Detecting chaos in pseudoperiodic time series without embedding

J. Zhang,* X. Luo, and M. Small
Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
(Received 5 October 2005; published 24 January 2006)

A different method is proposed to detect deterministic structure from a pseudoperiodic time series. By using the correlation coefficient as a measure of the distance between cycles, we are exempt from phase-space reconstruction and further construct a hierarchy of pseudocycle series that, in turn, preserve less determinism than the original time series. Appropriate statistics are then devised to reveal the temporal and spatial correlation encoded in this hierarchy of the pseudocycle series, which allows for a reliable detection of determinism and chaos in the original time series. We demonstrate that this method can reliably identify chaos in the presence of noise of different sources for both artificial data and experimental time series.

DOI: 10.1103/PhysRevE.73.016216 PACS number(s): 05.45.Tp

An accurate identification of the dynamics underlying a complex time series, i.e., whether the dynamics is generated by a deterministic process or a stochastic one, or both, is of crucial importance in understanding the corresponding physical process, and in turn affects the subsequent model development. To this end, traditional methods primarily rely on the estimation of physically meaningful data observables such as Lyapunov exponents, K_2 entropy, and correlation dimension. Recently, it has been shown that the smoothness or continuity of the vector field is a clear hallmark of determinism in the time series, and various nonlinear methods [1–4] are proposed to investigate the properties of the vector fields from different aspects. In contrast to these nonlinear approaches, the surrogate data method [5] provides a linear statistical test in search of the nonlinear determinism.

One class of real world data that exhibits strong periodic behavior such as the electrocardiogram (ECG), laser output, and the annual sunspot number has aroused great interest due to their close relation to some important natural and physiological systems. And recently, surrogate data generation algorithms [6,7] have been developed for such pseudoperiodic time series. However, many available methods are often incapable of detecting chaos in the presence of strong periodicity, which tends to hide underlying fractal structures [8]. Moreover, the existence of noise, which may mask or mimic the deterministic structure of the time series [9–11], can lead to spurious results for conventional chaotic invariants. The uniquely developed nonlinear approaches [1–4], though more reliable, are usually insufficient for application as direct tests of chaos. Finally, most of these approaches [1–4,6] depend heavily on a good reconstruction of the phase-space geometry of the dynamical system. Since there is no unique way to choose the embedding dimension and the time lag, the accuracy of such methods is hard to guarantee.

In this paper, we present a method to detect determinism from a pseudoperiodic time series, and to determine whether this determinism has chaotic attributes. Unlike most methods that require the reconstruction of the phase space and the Euclidian distance between the phase-space points, with our method the pseudoperiodic time series is divided into successive, nonoverlapping cycles that serve as the basic processing units. We use the correlation coefficient as a measure of the distance between different cycles and rearrange them accordingly, thereby building a hierarchy of pseudocycle series that effectively encodes the innercorrelation in the original time series. Several statistics are then devised to extract information within and across these newly created pseudocycle series, which enables the detection of determinism and chaos. This approach is demonstrated to be robust to both measurement noise (white and colored noise) and dynamical noise, and we show the effectiveness of our method on both artificial and experimental time series.

Throughout the paper, we use the x component of the well-known Rössler system and an experimental laser dataset for illustration, both of which are chaotic and contain obvious periodic component. The laser dataset is the record of the output power of the NH_3 laser available in Santa Fe Competition (Data Set A). Given a pseudoperiodic time series $\{x_i\}_1^n$ of n observations, the first step is to segment the pseudoperiodic time series into m consecutive cycles according to the local minimum (or maximum), denoted as $\{C_1, C_2, \ldots, C_m\}$. For each pair of cycles C_i and C_j $(i,j=1,2,\ldots,m, i \neq j)$ with length l_i and l_j, respectively, we then define the correlation coefficient as follows (without loss of generality, suppose $l_i < l_j$)

$$\rho_{ij} = \max_{l=0,1,\ldots,l_j-l_i} \frac{Cov[C_i(1:l_i), C_j(1+l:l_i+l)]}{\sqrt{V[C_i(1:l_i)]}\sqrt{V[C_j(1+l:l_i+l)]}}, \qquad (1)$$

where $C_i(a:b)$ denotes the segment between the ath and the bth element in C_i and Cov stands for covariance. This definition means that if two cycles are not of the same length, we will shift the shorter cycle C_i onto the longer one C_j for $(l_j - l_i)$ steps, calculate one correlation coefficient between C_i and the corresponding part of C_j in each step, and pick out the largest one as the correlation coefficient between C_i and C_j.

The correlation coefficient ρ_{ij} characterizes the similarity between cycle C_i and C_j. The larger the ρ_{ij}, the higher the level of similarity. Considering the continuity and smoothness of the vector fields of deterministic systems, two cycles with a larger ρ_{ij} will also be close in the phase space, see Fig. 1(a) for an illustration. The Rössler system is given by

*Electronic address: enzhangjie@eie.polyu.edu.hk

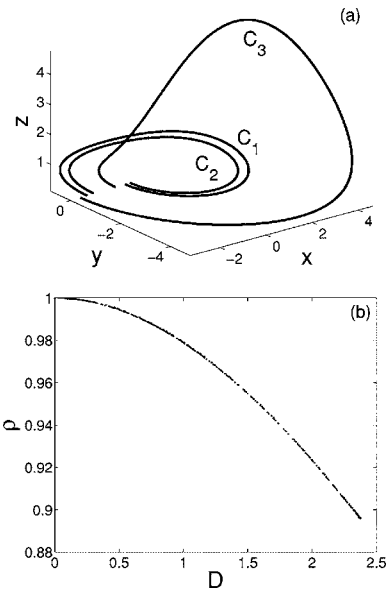

FIG. 1. (a) Three cycles C_1, C_2, C_3 in the phase space for the Rössler system. $\rho_{12}=0.9961$ and $\rho_{13}=0.9192$. Note that $\rho_{12}>\rho_{13}$, therefore C_2 is closer to C_1 than C_3. (b) The relation between ρ_{ij} and D_{ij}.

$$x' = -(y+z), \quad y' = x + 0.398y, \quad z' = 2 + z(x-4). \quad (2)$$

In Fig. 1(b), we further examine the relation between ρ_{ij} and D_{ij}. D_{ij} is the distance between cycle C_i and C_j in phase and is defined as

$$D_{ij} = \min_{l=0,1,\ldots,l_j-l_i} \frac{1}{l_i} \sum_{k=1}^{l_i} \|X_k - Y_{k+l}\|,$$

where X_k and Y_k are the kth point of C_i and C_j in the reconstructed phase space, respectively. We find that ρ_{ij} decreases smoothly and monotonously with D_{ij}. Therefore, when we check the relation between cycles, the correlation coefficient can be used equivalently as the phase-space distance, which spares the effort involved in the phase-space reconstruction.

Now we study the correlation coefficient between each pair of cycles in detail. For cycle $C_i, (i=1,2,\ldots,m)$, we first calculate its correlation coefficients with the remaining $(m-1)$ cycles C_j $(j \neq i)$. Then we sort these ρ_{ij}s in descending order, and the $(m-1)$ cycles are also rearranged correspondingly. We denote the sorted cycle sequence as a column vector $R_i=\{C_{S_{i1}}, C_{S_{i2}}, \ldots, C_{S_{i(m-1)}}\}'$, where S_{ij} is the index of the jth most similar cycle to C_i. Then, by picking out the pth$(1 \leqslant p \leqslant m-1)$ element from each column $R_1, R_2, \ldots, R_m$ and linking them together in order, we can build a (row) sequence of m cycles, denoted as $T_p=\{C_{S_{1p}}, C_{S_{2p}}, \ldots, C_{S_{mp}}\}$. For consistency, the original cycle series is denoted as $T_0=\{C_1, C_2, \ldots, C_m\}$.

Note that each cycle in T_p is the pth "closest" to the corresponding cycle in T_0, therefore as p gradually increases, T_p will grow less and less similar to T_0, i.e., T_1 is the most similar and nearest cycle series to T_0, while T_{m-1} is the most different and farthest one. We call these $m-1$ cycle series *pseudocycle series* as opposed to the "real" cycle series T_0. The construction of this hierarchy of pseudocycle series T_p $(p=1,2,\ldots,m-1)$, as we shall see later, provides a different way to examine the deterministic and chaotic structure in the original time series. It should be noted that the T_ps are not necessarily reorderings of T_0, since, for example, two cycles may have the same cycle as nearest neighbors which would result in a double entry in T_1.

Now we demonstrate how to extract useful information from the T_ps. For clarity of the notation, we use S_p to represent the sequence of the cycle indexes in T_p, i.e., $S_p=\{S_{1p}, S_{2p}, \ldots, S_{mp}\}$. Then we count the number of cycle pairs in T_p that satisfy the following condition;

$$S_p(i+k) - S_p(i) = k \quad (1 \leqslant i \leqslant m-k; k \geqslant 1), \quad (3)$$

where $S_p(j)$ represents the jth element in S_p. Physically, this means that we not only see the cycle C_i evolve into cycle C_{i+k} after k cycles in T_0 (which is trivial), but also see C_i's pth closest cycle $C_{S_{ip}}$ evolve into C_{i+k}'s pth closest cycle $C_{S_{(i+k)p}}$ after k cycles in T_p. Intuitively, this indicates that the two cycles C_i and $C_{S_{ip}}$ nearby in phase space are strongly correlated by sharing similar dynamical evolution, and the correlation time lasts for k cycles. We use N_{pk} to denote the number of cycle pairs in T_p that satisfy the condition (3). Note that for higher p and higher k the condition (3) can be fulfilled by chance. To avoid this, we use a more stricter condition, i.e., for each cycle span k only the pairs of cycles should be counted that fulfill condition (3) for all cycle span $\leqslant k$. Considering it is hard for this condition to be strictly met in the presence of noise, we suggest using this condition for clean data, and condition (3) for noisy data.

For chaotic systems, the distance between two nearby cycles will increase exponentially over time due to the very nature of sensitivity to initial conditions. Therefore, the correlation between two cycles, which is reflected in N_{pk}, is also expected to drop exponentially with the cycle span k. The semilog arithmic plot $\ln(N_{pk}) \sim k$ thus appears to be a straight line [see Fig. 2(a)], whose slope is actually related to the largest Lyapunov exponent. The larger the $|\Delta \ln N_{pk}/\Delta k|$, the higher the level of chaos. So we can use $|\Delta \ln N_{pk}/\Delta k|$ as an indicator of chaos, which we call *cycle divergence rate* (CDR). Usually we use $p=1$ for CDR, because T_1 is the most similar cycle series to T_0, and therefore maintains most of the determinism and chaos. When p increases, T_p will inherit less determinism and chaos from T_0, therefore $\ln(N_{pk}) \sim k$ will be subject to more statistical fluctuations, and $|\Delta \ln N_{pk}/\Delta k|$ [a linear fit to $\ln(N_{pk}) \sim k$] will also drop because the level of chaos decreases.

It is interesting to see that by summing the $N_{pk} \sim k$ curve for the first θ pseudocycle series, i.e., $N_k=\Sigma_{p=1}^{\theta} N_{pk}$ $(\theta=0.15m)$, we can find a power law relation between N_k and k, i.e., a linear dependency between $\ln(N_k)$ and $\ln(k)$. This is similar to the combination of a number of stochastic processes whose autocorrelation functions decay exponentially with different time constants, and the resultant process assumes a power law relation in its autocorrelation function. We call $|\Delta \ln N_k/\Delta \ln k|$ the *average cycle divergence rate* (ACDR), see Fig. 5(a) (the line with symbol "ring") for an illustration. And a larger ACDR indicates a higher level of

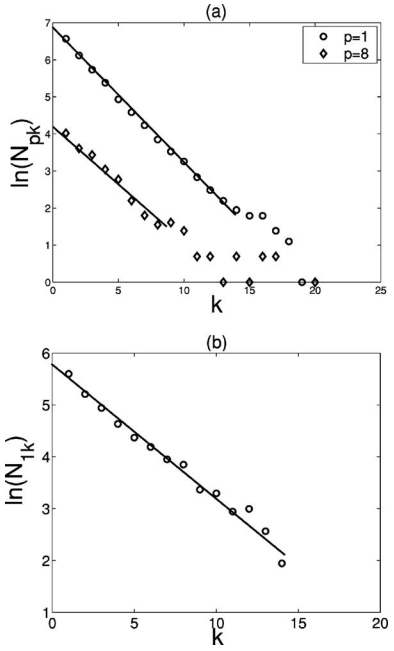

FIG. 2. Cycle divergence rate for (a) the x component of Rössler system (2) ($m=1596$). (b) Laser data set ($m=1224$).

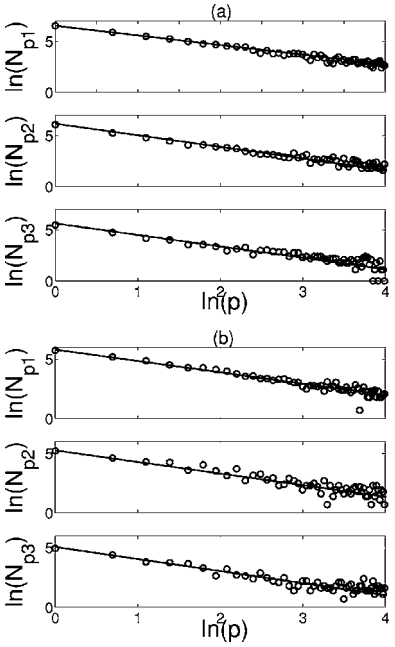

FIG. 3. Spatial decorrelation rate for (a) the Rössler system. (b) Laser data set.

chaos. Experimentally, ACDR is more robust to noise.

The third statistic we will address is the $|\Delta \ln N_{pk}/\Delta D_p|$ (k being fixed), called the *spatial decorrelation rate*, where D_p is the distance between T_p and T_0 in phase space. This statistic characterizes the decrease of determinism in T_ps. Since D_p is usually unavailable without embedding, we try to find the relation between N_{pk} and p instead. As is mentioned, T_p inherits less determinism from T_0 when p increases, hence N_{pk} is expected to drop exponentially with D_p, i.e., $N_{pk} \propto \exp(\gamma_1 D_p)$. Experimentally, we also find $(1-\rho_p) \propto \exp(\gamma_2 D_p)$, where ρ_p is the averaged correlation coefficients between corresponding cycles in T_0 and T_p. Combining these two equations, we have $N_{pk} \propto (1-\rho)^{\gamma_1/\gamma_2}$. Since $(1-\rho)$ is found to increase linearly with p (for $p \leqslant 0.15m$), we finally have $N_{pk} \propto p^{\gamma_1/\gamma_2}$, i.e., we expect N_{pk} to assume a power law relation with p. This is verified in Fig. 3.

To give a more vivid representation of the correlation between the cycles, we now follow the spirit of the recurrence plot (RP) [12] to provide a graphical illustration. The idea is that for each pair of cycles C_i and C_j, if ρ_{ij} is beyond a certain threshold, we then draw a black dot at position (i,j) in the 2D plane. We call this the *cycle recurrence plot* (CRP). Note that all cycles have already been sorted in the T_ps, the CRP thus can be easily obtained by drawing black points at $[i, S_p(i)]$, where $i=1,2,\ldots,m, p=1,2,\ldots,\theta$, and $\theta=0.15m$, see Fig. 4. Here we use θ because the threshold of the correlation coefficient may differ for distinct time series.

The observation of the points consecutive in time (forming the line parallel to the diagonal) in RP is an important signature of determinism. Similarly, in CRP the parallel lines to the diagonal that connects consecutive cycles have the same significance. Unlike RP, in CRP there is almost no horizontal or vertical lines. The reason is that the correlation coefficient ρ measures the distance between segmented cycles, rather than the phase-space points. Therefore the correlation reflected in CRP is on a much larger time scale, and the trivial temporal correlation of the phase-space points, typical in RP, is largely eliminated. This makes CRP much more robust to noise. Moreover, the cycle recurrence plot does not need phase-space reconstruction.

Based on the graphical recurrence plot, various statistics can be further calculated to give a more quantitative description of the deterministic structure, such as those designed in recurrence quantification analysis, including the percent recurrence, the percent determinism, the Shannon entropy, the trend, and the maximum diagonal line length [13]. These statistics can also be applied in our cycle recurrence plot. Among them, the maximum diagonal line length L_{max} is related to the largest Lyapunov exponent. However, the diagonal line in RP is easily broken into shorter fractions due to the influence of noise and the round-off error. Therefore $1/L_{max}$ may be imprecise in many cases. In comparison, what the cycle divergence rate reflects is the decreasing rate of the number of strongly correlated cycles N_{pk} over the

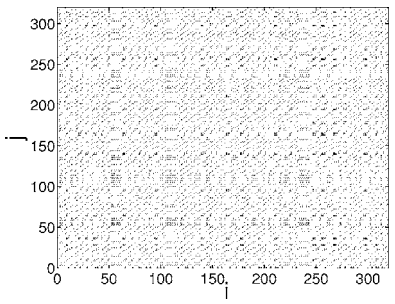

FIG. 4. Cycle recurrence plot for x component of the Rössler system with 5% measurement noise (i.e.; noise with standard deviation that is 5% of that of the data).

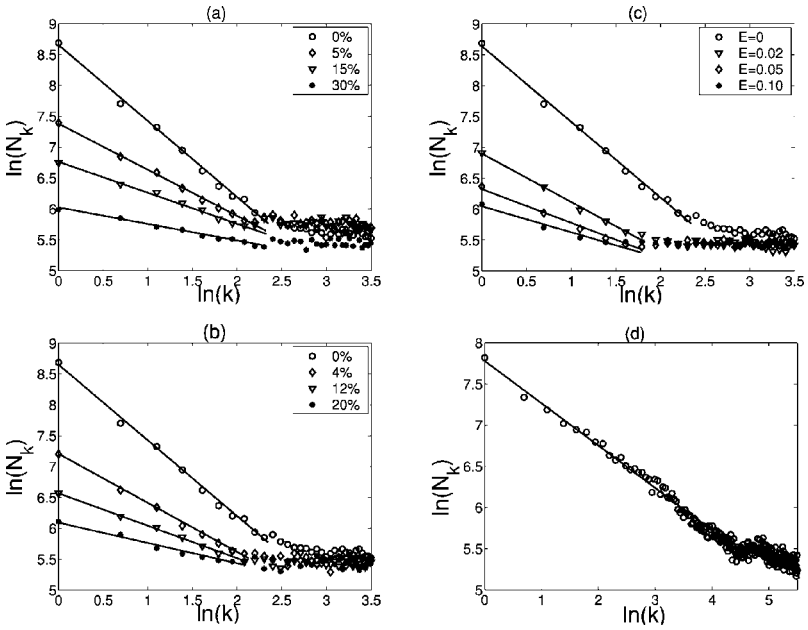

FIG. 5. ACDR for the Rössler system with (a) additive Gaussian noise of different levels. (b) Additive colored noise from the AR process $r_{n+1}=ar_n+b\eta_n$ (η_n is the Gaussian noise term) when $a=0.1, b=0.3$ (noise level 4%), $a=0.5, b=0.8$ (noise level 12%), and $a=0.8, b=0.9$ (noise level 20%). (c) Dynamical noise of different levels. Noise term $E\eta(t)$ is added to the right-hand side of the first equation in (2), which is integrated at a time span of 0.2. (d) ACDR for the laser data set.

cycle span k, rather than the absolute number of cycles in the diagonal line in CRP. Clearly, the cycle divergence rate is a more reliable statistics.

At last we consider the influence of different types of noise on the statistics we have defined. In the case of measurement noise, all the pair-wise correlation coefficient ρ_{ij}s will decrease. However, since the measurement noise has no preference in influencing different cycles in the time series, ρ_{ij}s will decrease roughly to the same extent, and their relative order remains nearly unchanged. Experimentally, when the noise level is low (less than 5%), the CDR from T_1 works well for chaos detection. For medium and high level noise (less than 30%), however, ACDR is more robust. In Fig. 5, we can see that ACDR can successfully detect chaos in the presence of additive noise (including white and colored noise) and dynamical noise. Though ACDR will drop as the noise level increases, it can bear a noise level (white Gaussian noise) of up to 30%. Moreover, we have also tested the above three statistics on the experimental laser data set. The results [see Figs. 2(b), 3(b), and 5(d)] clearly indicate the presence of chaos. We note that the minimum m needed to produce reasonable results for ACDR is about 100 for data with noise up to 5%. Larger m is needed for more noisy data. In this paper, larger m ($m=1596$ for the Rössler system and $m=1224$ for laser dataset) is used to produce a wider scaling region for visual inspection.

Through this approach, we can also discriminate between a low-dimensional chaotic signal and a periodic signal with noise. For the chaotic time series, as we have seen, N_{1k} will decrease exponentially with k, and a scaling region is present in the plot of $\ln(N_k) \sim \ln(k)$. While for periodic signals with dynamic noise, there are no such relations. For example, in the case of additive white noise, the correlation coefficients between different pairs of cycles are totally random, therefore the N_{pk} has very small values, and N_k will assume statistically the same value for different k. For the colored noise, though cycles of the noisy periodic signal might appear "correlated" due to the intrinsic correlation of the colored noise, we cannot find the scaling region in the plot of $\ln(N_k) \sim \ln(k)$, since N_k are roughly the same for k shorter than the decorrelation time of the noise. As a comparison to the embedding-based methods, we have calculated the correlation dimension (D_c) for a periodic time series $\sin(10\pi t)$ contaminated by colored noise $r_{n+1}=0.8r_n+0.9\eta_n$. The result $D_c=1.38$ shows that the traditional method cannot reliably differentiate a noisy periodic time series from a chaotic one. In contrast, our method works for this case, as discussed above.

In summary, we have proposed a method to detect deterministic structure and chaos for time series data exhibiting strong pseudoperiodic behavior. The intrinsic correlation of the data set is studied on the scale of single cycles by using a similarity measure, thus phase-space reconstruction can be avoided. The primary uniqueness lies in the hierarchy of the pseudocycle series that in turn shares decreasing similarity with the original signal. From these, reliable statistics can be defined for robust detection of determinism and chaos. For many real world systems that may contain both deterministic component and random noise of different sources, on which traditional invariants may fail, the proposed approach is a more robust and reliable alternative.

We wish to thank Kai Zhang and Tomomichi Nakamura for their valuable suggestions and help. This research was funded by a Hong Kong University Grants Council Grant Competitive Earmarked Research Grant (CERG) No. PolyU 5353/03E.

[1] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).
[2] L. W. Salvino and R. Cawley, Phys. Rev. Lett. **73**, 1091 (1994).
[3] Guillermo J. Ortega and Enrique Louis, Phys. Rev. Lett. **81**, 4345 (1998).
[4] R. Wayland *et al.*, Phys. Rev. Lett. **70**, 580 (1993).
[5] J. Theiler *et al.*, Physica D **58**, 77 (1992).
[6] M. Small *et al.*, Phys. Rev. Lett. **87**, 188101 (2001).
[7] X. Luo *et al.*, Phys. Rev. E **71**, 026230 (2005).
[8] B. Cazelles and R. H. Ferriere, Nature (London) **355**, 25 (1992).
[9] A. Osborne and A. Provenzale, Physica D **35**, 357 (1989).
[10] J. Theiler, Phys. Lett. A **155**, 480 (1991).
[11] H. Abarbanel *et al.*, Rev. Mod. Phys. **65**, 1331 (1993).
[12] J.-P. Eckmann *et al.*, Europhys. Lett. **4**, 973 (1987).
[13] L. Trulla *et al.*, Phys. Lett. A **223**, 255 (1996).