Using topological statistics to detect determinism in time series

Guillermo J. Ortega

Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes, Roque Saenz Peña 180, 1876 Bernal, Argentina

Enrique Louis

Departamento de Física Aplicada, Facultad de Ciencias, Universidad de Alicante, Apartado 99, E-03020 Alicante, Spain

Statistical differentiability of the measure along the reconstructed trajectory is a good candidate to quantify determinism in time series. The procedure is based upon a formula that explicitly shows the sensitivity of the measure to stochasticity. Numerical results for partially surrogated time series and series derived from several stochastic models, illustrate the usefulness of the method proposed here. The method is shown to work also for high–dimensional systems and experimental time series.

PACS number(s): 05.45.−a, 02.30.Cj, 07.05.Kf

I. INTRODUCTION

This paper deals with the problem of determining whether the complex behavior of a single time series may be explained in terms of a deterministic or a stochastic dynamical system. Although the idea was present from the very first days of the nonlinear time series analysis [1], methods explicitly aiming to detect determinism in time series [2–5] have only been published rather recently. The fact that noise can mimic or mask deterministic (e.g., chaotic) behavior in classical measures of chaos (Lyapunov exponents, K_2 entropy and correlation dimension) [6–8] has urged the need for more specific methods to discriminate deterministic from stochastic behavior.

Measures and vector fields [9], densities and trajectories [10], metric and topological [11], statistical and geometrical, these are, roughly speaking, different terms used to denominate the two broad approaches for investigating the chaotic behavior of dynamical systems. It is a remarkable fact that most of the aforementioned methods to detect determinism are based in the study of trajectories or vector fields. Considering the noisy character of the inverse problem of time series analysis (from the point of view of nonlinear dynamics), the statistical approach must lie at the heart of any proposed methodology. This becomes relevant as soon as the system underlying a time series is ''excited'' beyond the simple bifurcations, so that the geometrical information about the shape of the attractor, or the motion on it, is no longer available. This is also true in the case of a stochastic dynamical system, where noise can make it impossible to reconstruct the geometrical information [12]. Moreover, trajectories and the geometry of the flow *on* the attractor are well defined in the case of a deterministic system, in contrast to the case of a stochastic dynamics.

Bearing that in mind, we have developed a novel approach based on the study of the differentiability of the natural measure. The statistical character of our methodology is twofold. The first one concerns the use of the natural (or physical) measure. Differentiability of this invariant measure will be considered in the direction of the evolution, that is, along any typical trajectory. In this way, ''smoothness'' of the trajectory *and* measure are considered in a single step.

This is a general result of dissipative systems, regardless of its application to time series analysis. The second one is the use of statistics for analyzing mathematical properties [13], which allow us to test the differentiability of the measure in a statistical sense. In this way we can extend our methods to signal analysis.

The procedure is based upon a formula that will be derived in Sec. II, and that will allow us to discriminate between deterministic and stochastic behaviors. This problem was tackled by us in [14], whereas the basic idea of the methodology was proposed in [15,16].

The paper is organized as follows. Section II is devoted to the mathematical background of our approach. At the end of the section we have included a brief summary of the concepts introduced in this section. This is addressed to the less inclined mathematical reader, which can skip most of the mathematical details given in the section. In Sec. II A we introduce the fundamentals of our approach. The extension to stochastic dynamics is considered in Sec. II B. General concepts about measure projections and time series are explained in Sec. II C, whereas Sec. II D is devoted to a brief discussion of the statistics of topological properties used in this work. Section III is devoted to discussing the models to which we apply our method and the numerical procedures followed in each case. We also propose a variant of the surrogation method, partial surrogation, as a way to tune the degree of stochasticity. The results are presented in Sec. IV. The sensitivity of the time derivative of the measure to stochasticity, the dependence of the results on the embedding dimension, and the application of our approach to mixed time series are illustrated on the models discussed in Sec. III. We also present a simple test on an experimental time series like the Belousov–Zhabotinski chemical reaction.

II. THEORETICAL APPROACH

Here we discuss in detail the mathematical background of our method. The natural measure is first defined and the equation of motion that it obeys explicitly given. Then we show how stochasticity triggers a *wild* behavior of the time derivative of the natural measure. This is the feature upon which our proposal is based. In order to evaluate quantita-

tively this behavior we borrow the method proposed by Pecora, Carroll, and Heagy [13]. In particular we evaluate (statistically) the continuity of the logarithmic derivative of the natural measure. Those readers more interested in applications may go directly to Sec. II E where the most important features of the method are highlighted with the aid of the equations discussed hereafter.

A. The natural invariant measure along the trajectory

Consider a dissipative dynamical system described by n-first-order differential equations $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$. The corresponding flow f^t maps a "typical" initial condition $\mathbf{x}_0$ into $\mathbf{x}(t) = f^t(\mathbf{x}_0)$ at time t. Once transients are over, the motion settles over the attractor $\mathcal{A}$.

Given such a system, we can define a probability space (measure space) $(\mathcal{A}, \mathcal{B}, \mu)$, where $\mathcal{B}$ is the σ algebra generated by the open sets of the invariant set $\mathcal{A}$, and μ is an invariant measure defined over the subsets of $\mathcal{B}$ such that $\mu: \mathcal{B} \to (0,1)$.

Of all the invariant measures which can be defined, only one is relevant from the point of view of the experimentalist or in computer simulations. This is the *natural* invariant measure, which gives the limiting distribution of almost all starting initial conditions. Using the indicator function $1_B(\mathbf{y}) = 1$, if $\mathbf{y} \in B_\epsilon$ and 0 otherwise, we can define this measure for a set $B_\epsilon(\mathbf{x}) = \{\mathbf{y}: d(\mathbf{x}, \mathbf{y}) \leq \epsilon\}$ as

$$\mu(B_\epsilon(\mathbf{x})) = \lim_{t \to \infty} \frac{1}{t} \int_0^t 1_B(f^\tau(\mathbf{y}_0)) d\tau$$

for almost all $\mathbf{y}_0$ in the basin of attraction. The so-defined measure is invariant in the sense that it remains constant upon application of the evolution operator f^τ,

$$\mu(B_\epsilon(\mathbf{x})) = \mu(f^\tau(B_\epsilon(\mathbf{x})).$$

In the present case what we calculate is $\mu(B_\epsilon(\mathbf{x}(t))$ along the trajectory. This means that the evolution operator changes the point at which the set B_ϵ is centered, namely, $\mu(B_\epsilon(f^\tau(\mathbf{x})))$, leaving unchanged the ball B_ϵ. This is the so-called Lagrangian evolution (see Fig. 1), that has to be distinguished from the Liouvillian evolution $\mu(B_\epsilon(\mathbf{x}(t)) \to f^\tau(\mu(B_\epsilon(\mathbf{x}(t))$, which contracts the volume along the trajectory, as illustrated in Fig. 1.

Calling $\mu(\mathbf{x}(t)) = \mu(B_\epsilon(\mathbf{x}(t)))$, the "material derivative" of the natural measure can be expressed as

$$\frac{d\mu(\mathbf{x}(t))}{dt} = \dot{\mathbf{x}} \cdot \nabla \mu \qquad (1)$$

or

$$\frac{d\mu(\mathbf{x}(t))}{dt} = \nabla \cdot (\mu \mathbf{F}) - \mu \nabla \cdot \mathbf{F}. \qquad (2)$$

Equation (2) only involves partial derivatives of $\mu(\mathbf{x}(t))$ along the trajectory, and since each trajectory in the attractor is contained in the support, $\mu(\mathbf{x}(t))$ is smooth along any trajectory, and is therefore well defined. It is known that for axiom A systems, hyperbolic and with a dense set of periodic orbits, it is possible to decompose the tangent bundle at each
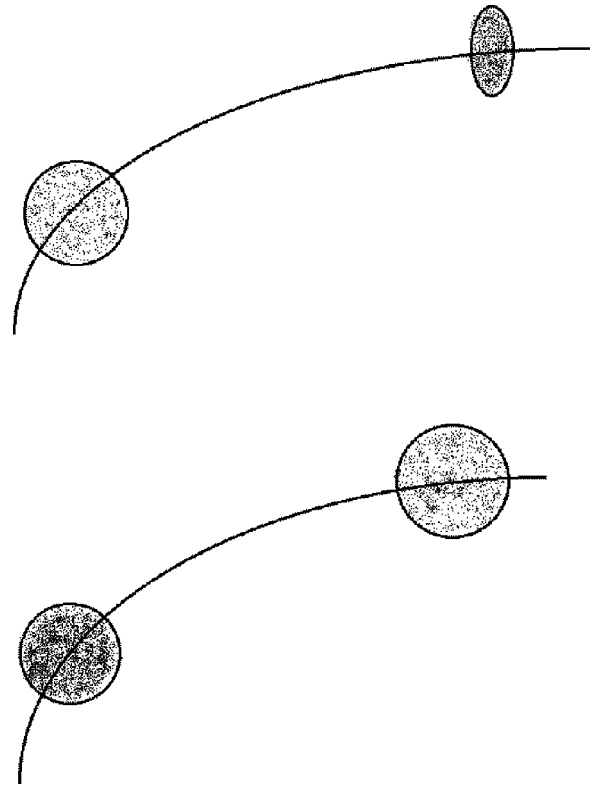

FIG. 1. Illustrates two alternative evolutions in dissipative systems: Liouvillian evolution (upper) and Lagrangian evolution (lower).

$\mathbf{x} \in \mathcal{A}$ in two linear subspaces; stable and unstable. It is also possible to define a measure in these systems, known as a Sinai–Ruelle–Bowen (SRB) which have the property of being smooth along the unstable manifold [9], due fundamentally to the stretching of the trajectories. On the other side, it is expected to have a wild behavior along the stable direction. In the present case, considering that Eq. (2) only involves derivatives along the trajectory and that in this direction no expansion or compression of the flux occurs (the associated Lyapunov exponent is equal to zero), we expect a smooth behavior of the measure.

B. Stochastic dynamics

Consider now a stochastic dissipative dynamical system with additive noise, described by n-first-order differential equations

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \eta \mathbf{G}(t), \qquad (3)$$

where $\eta > 0$ is a small number (noise intensity) and $\mathbf{G}(t)$ is a vector of independently and identically distributed random Gaussian variables, of zero mean and correlations $\langle G_i(t) G_j(t') \rangle = \delta_{ij} \delta(t - t')$. A physical system will normally have a small level η of random noise, so that it can be considered a stochastic process rather than a deterministic one. In a computer study, roundoff errors should play the role of the random noise. For suitable noise and η, the stochastic time evolution, Eq. (3), has a unique stationary measure μ [9]. This is the *natural* (or *physical*) invariant measure defined above.

Inserting Eq. (3) in to Eq. (1)

$$\frac{d\mu(\mathbf{x}(t))}{dt} = [\mathbf{F} + \eta\mathbf{G}(t)] \cdot \nabla\mu. \tag{4}$$

Whenever the vector field $\mathbf{F}(\mathbf{x})$ can be expressed as $\mathbf{F}(\mathbf{x}) = -\nabla\phi(\mathbf{x}) + \mathbf{f}(\mathbf{x})$, with $\mathbf{f}(\mathbf{x})$ being orthogonal to the gradient term and having no divergence, the measure is given by [17,18]:

$$\mu(\mathbf{x}) = N\exp\left(-\frac{\phi(\mathbf{x})}{\eta^2}\right). \tag{5}$$

Then, introducing Eq. (5) in to Eq. (4) we arrive at

$$\frac{d[\ln\mu(\mathbf{x}(t))]}{dt} = \frac{1}{\eta}\left(\frac{1}{\eta}|\mathbf{F}(\mathbf{x})|^2 + \mathbf{G}(t)\cdot\mathbf{F}(\mathbf{x})\right). \tag{6}$$

In Sec. II A we discussed the smoothness of the measure along the trajectory for a deterministic system, at least in the case where there exists a SRB measure. In the present case, that is, a stochastic process, it is possible to define a unique stationary measure that will tend to the SRB for $\eta \to 0$. This is the *Kolmogorov measure* [9]. At each time step of the evolution, we add some noise (of amplitude η). If we repeat these operations (evolving by time evolution and putting some noise) again and again then we will get an invariant measure which is smooth along the unstable direction. This is because the deterministic part of the time evolution will improve the continuity of the density in the unstable direction by stretching, and roughening it in the other directions due to contraction. Thus, in the zero-noise limit $\eta \to 0$ we will get a measure μ that satisfies SRB conditions.

Equation (6) provides an alternative tool to investigate the findings of [4], according to which smoothness in phase space implies determinism in the time series. For weak noise levels, the first term of the right hand side of Eq. (6) is dominant over the second term. In this case, smoothness in phase space implies ''continuity'' in the left hand side of Eq. (6), or differentiability of the measure along the trajectory. On the other hand, in the case of strong noise levels, the second term is the dominant one and a wild behavior in the measure must be expected. This is so because the vector $\mathbf{G}(t)$ is uncorrelated with the actual position in the phase space. Although Eq. (6) was derived for a vector field obeying some restrictions (see above), our numerical results indicate that it can be applied to more general systems.

C. Time series and measure projection

In the case where dynamical invariants are to be estimated from the knowledge of a single observable, we can rely upon Takens' Delay Coordinate Map Theorem [19]. According to it, the structure of the attractor (including differential information) can be preserved by using delay coordinate maps [20], as long as the embedding dimension m is large enough to fully unfold the attractor structure. Furthermore, the Fractal Whitney Embedding Prevalence Theorem [21] tells us that almost every smooth map will be an embedding (one-to-one and differential structure preserving) provided that $m > 2D$, where D is the box-counting dimension of the at-
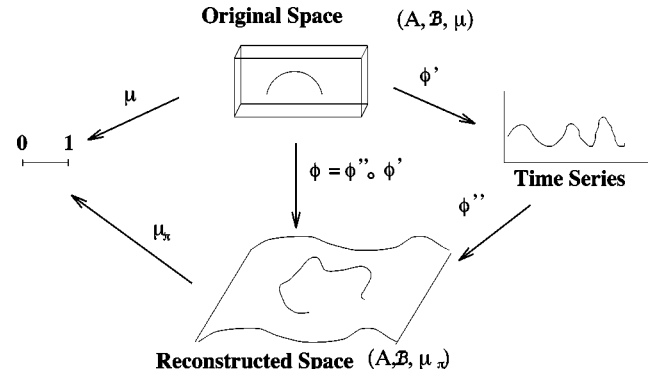

FIG. 2. Scheme followed to analyze a dynamical system and to calculate the natural measure μ_π on the reconstructed space.

tractor. However, as shown in [22,23], when dealing with functions of the measure, embedding dimensions greater than the correlation dimension are enough.

From the statistical point of view, we must ask how the embedding process in the measure space $(A, \mathcal{B})$ is associated with the attractor A. That is, if μ is a measure in $\mathfrak{R}^n$ and $\phi: \mathfrak{R}^n \to \mathfrak{R}^m$ is a function, we will have a projected or *induced measure* $\mu_\pi = \phi(\mu)$ over a subset $S \in \mathfrak{R}^m$, where $\phi(\mu)(S) = \mu(\phi^{-1}(S))$. This is important because we are actually transforming the measure in the embedding process. Most of the existing literature is related to projections and reconstructions of the invariants sets and the dimensions of the corresponding probability measures [24–26]. In our case, if μ is the physical measure describing the original system (μ is carried by an attractor in phase space) then the points in the reconstructed space are equidistributed with respect to the projected measure μ_π [9] (except for particular cases avoided explicitly by the embedding theorems [21]). This is a sufficient condition that allows us to extract information about the original system working in its corresponding projection.

If we have a time series, we can construct an m-dimensional vector

$$\bar{x}_i = (x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\cdot\tau}),$$

where τ is the time delay, chosen by one of the standard methods [8]. Then, from a single set of observations, multivariate vectors in m-dimensional space are used to trace out the orbits of the system. Figure 2 shows the concept of the reconstruction method. Explicitly shown is the reconstruction of the measure carried by the attractor μ_π.

D. Statistics of topological properties

Taken's theorem [19], and its sequels, give a rigorous justification for state space reconstruction. The essence of the mathematical proof is that the trajectory formed from the time series is diffeomorphically related to the actual phase-space trajectory of the dynamical system. In order to test the mathematical properties embodied in the diffeomorfism, that is, continuity, differentiability, inverse differentiability and injectivity, Pecora, Carroll, and Heagy [13] have developed a set of statistics intended to test quantitatively these properties. Their algorithms are of general use and can in particular be applied to test topological properties in any pair of sets of

points. We will take advantage of this procedure by implementing the aforementioned algorithms to test numerically the continuity of the logarithmic derivative of the measure along the trajectory. Basically, the method is intended to evaluate, in terms of probability or confidence levels, whether two data sets are related by a mapping having the continuity property: A function f is said to be continuous at a point $\mathbf{x}_0$ if $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\|\mathbf{x} - \mathbf{x}_0\| < \delta \Rightarrow \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \epsilon$. The results are tested against the null hypothesis, specifically, the case in which no functional relation between points along the trajectory and the measure exists. This is done by means of the statistics proposed by Pecora, Carroll, and Heagy [13]

$$\Theta_{C^0}(\epsilon) = \frac{1}{n_p} \sum_{j=1}^{n_p} \Theta_{C^0}(\epsilon, j) \qquad (7)$$

and

$$\Theta_{C^0}(\epsilon, j) = 1 - \frac{p_j}{p_{max}}, \qquad (8)$$

where p_j is the probability that all of the points in the δ set, around the point $\mathbf{x}_j \in \mathbf{x}(t)$, fall in the ϵ set around $d \ln \mu(\mathbf{x}_j)/dt$. The likelihood that this will happen must be relative to the most likely event under the null hypothesis, p_{max} (see [13]). When $\Theta_{C^0}(\epsilon, j) \approx 1$ we can confidently reject the null hypothesis, and assume that there exists a continuous function. As in the work of Pecora, Carroll, and Heagy [13] the ϵ scale is relative to the standard deviation of the density time series, and thus, $\epsilon \in [0,1]$. Plots of $\Theta_{C^0}(\epsilon)$ vs ϵ can be used to quantify the degree of statistical continuity of a given function. In order to characterize the continuity statistics by means of a single parameter we have also calculated

$$\theta = \int_0^1 \Theta_{C^0}(\epsilon) d\epsilon. \qquad (9)$$

The limiting values of θ, namely, 0 and 1, correspond to a strongly discontinuous and a fully continuous function, respectively.

E. Summary

A naive test to quantify noise in signals is to check how smooth they are. As long as more noise contaminates the signal, the more discontinuous it will become. This is the case for example of additive noise, e.g., noise added to the signal. However, this is by no means a general rule. For instance, intrinsic noise, that is, noise added in the equation of motion, is not expected to affect the smoothness of the signal. This can be clearly seen in the case of surrogate time series: two time series (the original series and its surrogate) having the same correlation structure, may have very different underlying dynamics, i.e., one deterministic and the other stochastic. We can overcome the above drawback by using the trajectory of the system, instead of a single variable. Smoothness or continuity of the trajectory in phase space has been used before [2] in this context. What we propose here is to use the distribution of points on the trajectory (or the

natural measure) as a way to evaluate the the degree of noise in the system. Our proposal is based upon Eq. (6). This equation tells us that the amount of noise present in the system is directly related to the differentiability of the natural measure, evaluated along a typical trajectory. The noisier the system is the less differentiable it becomes. It remains a question of how to apply a fundamental analysis concept, like differentiability, to a magnitude defined numerically. The ''topological statistics'' tools developed by Pecora, Carroll, and Heagy [13] were helpful. They were devised to evaluate in a statistical fashion topological properties of functions. Equation (8) is the statistics used to test, against the null hypothesis, the degree of continuity of our function. As long as this statistic approaches unity we are confident that we have a continuous function. We should note that we have evaluated the continuity of the numerical derivative of the measure, which is equivalent to test differentiability, a computationally easier procedure. Continuity is directly related to the resolution with which we are looking at the function, that is ϵ, so the statistics is actually ϵ dependent $\Theta_{C^0}(\epsilon)$. In order to provide us with an overall continuity test, we summed up over the whole range of ϵ obtaining a single parameter θ [see Eq. (9)] which we use hereafter to characterize continuity.

III. MODELS AND NUMERICAL PROCEDURES

Here we discuss the various stochastic dynamical systems on which we have investigated the efficiency of our approach. The numerical procedures followed to reconstruct the space from time series related to some of the coordinates of those models and to evaluate the statistical differentiability of the natural measure are also described. We also discuss a variant of the surrogation process which consists of producing a partially surrogate series. The method is a way to vary the degree of stochasticity of the time series.

A. The stochastic Van der Pol oscillator

The simplest case in which we can apply our ideas is in a nonlinear oscillator. We have used the Van der Pol (VdP) oscillator [27] with an additive stochastic term

$$\dot{x} = y + \eta G_1,$$
$$\dot{y} = -(x^2 - 1)y - x + \eta G_2. \qquad (10)$$

The parameter η represents the noise level, and $G_i(t)$ are uncorrelated Gaussian noises, such that $G_i(t) \in$ normal$(0, \sigma)$, zero mean and standard deviation σ. Without loss of generality we take $\eta = 1$, and tune the degree of noise only by the standard deviation σ.

Numerical integration was carried out by means of an Euler algorithm. 20000 data points have been generated. Using the x coordinate as our ''experimental'' time series, we have made a reconstruction with an embedding dimension in the range 2–20, and a τ lag of 10 (in sampling units). Every reconstruction has been rescaled to the unit hyper-square.

Density time series have been obtained for each reconstruction. Starting with the first point in the reconstructed phase space, we follow the trajectory recording the density of points around each of the trajectory's points. In order to estimate this density, we have used the Epanechnikov kernel
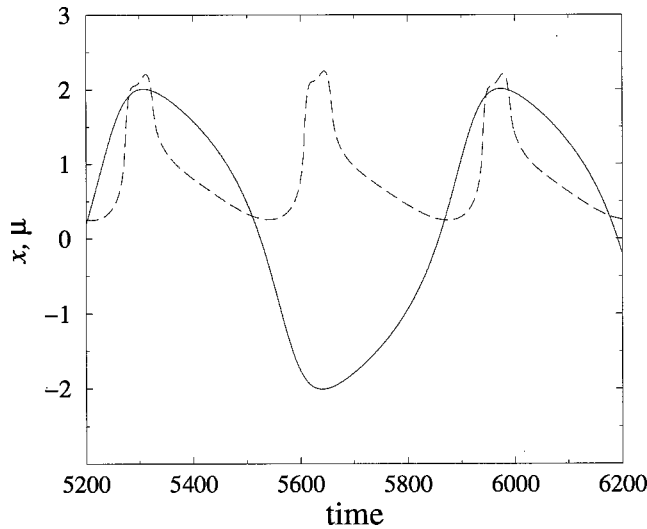
FIG. 3. Typical records of the x coordinate (continuous line) and of the natural measure (broken line) time series (in arbitrary units) from the x coordinate of the Van der Pol oscillator. The results correspond to the parameters given in the text whithout stochastic term, and an embedding dimension of 4.

[28], which, roughly speaking,''weighs'' the points according to its distance to the center. This is preferable to the Gaussian kernel, which is of infinite support (and therefore has a lower computational efficiency) and the ''square'' kernel which gives equal status to the different points in the ball around the reference point. A parameter that has to be chosen carefully is the ball radius used to estimate the density. If the radius is too small, the low density regions will be practically depopulated and the measure will be underestimated. On the other hand, if the radius is too big, the estimation will capture points which are not really part of the neighborhood of the reference point. Of course, the ball radius is a function of the data points being used in the reconstruction process and must be chosen according to this fact. We have used a radius of the ball of 5% of the total attractor extent. In evaluating the continuity statistics, we average $\Theta_{C^0}(\epsilon, j)$ over n_p points [see Eq. (6)] randomly distributed in the trajectory, typically 10% of the total record.

Figure 3 shows a typical density time series from the x coordinate of the VdP oscillator. Albeit qualitative, the smooth behavior of this density along the trajectory is readily noted.

B. The stochastic Lorenz system

The VdP oscillator will allow us to get a closer look at the procedure we are implementing. However, a nonlinear oscillator is a somewhat trivial example and we want to apply the method to more complicated cases. In fact, the ultimate objective of the methodology is to discriminate random behavior from a deterministic one, and this is specially important in the case of chaotic behavior.

The Lorenz system [29] with an additive stochastic term is an adequate choice. The related system of differential equations can be written as

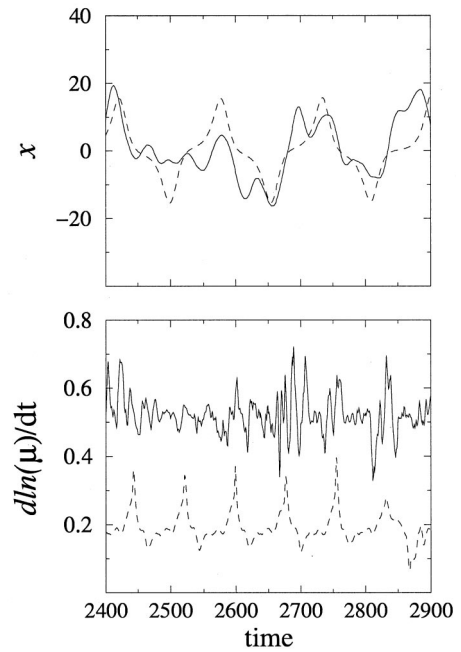$$\dot{x} = -sx + sy + \eta G_1,$$

FIG. 4. Illustrates the differences between using a coordinate or the time derivative of the natural measure in the evaluation of the continuity statistics. The results correspond to the Lorenz system without noise: (a) x coordinate (broken line) and its 100% surrogate (continuous line), (b) derivative of the reconstructed natural measure (broken line) from the x coordinate and from the x surrogate surrogate (continuous line).

$$\dot{y} = -y + rx - xz + \eta G_2,$$

$$\dot{z} = -bz + xy + \eta G_3. \tag{11}$$

The parameters used in the calculations are; $s = 10.0$, $r = 28.0$ y, $b = 2.66$, which give chaotic behavior in the case $\eta = 0$. $G_i(t)$ and η were defined above. When not specified the results for the Lorenz system discussed in the following section were obtained for an embedding dimension of 3, which is greater than the correlation dimension of the Lorenz system.

Numerical integration of the Lorenz system was carried out by means of the Euler method. The time integration step was 0.01. Time series with 16 384 data points and their respective surrogates were subsequently generated. The reconstruction was performed by the usual time–delay method [19–21], with a time delay given by the first zero of the autocorrelation estimate (10 in units of the integration step) on an embedded phase space of dimension 3. The natural measure $\mu(\mathbf{x}(t))$ along the trajectory was calculated by means of the Epanechnikov kernel density estimator [28] with an sphere of radius 5% of the attractor extent. As in the VdP oscillator the continuity statistics was evaluated including up to 10% of the points in a given record.

The sensitivity of the time derivative of the measure to stochasticity is illustrated in Fig. 4. This figure shows that whereas the surrogate of the x coordinate time series is as ''smooth'' as the original series, the time derivative of the logarithm of the measure is much more spiked in the surrogate than in the original series.

C. The Mackey–Glass model

In order to investigate the effects of the embedding dimension we have also considered the high dimensional system introduced in [30]. The dynamical system is described by means of the following delay–differential equation:

$$\dot{x} = \frac{ax(t-\delta)}{1+x(t-\delta)^c} - bx(t). \tag{12}$$

This equation has been proposed to model nonlinear feedback control in physiology. We use the set of parameters that gives an attractor dimension of ≈ 7.5 [2], namely, $a=0.2$, $b=0.1$, $c=10$, and $\delta=100$. The resulting time series were analyzed with a time delay given by the first zero of the autocorrelation estimate, and the measure was evaluated on spheres of radius 10% of the attractor extent and the continuity statistics with 10% of the points in the total record.

D. Partial surrogation

The surrogation process is a well established method in the context of nonlinear time series analysis [31,32]. In typical applications a single time series is available. From this time series, an ensemble of the so-called surrogate series are generated that mimic certain properties of the original. For example, by simply scrambling the temporal order of the points in the original, one obtain surrogate time series which preserve the mean, variance, etc. One of the most popular methods of producing surrogate time series consists of shuffling the phases in the Fourier transform of the original data set [31]. In this way, each value of the Fourier transform of the original data is multiplied by a random phase $\exp(i\phi)$, with $\phi \in$ random $[0,2\pi]$. (In order to get a real time series in the antitransformation we multiply symmetrically with respect to the center of the transform.) The procedure generates a new time series with the autocorrelation structure of the original. Here we propose to introduce an additional factor, $\exp(i\phi\alpha)$, with $\alpha \in [0,1]$. This factor allows us to control the degree of stochasticity by tuning the parameter α.

IV. RESULTS

A. Sensitivity of the time derivative of the measure to stochasticity

In order to test the efficiency of the approach proposed here we have first evaluated the continuity statistics either on a coordinate or on the time derivative of the natural measure time series of the stochastic VdP oscillator. Figure 5 shows the continuity statistics for the x coordinate and for the time derivative of the reconstructed natural measure with and without noise. There is almost no modification in the statistics of the coordinate upon noise addition. However, the statistics of the density time series reflects very clearly the presence of the stochastic term. This illustrates the efficiency and novelty of our approach and supports its application to more complex cases. We must remark that we are using the commonly named ''dynamical'' noise, that is, a stochastic term added in the dynamical equations, instead of the ''measurement'' noise, which is added after the ''clean'' integration step. Preliminary results shows that an efficient method to discriminate both types of noise, can be achieved by using
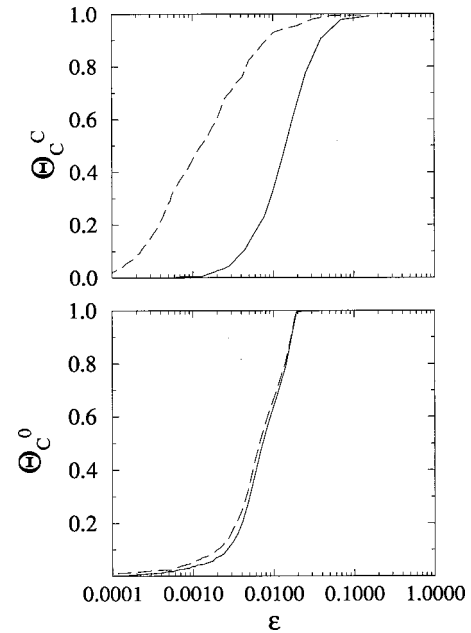

FIG. 5. Van der Pol oscillator: continuity statistics [see Eq. (7)] for the x coordinate (lower frame) and for the time derivative of the measure along the trajectory obtained from the x coordinate (upper frame). Results for the deterministic system (broken curves) and the stochastic system with $\sigma=0.06$ (continuous curves) are shown.

the continuity statistics over the density and the coordinates, but this deserve further research, and eventually will be published elsewhere.

The results for the continuity statistics of the time derivative of the reconstructed measure from the x coordinate of the Lorenz system are illustrated in Fig. 6. The results of Fig. 6(a) show that the time derivative of the measure in the original series is ''more continuous'' (in a statistical sense) than
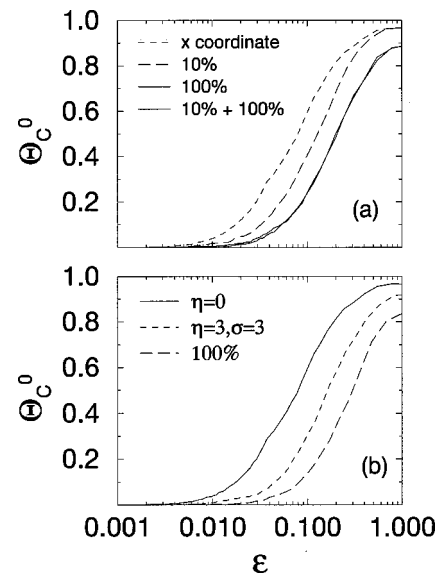

FIG. 6. Continuity statistics [see Eq. (7)] for the time derivative of the measure along the reconstructed trajectory from the x coordinate of the Lorenz system. (a) Results for the original time series, and for the series partially (10%) or totally surrogated (100%), and a combination of both. (b) Results for the Lorenz system with noise [see Eq. (8)] and for its surrogate series.
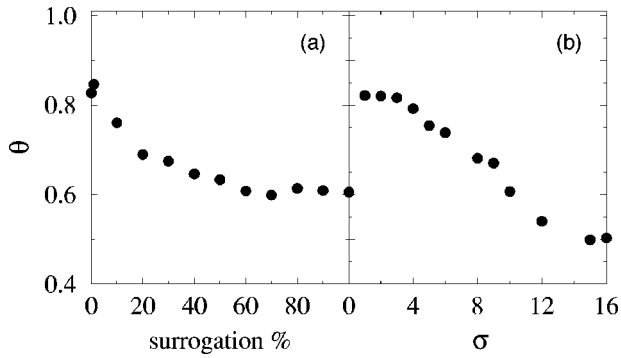
FIG. 7. Integral of the continuity statistics [as defined in Eq. (8)] for time series derived from the Lorenz system. The results correspond to: (a) partially randomized series with increasing degree (percentage) of randomization and (b) stochastic Lorenz system with increasing standard deviation of the Gaussian noise in Eq. (8).

its surrogate. Partial surrogation (10%) decreases the degree of continuity of the time derivative of the measure in an extent lower than total surrogation, as expected. On the other hand, the results show that the continuity of the totally surrogated series shows almost no dependence on whether it has been derived from the original series or from a partially surrogated series (10% surrogated).

The results for the stochastic Lorenz system reported in Fig. 6(b) clearly show that the stochastic terms significantly decrease the statistical continuity of the time derivative of the measure. Surrogation of the stochastic series produces a further decrease of continuity, indicating that the series still has some degree of determinism. The degree of stochasticity of a time series can be quantified by calculating the integral of the continuity statistics as defined in Eq. (8). Figure 7(a) shows how steeply θ decreases with the percentage of surrogation. Similarly, θ decreases with the standard deviation of the Gaussian noise in the stochastic Lorenz system [Fig. 7(b)], as expected. Thus, the magnitude θ can be used to evaluate the relative stochasticities of a set of experimental time series.

B. Dependence on the embedding dimension

A point of crucial relevance is how the above results change with the embedding dimension m. We have investigated this question on the Lorenz system and on the high–dimensional system discussed in Sec. III D [30]. The results for the Lorenz system depicted in Fig. 8(a) show that θ decreases with the embedding dimension. This is a consequence of working with a fixed sphere radius for all m and of the numerical noise that should increase with m. The decrease of θ is stronger in the surrogate series, although it is likely that the difference between the two should decrease for large enough m. In any case, the difference in θ between the original and the surrogate series changes only from 0.35 to 0.49 when m is varied in the range 3–10. The behavior of θ in the high–dimensional system is far more intricate [see Fig. 8(b)]. For m well below the attractor dimension the measure for the surrogate series seems to be more continuous than that for the original series. The reason for this rather odd behavior has to be found in the heavy crossing of trajectories that occur at m far below the attractor dimension [8]. In those cases, surrogation seems to have a smoothing effect. Instead,
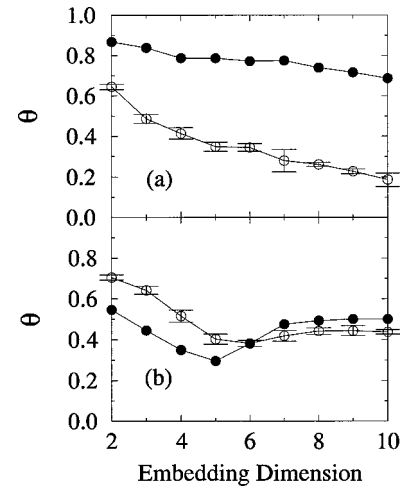

FIG. 8. Integral of the continuity statistics [see Eq. (8)] as a function of the embedding dimension for time series (filled symbols), and their surrogates (empty symbols), derived from: (a) the Lorenz system and (b) the high–dimensional system proposed in [30]. The error bars in the results for the surrogate series account for averages over five realizations.

for $m > 7$ the behavior is similar to that of the Lorenz system, although the difference between the original and the surrogated records is substantially smaller. This points to further study that is actually in progress.

Figure 9 illustrates the dependence of the continuity statistics of the measure on the embedding dimension m for the stochastic VdP oscillator. In this case the simplicity of the attractor gives an almost null dependence on m for no noise or very low noise levels. When the noise is increased the behavior is in line with that found in the Lorenz system, namely, a decrease of θ as m increases. This dependence on m is more noticeable the greater the noise level.

The study of the dependence of the continuity statistics of the measure derivative on the embedding dimension involves numerical difficulties that deserve some comments. In the case of low enough embedding dimension and moderate
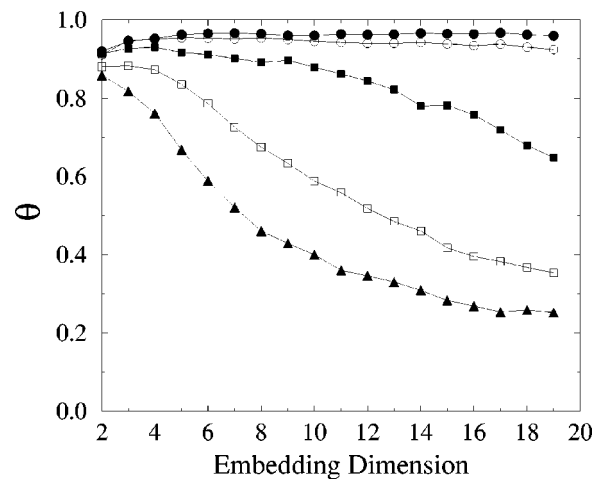

FIG. 9. Integral of the continuity statistics [see Eq. (8)] as a function of the embedding dimension for time series derived from the the x coordinate of the Van der Pol oscillator with various degrees of noise: $\sigma = 0$ (filled circles), 0.01 (open circles), 0.03 (filled squares), 0.06 (open squares), and 0.09 (filled triangles).
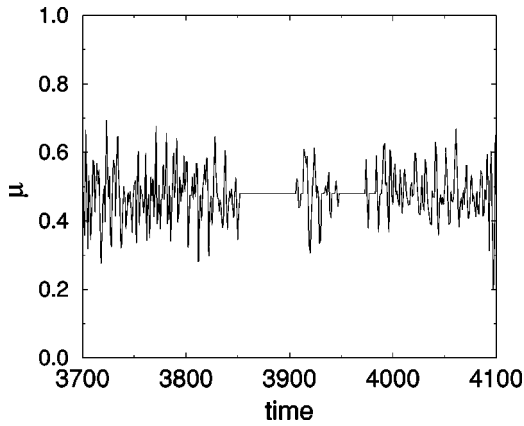
FIG. 10. Measure estimate in the case of a stochastic Van der Pol oscilator (x coordinate) for a embedding dimension of 19, and $\sigma = 0.5$.

number of data points, the estimation can be achieved confidently for almost all the points in the trajectory. Of course, ''moderate'' and ''long enough'' are terms that depend on the attractor dimension. We think that the same criteria followed, to get for example a reliable estimation of the correlation dimension [8], apply in this case. In our approach we have introduced noise explicitly, adding another factor to be considered in the estimation step. As it is well known, as ''extra'' dimensions are included in the embedding process, noise populates them more or less uniformly. This is specially problematic in the case of ''simple'' systems, like the VdP oscillator, as the trajectory in a stochastic oscillator may ''wander'' far from the zero-noise limit cycle. In these excursions, the measure swept by the trajectory is unavoidably constant, because no other points are in the neighborhood of the evolution, except those which are time correlated. In such a case, a constant measure results, and thus a high value of the continuity statistics. This effect is more noticeable as the embedding dimension is increased, because ''more'' space is available (see Fig. 10).

C. Application to mixed time series

A distinctive feature of our method is the possibility of using it in different ranges of a given time series. In this way we can examine short records and evaluate their stochasticity. Bearing this in mind we have devised the following example: Suppose we have a time series which is half deterministic and half stochastic. Could our method discriminate both behaviors in the same time series?. In order to answer this question, we have generated a single time series (16 384 points) with the first half coming from the x coordinate of the deterministic Lorenz system, and the second half coming from its surrogate (100% randomization) time series. We have applied the continuity statistics over four regions in the time derivative of the density record (two randomly selected in the first half and two in the second half). Figure 11 shows the results. It is then clear that the statistics utilized here can discriminate stochastic from deterministic behavior. Figure 11 also shows the statistic for the whole time series (same number of reference points randomly selected along the time series). The results are midway between those for the stochastic and deterministic ranges.
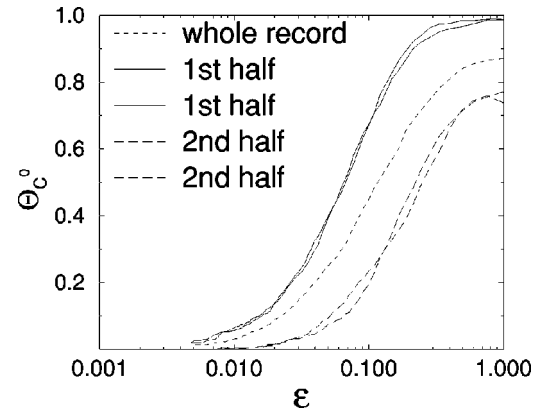

FIG. 11. Continuity statistics for the time derivative of the measure corresponding to the x coordinate of the Lorenz system. The time series was formed by joining the original time series (first half) to its fully randomized series (second half).

As another example of the use of our methodology we have investigated the effect of a noise burst in the VdP oscillator. The dynamic equations are rewritten as

$$\dot{x} = y + A(t)\eta G_1,$$
$$\dot{y} = -(x^2 - 1)y - x + A(t)\eta G_2, \qquad (13)$$

where $A(t) = 1$ if $t_1 < t < t_2$ and 0 otherwise. The interval $[t_1, t_2]$ is a small interval where we ''turn on'' the stochastic term. The idea behind this system is to test the capabilities of the method to detect the noise introduced. Using 20 000 data points, we have used 500 consecutive points with the stochastic term added. In Fig. 12 we show a typical time series (and the measure time derivative) where noise has been turned on in the time interval 10 000–10 500, with a strength of $\sigma = 0.05$. Figure 13 shows Pecora, Carroll, and Heagy [13] statistics applied to five regions of the whole series, one
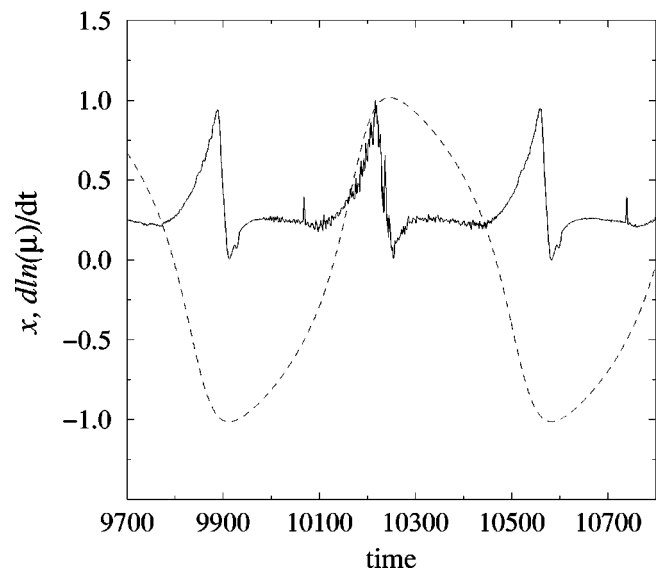

FIG. 12. x coordinate (broken line) and derivative of the natural measure (continuous line) time series of the Van der Pol oscillator in which a noisy burst, of $\sigma = 0.05$ has been introduced in the time interval 10 000–10 500.
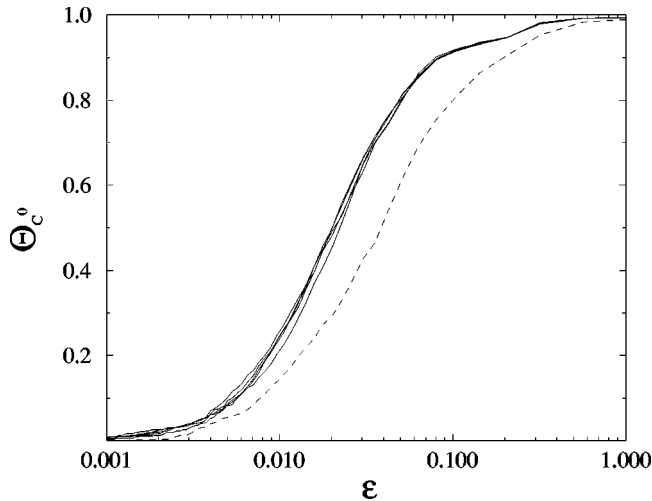
FIG. 13. Continuity statistics for the time series of Fig. 12, applied to five regions of the series one of which is the stochastic region (broken line). The noise intensity is $\sigma = 0.05$.

FIG. 14. Continuity statistics for the BZ reaction time series (solid line) and its surrogate (broken line).

of them being the stochastic region. Again, our method clearly discriminates between noisy and clean regions.

D. Application to experimental time series

In order to test our method in actual experimental time series, we have used the data set from the Belusov–Zhabotinski (BZ) chemical reaction [33]. As shown by those authors the apparently random behavior of the amplitude of the concentration of bromide ions can in fact be explained by deterministic laws. We will use here our approach to confirm the above finding.

Using the bromide concentration time series, as in the work of Roux, Simoyi, and Swinney [33] we repeat the procedure explained above. We have used a $\tau = 30$ (in sampling units) for the reconstruction and an embedding dimension of 4. Figure 14 shows the continuity statistics for both the time derivative of the reconstructed measure using the BZ time series and for its 100% surrogate. The large difference between the two and the rather high value of Θ_{C^0} confirm our proposal in the sense that one can explain the behavior of this record as the output of a deterministic dynamical system.

V. CONCLUDING REMARKS

In brief, we have proposed a method to identify determinism in time series, which exploits the continuity of the logarithmic time derivative of the natural measure along the trajectory, that is, its differentiability. The method is based upon a formula which explicitly shows the sensitivity of the measure to stochasticity. In the present work we have adapted the statistical method of Pecora, Carroll, and Heagy [13] to investigate the continuity of the time derivative of the measure. Results of partially surrogated series and series derived from two stochastic dynamical systems and a high–dimensional system, clearly illustrate the suitability of the present method to the problem at hand. As we have shown,
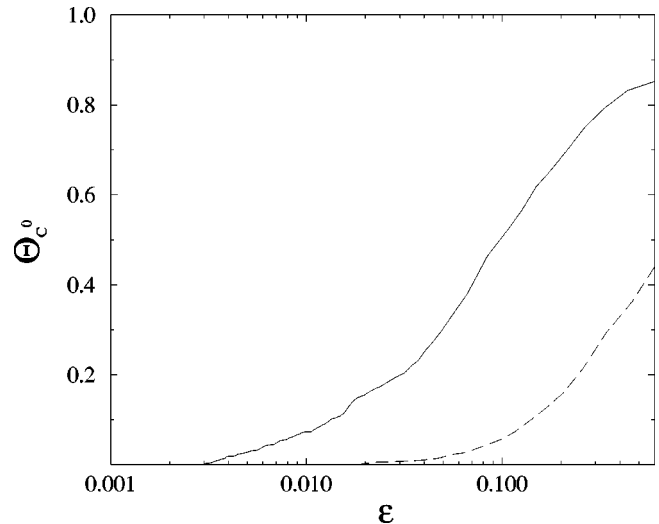
the method is very easily applied to any kind of time series, from the simplest one, as is the case of limit cycle oscillators, to the high–dimensional cases. Given a time series, and once done the standard reconstruction process, obtaining the density (or its time derivative) is a straightforward procedure. Then, one can use the Pecora, Carroll, and Heagy approach over the whole time series or in selected pieces of it. A warning against a blind application of the method is in order, particularly in what concerns the evaluation of the continuity of the measure time derivative. A careful inspection of the density time series must be done before any further operation is performed.

The dependence of the continuity statistics on the embedding dimension in low- and high–dimensional systems, indicate that applications to real (experimental) time series would eventually require a thorough investigation of this point in each particular case, as is common in time series analysis. In any case, the fact that the method works reasonably well on short time series, supports its usefulness for the analysis of experimental series. We have shown this in a simple case as it is the BZ reaction, confirming previous findings.

In a broader sense, the application of continuity statistics over the density time series is a new aspect of the possibilities offered by the Lagrangian measures [16]. As we have shown previously, the use of more traditional tools on this density, such as Fourier transforms or histograms, may help in extracting new information on the underlying dynamical system.

[1] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).

[2] D.T. Kaplan and L. Glass, Phys. Rev. Lett. **68**(4), 427 (1992); Physica D **64**, 431 (1993).

[3] R. Wayland, D. Bromley, D. Pickett, and A. Passamante, Phys. Rev. Lett. **70**(3), 580 (1993).

[4] L.W. Salvino and R. Cawley, Phys. Rev. Lett. **73**(8), 1091 (1994).

[5] G. Sugihara and R. May, Nature (London) **344**, 734 (1990).

[6] A. Osborne and A. Provenzale, Physica D **35**, 357 (1989).

[7] J. Theiler, Phys. Lett. A **155**, 480 (1991).

[8] H. Abarbanel, R. Brown, J. Sidorowich, and L. Tsimring, Rev. Mod. Phys. **65**(4), 1331 (1993).

[9] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**(3), 617 (1985).

[10] A. Lasota and M. Mackey, *Probabilistic Properties of Deterministic Systems* (Cambridge University Press, Cambridge, 1985).

[11] R. Gilmore, Rev. Mod. Phys. **70**(4), 1455 (1998).

[12] M. Casdagli, S. Eubank, D. Farmer, and J. Gibson, Physica D **51**, 52 (1991).

[13] L. Pecora, T. Carroll, and J. Heagy, Phys. Rev. E **52**(4), 3420 (1995).

[14] G. Ortega and E. Louis, Phys. Rev. Lett. **81**, 4345 (1998).

[15] G. Ortega, Phys. Lett. A **209**, 351 (1995).

[16] G. Ortega, Phys. Rev. Lett. **77**(2), 259 (1996).

[17] R. Graham and T. Tél, Phys. Rev. Lett. **52**(1), 9 (1984).

[18] M. San Miguel and R. Toral, in *Instabilities and Nonequilibrium Structures*, edited by E. Tirapegui (Kluwer Academic, 1997) Vol. IV.

[19] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics, Vol. 898 (Springer, Berlin, 1981).

[20] N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).

[21] T. Sauer, J. Yorke, and M. Casdagli, J. Stat. Phys. **65**(3/4), 579 (1991).

[22] M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. Yorke, Physica D **69**, 404 (1993).

[23] C. Schroer, T. Sauer, E. Ott, and J. Yorke, Phys. Rev. Lett. **80**(7), 1410 (1998).

[24] R. Kaufman, Mathematika **15**, 153 (1968).

[25] P. Mattila, Ann. Acad. Sci. Fenn., Ser. AI: Math. **1**, 227 (1975).

[26] T. Sauer and J. Yorke, Ergodic Th. Dyn. Syst. **17**, 941 (1997).

[27] B. Van der Pol and J. Van der Mark, Philos. Mag. **6**, 763 (1928).

[28] V.A. Epanechnikov, Theor. Probab. Appl. **14**, 153 (1969).

[29] E. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

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

[31] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer, Physica D **58**, 77 (1992).

[32] T. Schreiber, Phys. Rev. Lett. **80**, 2105 (1998).

[33] J.C. Roux, R.H. Simoyi, and H.L. Swinney, Physica D **8**, 257 (1983).