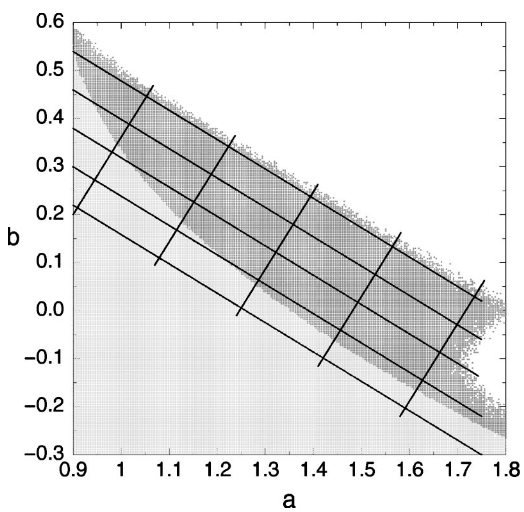

FIG. 4. The parameter space (a,b) of Eq. (6) including the noise term; the gray shading is as in Fig. 3. The black grid indicates the parameter set underlying the standardizing sample of Fig. 7.

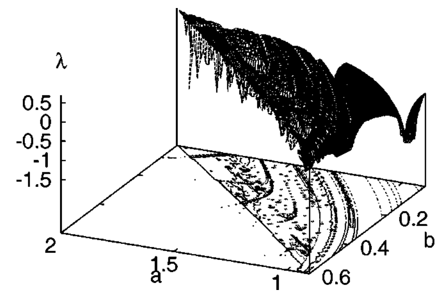

FIG. 5. The maximal Lyapunov exponent λ of the unperturbed Hénon map for the parameter values (a,b) on a 120×120 grid in the parameter plane for which a bounded solution exists.

The effect of interactive noise on the Lyapunov exponent is illustrated in Figs. 5 and 6. We plot the maximal Lyapunov exponent λ of Eq. (6) for the parameter values (a,b) on a grid in the parameter plane for which a bounded solution exists. Without a noise term coupled to the Hénon map the very intricate structure of λ due to the large number of bifurcations is evident. The smooth but very rapid fluctuations of λ in Fig. 5 can be traced to arbitrarily small length scales, as shown in Fig. 2. Other features, such as the expectation value or the standard deviation of a time series, show a similar behavior. In clear contrast to this, the Lyapunov exponent (and as well the other features) calculated from Eq. (6) varies only slowly due to the smoothing effect of the dynamical noise on small length scales in parameter space.

The identification of the varying parameters can be performed for this smoothed situation. Selecting 2000 pairs of parameter values (a,b) on each line of the grid plotted in Fig. 4, we produce a time series of length 2000 for each of

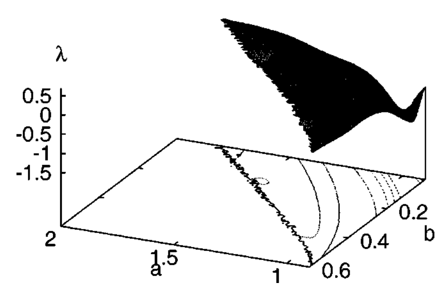

FIG. 6. The maximal Lyapunov exponent λ of Eq. (6) for the same set of parameters as in Fig. 5.

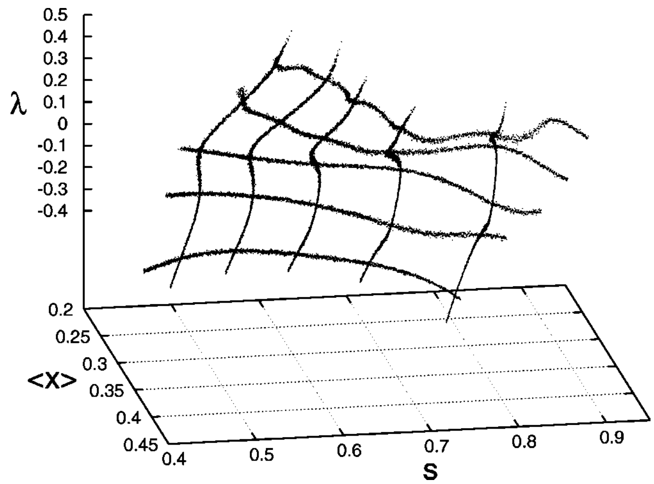

FIG. 7. The feature vectors $(\langle x\rangle_i, s_i, \lambda_i)$ of the standardizing sample plotted in the feature space; compare to Fig. 4.

them. This set forms our standardizing sample, and, since both a and b were evidently varied, the effective number of parameters is 2. As features, here we choose the expectation value $\overline{x}=\langle x\rangle$, the standard deviation $s=\sqrt{\langle(x-\overline{x})^2\rangle}$, and the maximal Lyapunov exponent λ. In Fig. 7, every time series of the standardizing sample is represented by a dot showing the values of the triple $(\overline{x},s,\lambda)$.

As can clearly be seen, the topology of the grid in the parameter space can be nicely recovered in the feature space. The lines are blurred as a consequence of finite samples: The features on every time series segment suffer slightly from statistical errors and thus characterize, to a very small extent, the particular finite time series and not only the underlying parameters. However, these uncertainties of the features are small enough not to deteriorate the usefulness of this concept.

C. Estimating the number of parameters—dimensions in feature space

It is desirable to determine the number of parameters to describe the variations in a given standardizing sample, i.e., to estimate the number of effective parameters d_e in the context of statistical pattern recognition also known as the intrinsic dimension [7]. Under certain conditions, which we will discuss below, methods normally used to estimate fractal dimensions of attractors can be used for this task. We will use the famous Grassberger-Procaccia correlation sum for the estimation of the correlation dimension, but local (because of curvatures) linear methods such as the local single value decomposition (SVD) method [13] can also be employed.

As an example, we use two samples of time series produced by the noisy Hénon map [Eq. (6)]. The parameters were chosen according to $a=0.9+0.9p_1$ and $b=-0.3+0.9p_2$, 10000 values for each set. If a trajectory diverges to infinity, the corresponding parameters were discarded and replaced by a new random pair. For the first sample, p_1 was chosen randomly between 0 and 1, while $p_2=1-p_1$, so we expect that $d_e=1$. For the second set both p_1 and p_2 were independent random numbers uniformly distributed in [0,1] such that $d_e=2$. For every time series in these sets we again estimated the features: the largest Lyapunov exponent λ, the mean $\overline{x}$, and the standard deviation s. By this procedure the two samples of time series were represented by two sets of points in the three-dimensional feature space, similar to Fig. 7. The $d_e=1$ case yields a blurred curve in this space, and the $d_e=2$ case a blurred and bent surface.

We can numerically extract a dimensionality from these data. Given a set of N vectors $\vec{y}_i\in\mathbf{R}^m$, the Grassberger-Procaccia correlation sum is defined as

$$C(\epsilon)=\frac{2}{N(N-1)}\sum_{i<j}^{N}\Theta(\epsilon-d(\vec{y}_i-\vec{y}_j)), \tag{8}$$

where Θ is the Heaviside step function, and d is a metric. When the vectors $\vec{y}_i$ are a random sample taken from a set with (fractal) dimension $D_f\leqslant m$, then there will be (under suitable conditions, e.g., N sufficiently large) a scaling range in ϵ where $C(\epsilon)\propto\epsilon^{D_f}$. In our context the vectors $\vec{y}_i$ would naturally be the feature vectors of the time series S_i. We then expect the correlation sum to scale as

$$C(\epsilon)\propto\epsilon^{d_e}. \tag{9}$$

The local slopes of $\ln C$ vs $\ln\epsilon$,

$$d(\epsilon)=\frac{d\ln C(\epsilon)}{d\ln\epsilon},$$

are usually regarded as length scale dependent dimensions which can be useful for the characterization of noisy attractors.

In the present case d_e also typically depends on ϵ. The influence of parameter variation on the selected features can be different, so that the extension of the data cloud is different in different spatial directions, so that on large scales fewer dimensions are visible. Moreover, the estimation of each feature vector from a finite time series produces statistical fluctuations which look like noise and increase the dimension on the small scales, as can be seen in Fig. 8.

Note that it is not the feature vectors themselves that enter but only their pairwise distances. If we are able to determine the distances between pairs of time series, $d(S_i,S_j)$, directly, in principle we have the possibility of estimating d_e without constructing any feature vectors, and thus circumventing the delicate task of feature selection. Unfortunately, there are still several unsolved problems with this approach. The main problem is to find a good metric for the time series. Some candidates based on cross-prediction errors [14,4], which are attractive from the physical and computational points of view, are only dissimilarity measures and do not possess all the properties of a metric; also see Refs. [15,16] for a discussion of some other measures. Up to now we have not succeeded in finding a metric which produced robust and easy interpretable results, comparable to these shown in Fig. 8.

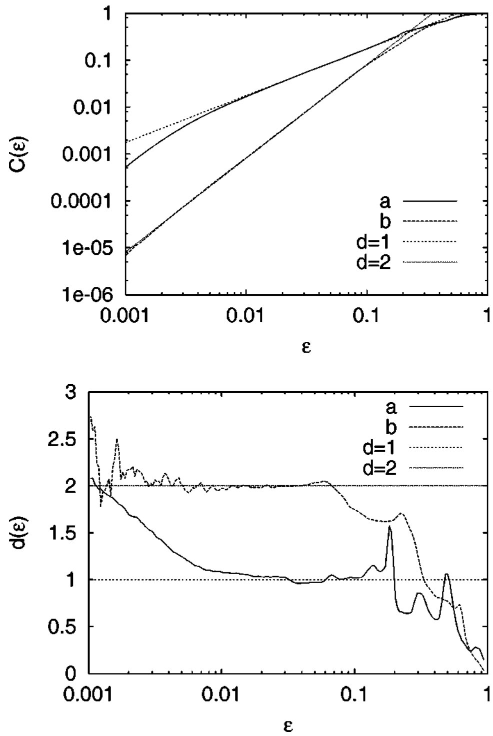

FIG. 8. $C(\epsilon)$ (top) and the local slopes $d(\epsilon)$ (below) estimated in the feature space spanned by $(\langle x\rangle_i, s_i, \lambda_i)$ (also compare Fig. 7) varying one parameter (a) and two parameters (b).

V. NONSTATIONARITY AND PATHS IN FEATURE SPACE

The previous discussions were based on standardizing samples where each single time series was stationary by construction. In a nonstationary setting, typically a single long time series with time dependent parameters will be cut into segments which, for the sake of identical statistical errors of the features, should have equal lengths. These segments are at best pseudostationary, i.e., the parameter variation inside each given segment can be neglected. To increase the number of feature vectors, one can use overlapping segments which form the standardizing sample. Since now the sample elements possess a natural time ordering, it makes sense to study the path in the feature space thus created. For an illustration here we use data from the Mackey-Glass delay differential equation [17], which for our parameter values and the time lag $\tau = 55$ has attractor dimensions between unity and about 10:

$$\dot{x}(t) = \frac{a x(t-\tau)}{1 + x(t-\tau)^{10}} - b x(t). \qquad (10)$$

The parameters a and b were varied on an elliptic path in this two-dimensional parameter space, $a(t) = 0.3 + 0.1\cos(\Omega t)$ $b(t) = 0.125 + 0.025\sin(\Omega t)$, and $\Omega = 1/(2000\tau)$. The complete nonstationary time series sampled with $\Delta t = \tau/2$ is plotted in Fig. 9. As features, we use the two-point correlation $m_{02} = \langle x_t x_{t-2}\rangle$, the mean $\langle x\rangle$, and the standard deviation s^2 of the data on 50 disjoint moving windows of length 200τ each. Although now the dynamics of each time series segment is not strictly stationary, we can of course compute the

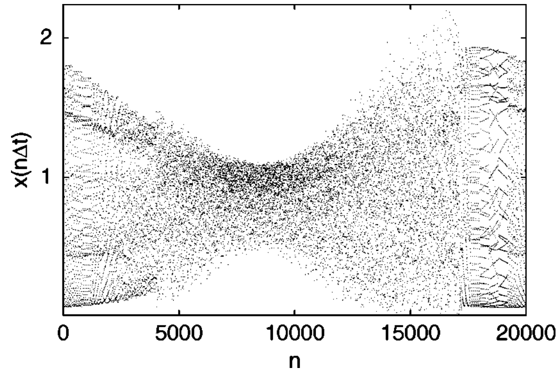

FIG. 9. Time series of the nonstationary Mackey-Glass system [Eq. (10)].

values of these features, which can be interpreted as a kind of average over the small remaining nonstationarity inside each segment. We see a clear loop in feature space in Fig. 10 with the topology of a circle. Longer time series yield even clearer results, but we purposely restrict ourselves here to a small amount of data in order to demonstrate the applicability of the method in realistic situations.

The situation of nonstationarity requires an additional discussion of time scales. In order to be able to compute the features in a robust way, the time dependence of the parameters should be slow compared to the internal time scale of those variables which are to be eliminated. The shorter the time series segments on which features are computed, the large are the statistical errors of the values of these features. Overly large statistical fluctuations may conceal the structure one is searching for. If no clear time scale separation exists, everything is difficult and the concept might be unapplicable.

In our example for nonstationarity, a second issue is hidden: the harmonic time dependence of $a(t)$ and $b(t)$ can be created by two additional deterministic degrees of freedom (e.g., $\dot{a} = -\omega_1^2 b$, $\dot{b} = -\omega_2 a$, with suitable ω_1 and ω_2), in which case there is no nonstationarity at all. However, in order to reconstruct these in a time delay embedding space á la Takens, one should use a time lag which is huge compared to a suitable lag for the Mackey-Glass dynamics, so that it is practically impossible to reconstruct both parts of the dynamics, the Mackey-Glass dynamics, and the $a-b$ dynamics, in the same time delay embedding space. Hence the point of view of nonstationarity is more appropriate. None-

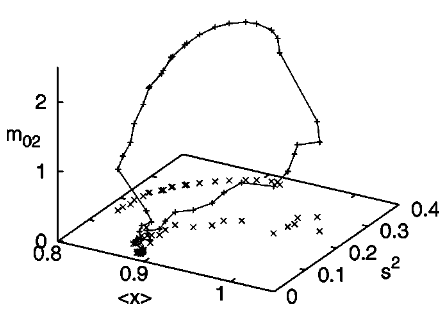

FIG. 10. Loop in feature space, representing the nonstationary Mackey-Glass time series of Fig. 9. Crosses: projection of the feature vectors onto the bottom plane.

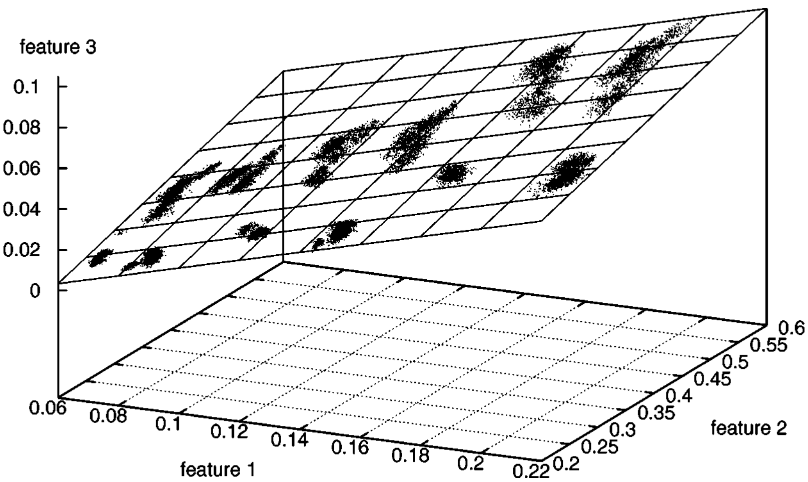

FIG. 11. A set of feature vectors $\mathbf{v}_2 \in V_2$ calculated from the time series of the stator current of an induction motor. The entries are three components of windowed Fourier spectra of sections of the time series which are harmonic to the electric supply frequency.

theless, it shows that there is (naturally) some ambiguity in the distinction between system variables and parameters. In practice, the interpretation of what can be considered as noise is also vague. Often, drifting parameters and noise have a clear time scale separation, but certain types of processes such as $1/f$ can yield complications.

VI. FAILURE DETECTION FOR INDUCTION MOTORS

Finally we want to show a practical application of our concept of parameter identification to a problem of signal processing, the failure detection of induction (electro)motors, by monitoring a single phase of the stator current. This problem from electrical engineering has the advantage of being a real world problem instead of an experiment which may simply be designed to show low-dimensional behavior; see Ref. [18], and references cited therein, for a closer introduction to this subject.

The stator current of induction motors can easily be measured at the power supply, and offers a cheap method for monitoring these machines which are very widespread in industry applications (e.g., to drive conveyor belts, assembly lines, air-condition systems, etc.). The difficulty is to isolate changes of the stator current caused by (developing) failures from all other influences where, of course, possible failures have to be detected well before the motor breaks down. The quantities which mainly influence the stator current are (besides the electrical supply frequency) the magnitude and time dependence of the torque of the motor (i.e., it makes a difference if the motor drives a constant torque or an oscillating load torque like a compressor), the production tolerances (the air gap between the stator and rotor and the winding distribution of the stator), the environmental conditions present during operation (essentially the temperature and the air humidity), and, last but not least, possible failures. Typical failures includes damage to the rotor due to overheating, imbalances of the driven loads, and damage to the bearing races due to (continuous) abrasion. The impact of the load torque and the production tolerances on the stator current is typically about one magnitude larger than that of the environmental conditions and possible failures. The influences of these last two quantities are about the same. The rotational frequency of induction motors cannot be regulated, and is slightly less (1– 2 %, depending on the load torque) than one half or one time (depending on the construction of the motor) the electrical supply frequency. Therefore, this quantity does not independently enter the stator current.

From this it follows that a reliable monitoring of induction motors requires training a fault sensing algorithm with data recorded from the particular (healthy) motor which should be monitored in order to learn about the load states and production tolerances of this motor. Additionally, a selection of the environmental conditions present during normal operation can be covered. However, since in practice one cannot expect that all potential environmental conditions are contained in the training set, the failure detection algorithm must be able to distinguish between previously unobserved motor states and actual faults. In Ref. [18] a method of geometric signal separation for the failure detection of induction motors was introduced, which successfully deals with this problem.

Here we only want to show one aspect of this problem, whose analysis is important for the development of a solution: the environmental degrees of freedom of the stator current of induction motors. Due to the quasiperiodicity of the stator current and theoretical arguments, the entries of the feature vectors are appropriately chosen components of windowed Fourier spectra of sections of the recorded time series. In Ref. [18] it was shown that there exist two feature spaces V_1 and V_2, where in V_2 almost only information about the average load torque, the production tolerances, and the environmental conditions enter, while in V_1 the information about motor failures and the time dependence of the torque is *additionally* contained. The components of the feature vectors $\mathbf{v}_2 \in V_2$ have to be harmonics of the electrical supply frequency. In Fig. 11 a set of feature vectors $\mathbf{v}_2$ calculated from data of an induction motor is plotted in the three-dimensional feature space V_2.

The data set consists of seven recordings from a four-pole induction motor with 8-kW power, where each recording contains 12 time series (of 9-min length each) which were generated during the operation of a single load. Four types of loads (constant, sinusoidal at a rotating frequency of the motor, sinusoidal at a half rotating frequency, and sinusoidal at a twice rotating frequency) were operated at each of three torques (half, three quarter, and full rated loads of the motor). The environmental conditions between the different recordings have partially changed, but we do not have any information about this. Four recordings were done with an unbalanced disk attached to the rotor shaft of the motor in order to simulate an imbalance. From each of the 84 time series, 204 feature vectors were calculated which are plotted in Fig. 11.

Calculating the covariance matrix of this set of feature vectors shows that, to a very good approximation, all vectors lie in a plane spanned by the eigenvectors with the largest eigenvalues of the covariance matrix. The eigenvalue of the eigenvector perpendicular to the plane turns out to be 63 times smaller than the second largest eigenvalue, and about 320 times smaller than the largest eigenvalue.

The 12 large clusters in Fig. 11, which are well separated from each other, correspond to the three load torques of the motor and to four environmental conditions which appear to be clearly different. We may conclude that two of the seven recordings which correspond to clusters lying pairwise close to each other are recorded under similar environmental conditions (i.e., within a short period of time); however, as mentioned above we do not have information about this. We now can conclude that the two visible degrees of freedom in the feature space V_2 correspond to the average load torque of the motor and to one environmental degree of freedom which is probably dominated by the temperature. We also expect that in the higher-dimensional, failure sensitive feature space V_1, only one environmental degree of freedom visibly enters. This is a very useful result, because the structure of V_1 is much more complicated than that of V_2. We want to mention that, despite the fact that induction motors may appear not to be very complicated systems, these results cannot be obtained from theoretical arguments or models of induction machines.

VII. OVEREMBEDDING AND INSTANTANEOUS PARAMETERS

Our approach in this paper is to rely exclusively on time series data representing the systems' dynamics. If, however, a training set of data is available, where in addition to observables system parameters are measured synchronously, and this training set covers a suitably large range of different parameter settings, an alternative approach suggested by Parlitz *et al.* [5] can be followed. Since some aspects are related to our approach, this shall be briefly revisited here. This approach can be formalized by the idea of overembedding introduced in Ref. [19]: An implicit knowledge of the equations of motion is represented by a delay vector and the next observation, $(\vec{s}_n, s_{n+1})$, since these vectors of the extended delay embedding space are confined to a manifold. Equations of motion corresponding to a modified parameter configuration yield a different manifold, to which the delay vectors of the modified dynamics are again confined. For its unambiguous distinction from the first manifold, one has to extend the phase space further. For each time dependent parameter, one needs two additional directions in phase space, as shown in Ref. [19]. In this extended phase space, neighbors of a given delay vector are states of the dynamical system observed under the same parameter setting referred to in the present delay vector, or others nearby. It was shown in Ref. [19] that for D variable parameters and an N-dimensional phase space, one has to reconstruct at most $N+D$ degrees of freedom, and thus use an embedding space of $m>2(N+D)$.

Together with this background, the idea of Parlitz *et al.* relies on the existence of a training set which contains direct measurements of those parameters which should be determined later on in situations where they are no longer measured. Thus we start from time series of tuples $(s_n, \vec{p}_n)$. The series of s_n is converted into delay vectors with a higher than minimal dimension, and through this procedure every delay vector $\vec{s}_n$ possesses a corresponding parameter vector $\vec{p}_n$. Thus there is a trivial map from $\vec{s}$ to $\vec{p}$. For every delay vector of a test set one searches for the N closest delay vectors from the training set, for which there exist N corresponding parameter vectors. The actual setting of the parameters is then a suitable average over these parameter vectors corresponding to the training vectors. Overembedding is needed in order to guarantee that the map from $\vec{s}$ to $\vec{p}$ is invertible.

Thus here there is no need to select suitable features; instead the training situation fixes which parameters can be identified from the test set. This concept can deal with much shorter time series segments in the test phase, since one can give an estimate of $\vec{p}$ for every single m-dimensional delay vector derived from the time series. The training set, however, has to be as large as or even larger than the whole data sets we are using in this paper, since for every delay vector in the test set one needs a "good" neighbor in the training set for a reasonable parameter identification. Thus the training data have to fill the high-dimensional embedding space reasonably well. In addition, as mentioned above, in the training period the setting of the parameter vectors $\vec{p}_n$ also has to be recorded. Thus this approach applies to a different setting of the problem.

VIII. CONCLUSIONS

We have shown that variations of parameters underlying a set of time series can be identified qualitatively and quantitatively in terms of their number and the manner of variation in feature spaces. This supplies an approach to characterizations of both nonstationary dynamical phenomena and sets of time series stemming from related dynamical phenomena. In nonstationary situations, one can follow paths in feature space representing the time evolution of system parameters, and thus identify their own dynamics; this amounts to a tremendous reduction of complexity, since the fast degrees of freedom, called $\vec{x}$ and $\vec{\xi}$ in Eq. (1), are eliminated.

Although in the motivation of the method we drew a parallel to the time delay embedding method for the reconstruction of state spaces from scalar time series, there are some essential differences which make the range of applicability of the parameter reconstruction different. This method applies to deterministic and stochastic processes, and a strict stationarity of the single time series segments is not required. On the other hand, we do not see a way to construct a feature space which is general enough to *guarantee* that all time dependent parameters will be visible. Presumably for each problem one has to optimize the feature space and to try different combinations of features to represent the changes of all parameters on approximately similar scales.

ACKNOWLEDGMENT

Part of this work was supported by BMBF Grant No. 13N7085.

[1] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).

[2] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

[3] J. Kadtke, Phys. Lett. A **203**, 196 (1995).

[4] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **79**, 1475 (1997).

[5] U. Parlitz, R. Zöller, J. Holzfuss, and W. Lauterborn, Int. J. Bifurcation Chaos Appl. Sci. Eng. **4**, 1715 (1994).

[6] H. Whitney, Ann. Math. **37**, 645 (1936).

[7] K. Fukunaga, *Statistical Pattern Recognition* (Academic Press, Boston, 1990).

[8] A. C. Tsoi, D. S.-C. So and A. A. Sergejew, in *Advances in Neural Information Processing Systems 6*, edited by J. D. Cowan, G. Tesauro and J. Alspector, (Kaufmann, San Francisco, 1994), pp. 1151–1158; J. Pardy, S. Roberts, L. Tarassenko, and J. Stradling, J. Sleep Res. **5**, 201 (1996).

[9] G. E. P. Box and G. M. Jenkins, *Time Series Analysis* (Holden-Day, San Francisco, 1976).

[10] J. D. Farmer and J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).

[11] J. P. Crutchfield and B. S. McNamara, Complex Syst. **1**, 417 (1987).

[12] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[13] T. Hediger, A. Passamante, and M. E. Farrell, Phys. Rev. A **41**, 5325 (1990).

[14] T. Schreiber, Phys. Rev. Lett. **78**, 843 (1997).

[15] R. Moeckel and B. Murray, Physica D **102**, 187 (1997).

[16] T. Schreiber, Phys. Rep. **308**, 2 (1999).

[17] M. C. Mackey and L. Glass, Science **197**, 287 (1977).

[18] S. Güttler and H. Kantz, Electr Machines Power Syst. **28**, 515 (2000).

[19] R. Hegger, H. Kantz, L. Matassini, and T. Schreiber, Phys. Rev. Lett. **84**, 4092 (2000).