Smoothness Implies Determinism in Time Series: A Measure Based Approach

Guillermo J. Ortega and Enrique Louis

Departamento de Física Aplicada, Universidad de Alicante, Apartado 99, E-03080 Spain

(Received 26 January 1998)

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 the stochastic Lorenz model illustrate the usefulness of the method proposed here. The method is shown to work also for high-dimensional systems. [S0031-9007(98)07686-8]

PACS numbers: 05.45.+b, 02.30.Cj, 07.05.Kf

There has been a great interest in the last years aimed to detect "determinism" in time series [1–4]. Though experimental noise is a common problem in many areas of physics, it can be especially troubling in the field of nonlinear time series analysis. For instance, attention has been called to situations where noise can mimic or mask deterministic (e.g., chaotic) behavior, when the dynamics of the system is characterized by means of classical measures of chaos such as Lyapunov exponents, K_2 entropy, and correlation dimension [5,6] (see also [7] for a review of these methods and some of their pitfalls). In searching more reliable and robust methods, different aspects of the vector field have been investigated. In particular, it has been suggested [1,3] that the continuity of the vector field is a clear hallmark of determinism. In this work we exploit continuity in phase space although from a different point of view.

In earlier work [8,9] we used the natural invariant measure along the system trajectory to detect hidden periodicities in the reconstructed phase space. Here we extend this technique to stochastic dynamical systems and show how one can quantify the degree of stochasticity in time series in terms of the continuity of that density. It should be noted that although the numerical estimation of the measure presents serious difficulties, it is more reliable and easier than a similar estimation of the vector field. The procedure is based upon a formula that explicitly shows the sensitivity of the measure to stochasticity.

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

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \eta \mathbf{G}(t), \tag{1}$$

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, round-off errors should play the role of the random noise. For suitable noise and η, the stochastic time evolution (1) has a unique stationary measure μ [10]. This is the *natural* (or *physical*) invariant measure, which gives the limiting distribution of almost all initial conditions.

In the deterministic case ($\eta = 0$) the "material derivative" of the measure (time derivative along the trajectory) can be expressed as [9]

$$\frac{d\mu(\mathbf{x}(t))}{dt} = \dot{\mathbf{x}} \cdot \nabla \mu \,. \tag{2}$$

Now, the same arguments can be applied to Eq. (1), namely,

$$\frac{d\mu(\mathbf{x}(t))}{dt} = [\mathbf{F} + \eta \mathbf{G}(t)] \cdot \nabla \mu \,. \tag{3}$$

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 [11,12]

$$\mu(\mathbf{x}) = N \exp\left(-\frac{\phi(\mathbf{x})}{\eta^2}\right). \tag{4}$$

Introducing (4) in (3) we arrive at

$$\frac{d[\ln \mu(\mathbf{x}(t))]}{dt} = \frac{1}{\eta} \left[\frac{1}{\eta} |\mathbf{F}(\mathbf{x})|^2 + \mathbf{G}(t) \cdot \mathbf{F}(\mathbf{x}) \right]. \tag{5}$$

Equation (5) provides an alternative tool to investigate the finding of [3], according to which smoothness in phase space implies determinism in time series. For weak noise levels, the first term of the right-hand side of (5) is dominant over the second term. In this case, smoothness in phase space implies "continuity" in the left-hand side of (5), 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 term must be expected. This is so because the vector $\mathbf{G}(t)$ is uncorrelated with the actual position in the phase space. Although Eq. (5) was derived for a vector field obeying some restrictions (see above), our numerical results indicate that it can be applied to more general systems.

In order to test numerically the continuity of the logarithmic derivative of the measure along the trajectory we have implemented the method proposed by Pecora *et al.* [13] to carry out statistical evaluations of continuity and differentiability of functional mappings. In our opinion this method is the best suited for our purposes. Basically, this is a statistics 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 *et al.* [13]

$$\Theta_{C^0}(\epsilon) = \frac{1}{n_p} \sum_{j=1}^{n_p} \Theta_{C^0}(\epsilon, j) \tag{6}$$

and

$$\Theta_{C^0}(\epsilon, j) = 1 - \frac{p_j}{p_{\max}}, \tag{7}$$

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 at random in the ϵ set around $\frac{d \ln \mu(\mathbf{x}_j)}{dt}$. The likelihood that this will happen must be relative to the most likely event under the null hypothesis, $p_{\max}$ (see Ref. [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 *et al.* [13] the ϵ scale is relative to the standard deviation of the density time series, and $\epsilon \in [0, 1]$. Plots of $\Theta_{C^0}(\epsilon)$ versus 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

$$\theta = \int_0^1 \Theta_{C^0}(\epsilon) d\epsilon \,. \tag{8}$$

The limiting values of Θ_{C^0}, namely, 0 and 1, correspond to a strongly discontinuous and a fully continuous function, respectively.

The preceding procedure has been first tested on the Lorenz system. We investigate the effects of tuned stochasticity either by introducing an additive stochastic term in the Lorenz system or by partial surrogation of time series derived from the deterministic Lorenz system. The latter is done by replacing the factor $\exp(i\phi)$, with $\phi \in$ random $[0, 2\pi]$, commonly introduced in the shuffling step of Fourier phases [14], by $\exp(i\phi\alpha)$, with $\alpha \in [0, 1]$. In this way, the "degree" of randomization of the Fourier phases is varied from 0% to 100%. This study is carried out for an embedding dimension of three, which is greater than the correlation dimension of the Lorenz attractor [15]. The effects of changing the embedding dimension are investigated on time series generated from either the Lorenz or a high-dimensional system and their surrogates. Finally, we show how the method can be applied to mixed series containing both deterministic and stochastic regions.

The Lorenz system [19] with an additive stochastic term can be written as

$$\begin{aligned} \dot{x} &= -\, sx + sy + \eta G_1 \,, \\ \dot{y} &= -\, y + rx - xz + \eta G_2 \,, \\ \dot{z} &= -\, bz + xy + \eta G_3 \,. \end{aligned} \tag{9}$$

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. The parameters used in the calculations are $s = 10.0$, $r = 28.0$, and $b = 2.66$, which give chaotic behavior in the case $\eta = 0$.

Numerical integration of the Lorenz system was carried out by means of the modified 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 [18,20,21], with a time delay given by the first zero of the autocorrelation estimate and on an embedded phase space of dimension three. The natural measure $\mu(\mathbf{x}(t))$ along the trajectory was calculated by means of the Epanechnikov kernel density estimator [22] with a sphere of radius 5% of the 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.

The results for the continuity statistics of the time derivative of the measure corresponding to the reconstructed attractor from the x coordinate of the Lorenz system are illustrated in Fig. 1. The results of Fig. 1a show

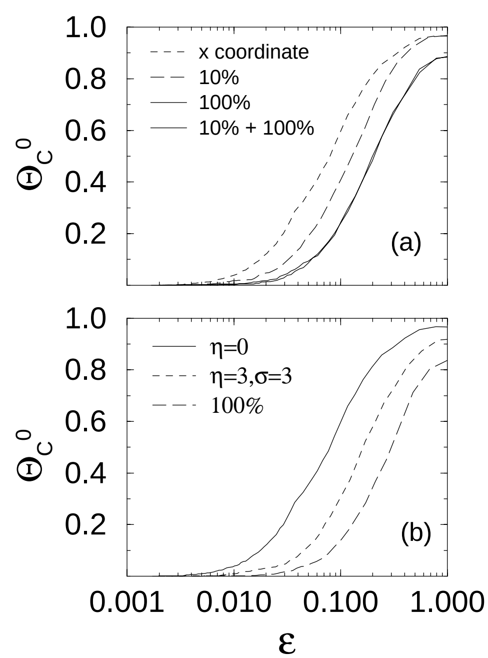

FIG. 1. Continuity statistics as defined in Eq. (7) for the time derivative of the measure along the trajectory corresponding to the reconstructed attractor from the x coordinate of the Lorenz system. (a) Results for the original time series, for the series partially (10%) or totally surrogated (100%), and for a combination of both. (b) Results for the Lorenz system with noise [see Eq. (9)] and for its surrogate series.

that the time derivative of the measure in the original series is "more continuous" (in a statistical sense) than in 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 show almost no dependence on whether it has been derived from the original series or from a partially surrogated series. The results for the stochastic Lorenz system reported in Fig. 1b clearly show that the stochastic terms in the Lorenz system 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 2a shows how steeply θ decreases with the percentage of surrogation. Similarly, θ decreases with the standard deviation of the Gaussian noise in the Lorenz system (Fig. 2b), as expected (without loss of generality we take $\eta = 1$, and tune the degree of noise only by the standard deviation σ). Thus, the magnitude θ can be used to evaluate the relative stochasticities of a set of experimental time series.

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 proposed in [23]. In the latter case we used the set of parameters that gives an attractor dimension of $\approx$7.5 [1]. 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. The results for the Lorenz system depicted in Fig. 3a 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. The behavior of θ in the high-dimensional system is far more intricate (see Fig. 3b). 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 [7]. In those cases, surrogation seems to have a smoothing effect. Instead, for $m > 6$ the behavior is similar to that of the Lorenz system, although the difference between the original and the surrogated records is substantially smaller. This point deserves further study that is actually in progress.

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 its stochasticity. Bearing this in mind we have devised the following example: Suppose we have a time series which is half deterministc 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 density record (two randomly selected in the first half and two in the second half). Figure 4 shows the results. It is then clear that the statistics utilized here can discriminate stochastic from deterministic

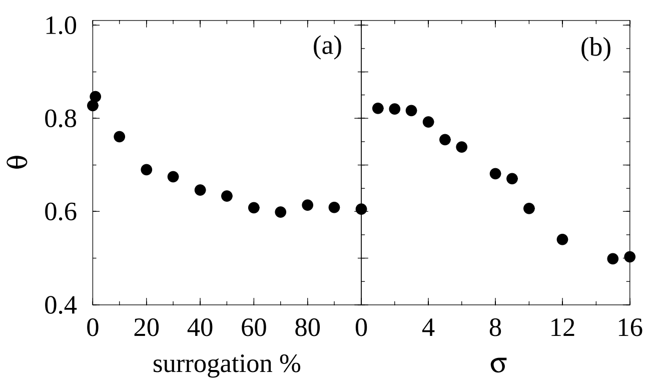

FIG. 2. Integral of the continuity statistics [as defined in Eq. (8)] for time series derived from the Lorenz system. The results correspond to (a) a partially randomized series with increasing degree (percentage) of randomization; and (b) a stochastic Lorenz system with increasing standard deviation of the Gaussian noise in Eq. (9) and $\eta = 1$.

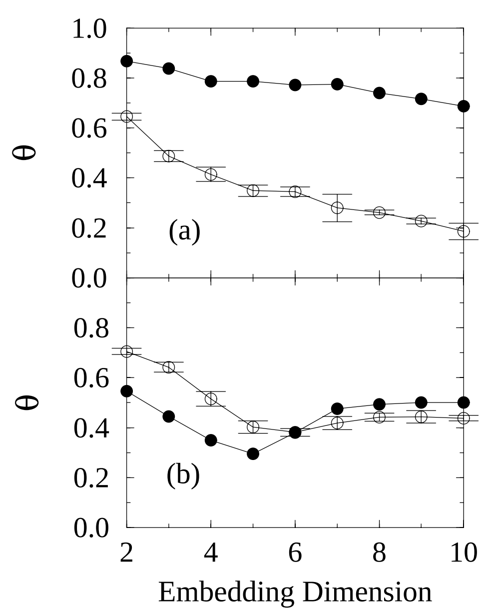

FIG. 3. Integral of the continuity statistics [see Eq. (8)] as a function of the embedding dimension for time series (filled symbols), and their surrogates (empty symbols), derived from (a) the Lorenz system, and (b) the high-dimensional system proposed in Ref. [23]. The error bars in the results for the surrogate series account for averages over five realizations.

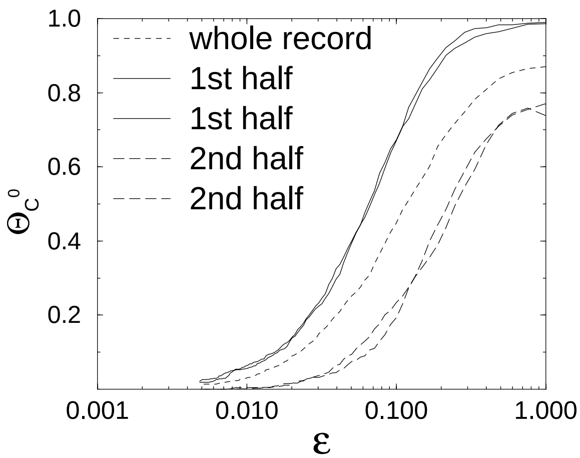

FIG. 4. Continuity statistics for the time derivative of the measure corresponding to the x coordinate of the Lorenz system. The time series was formed by joining the original time series (first half) to its fully randomized series (second half). See text.

behavior. Figure 4 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. This procedure can in principle be used to identify stochastic regions in experimental time series.

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. 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 *et al.* [13] to investigate the continuity of the measure. Results for two types of stochastic time series, namely, partially surrogated series and series derived from the stochastic Lorenz system, clearly illustrate the suitability of the present method to the problem at hand. The dependence of the continuity statistics on the embedding dimension in low- and high-dimensional systems indicates 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. The fact that the method works reasonably well on short time series supports its usefulness for the analysis of experimental series.

Thanks are due to E. Hernández, M. SanMiguel, and R. Toral for many useful comments and suggestions. The financial support of the Spanish CICyT (Grant No. PB96-0085) is gratefully acknowledged. G. O. also thanks the Spanish "Ministerio de Educación y Cultura" for support.

[1] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992); Physica (Amsterdam) **64D**, 431 (1993).
[2] R. Wayland, D. Bromley, D. Pickett, and A. Passamante, Phys. Rev. Lett. **70**, 580 (1993).
[3] L. W. Salvino and R. Cawley, Phys. Rev. Lett. **73**, 1091 (1994).
[4] G. Sugihara and R. May, Nature (London) **344**, 734 (1990).
[5] A. Osborne and A. Provenzale, Physica (Amsterdam) **35D**, 357 (1989).
[6] J. Theiler, Phys. Lett. A **155**, 480 (1991).
[7] H. Abarbanel, R. Brown, J. Sidorowich, and L. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[8] G. Ortega, Phys. Lett. A **209**, 351 (1995).
[9] G. Ortega, Phys. Rev. Lett. **77**, 259 (1996).
[10] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[11] R. Graham and T. Tél, Phys. Rev. Lett. **52**, 9 (1984).
[12] M. San Miguel and R. Toral, in *Instabilities and Nonequilibrium Structures IV*, edited by E.Tirapegui (Kluwer Academic, Dordrecht, 1997).
[13] L. Pecora, T. Carroll, and J. Heagy, Phys. Rev. E **52**, 3420 (1995).
[14] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica (Amsterdam) **58D**, 77 (1992).
[15] As shown in Refs. [16,17], when dealing with functions of the measure, embedding dimensions greater than the correlation dimension are enough, as opposed to the general rule according to which the embedding dimension has to be at least twice the box-counting dimension of the attractor [18].
[16] M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. Yorke, Physica (Amsterdam) **69D**, 404 (1993).
[17] Ch. G. Schroer, T. Sauer, E. Ott, and J. A. Yorke, Phys. Rev. Lett. **80**, 1410 (1998).
[18] T. Sauer, J. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[19] E. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[20] 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).
[21] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[22] V. A. Epanechnikov, Theory Probab. Appl. **14**, 153 (1969).
[23] M. C. Mackey and L. Glass, Science **197**, 287 (1977).