Kolmogorov-Smirnov test distinguishes attractors with similar dimensions

A. M. Albano
Department of Physics, Bryn Mawr College, Bryn Mawr, Pennsylvania 19010

P. E. Rapp
Department of Physiology, The Medical College of Pennsylvania, Philadelphia, Pennsylvania 19129

A. Passamante
Naval Air Warfare Center Aircraft Division, Warminster, Pennsylvania 18974
(Received 30 September 1994; revised manuscript received 3 April 1995)

Recent advances in nonlinear dynamics have led to more informative characterizations of complex signals making it possible to probe correlations in data to which traditional linear statistical and spectral analyses were not sensitive. Many of these new tools require detailed knowledge of small scale structures of the attractor; knowledge that can be acquired only from relatively large amounts of precise data that are not contaminated by noise—not the kind of data one usually obtains from experiments. There is a need for tools that can take advantage of "coarse-grained" information, but which nevertheless remain sensitive to higher-order correlations in the data. We propose that the correlation integral, now much used as an intermediate step in the calculation of dimensions and entropies, can be used as such a tool and that the Kolmogorov-Smirnov test is a convenient and reliable way of comparing correlation integrals quantitatively. This procedure makes it possible to distinguish between attractors with similar dimensions. For example, it can unambiguously distinguish ($p < 10^{-8}$) the Lorenz, Rössler, and Mackey-Glass (delay=17) attractors whose correlation dimensions are within 1% of each other. We also show that the Kolmogorov-Smirnov test is a convenient way of comparing a data set with its surrogates.

PACS number(s): 05.45.+b

I. INTRODUCTION

Recent advances in nonlinear dynamics have led to more informative characterizations of complex signals [1–4]. The realization that even simple dynamical systems are capable of exhibiting complex unstable behavior with broadband spectra has made it clear that descriptions in terms merely of time series, conventional statistical measures, and spectra are no longer sufficient. Chaotic trajectories starting from neighboring initial points diverge exponentially; simple nonlinear maps have spectra that are as broadband as those of the outputs of well-designed random number generators or of sequences as random as the time intervals between decays of radioactive nuclei [5].

Nonlinear dynamical tools have made it possible to probe correlations to which traditional linear statistical and spectral analyses were not sensitive. Some of these new tools make it possible, at least in principle, to determine if a signal is deterministic, to identify and classify seemingly random signals, to compare complex experimental results with theoretical or model predictions, or to distinguish signals generated by distinct but similar dynamical systems.

These new tools typically rely on the possibility of using a series of measurements of a single variable to reconstruct or embed the system's multidimensional trajectory. The embedding, or reconstruction, is usually done using the method of time delays [6–8]. If the reconstructed trajectory evolves on an attractor, the system's dynamics may be described in several ways, including (1) the rate at which nearby trajectories diverge from each other (Lyapunov exponents) [9–12], (2) the rate at which information is generated by each new datum (entropy production) [13], (3) the topological properties of the trajectory [14–18], or (4) by characterizing the geometry of the attractor by one or all of a hierarchy of dimensions [19–21].

Many of these new tools, however, require detailed knowledge of small scale structures of the attractor; knowledge that can be acquired only from relatively large amounts of precise data which are not contaminated by noise. Unfortunately, this is usually not the kind of data one gets from experiments.

When dealing with limited amounts of noise-contaminated data, there is a need for tools that can take advantage of "coarse-grained" information, but which nevertheless remain sensitive to higher-order correlations in the data. In many situations, the need is not for an absolute characterization of an attractor, but rather for a reliable way of comparing attractors. There is increasing realization that answering the question "is it chaos or is it noise?" is not as easy, or indeed as immediately important, as it once seemed. For instance, one often wishes to compare experimental results with the predictions of a theory or a model, or to compare an attractor with those of surrogate data—correlated noises with some statistical or spectral properties that mimic those of the original

1063-651X/95/52(1)/196(11)/$06.00

data [22,23]. In medical applications, one may wish to determine whether a particular signal more closely resembles those from a normal subject, or those from subjects afflicted by one of a variety of pathological conditions.

In Ref. [24] we illustrated the use of the (Gibbs) entropy and related measures to analyze sequences of time intervals between spikes in the electrical activity of human muscles (human electromyograms, or EMG's). We achieved statistically significant discrimination ($p < 0.001$) between normal and denervated muscle using measures related to the inertia tensor of two-dimensional embeddings of interspike interval data using less than 200 points.

In this contribution, we propose that the correlation integral, now much used as an intermediate step in the calculation of dimensions and entropies [18–21], can be used for comparing attractors, and that the Kolmogorov-Smirnov test (see, e.g., [25,26]) is a convenient and reasonably reliable way of quantifying the comparison. We emphasize that this procedure does not provide a quantitative characterization of individual attractors. Rather, it makes it possible to compare attractors even in some situations where the more commonly used nonlinear dynamical measures are not computable.

In Sec. II we review the use of the correlation integral in the calculation of dimensions and Kolmogorov entropies. We use the correlation dimension to illustrate problems that arise when dealing with limited amounts of noisy data, or when comparing dynamical systems with similar dimensions, or when dealing with moderately high-dimensional systems. In Sec. III we review the Kolmogorov-Smirnov test and discuss precautions that need to be taken when applying it to compare correlation integrals. In Sec. IV we illustrate some of its capabilities as well as its limitations by using it in the same situations that caused the problems discussed in Sec. II. We also illustrate its use in comparing a time series with its surrogates. Section V summarizes the paper and discusses some areas in which the procedures proposed here may profitably be used.

II. THE CORRELATION INTEGRAL

Let $X=\{x_1,x_2,\ldots,x_N\}$, $x_k=x(k\tau)$, be a series of measurements (a time series) of x, one of the variables that characterize a system. These measurements are taken at equal time intervals τ. The state of the system at any time is specified by the simultaneous values of all the system variables, and its temporal evolution is described by the trajectory of this representative point in phase space, the space which the variables span. The time series X may be embedded in an m-dimensional space by constructing the vectors

$$y_k^{(m)}=(x_k,x_{k+1},\ldots,x_{k+m-1}) , \quad k=1,2,\ldots,N_v , \tag{1}$$

where $N_v=N-m+1$ is the total number of vectors formed from the time series. For a sufficiently large value of the embedding dimension m, and if some additional conditions are satisfied, the reconstructed trajectory has the same topological and geometrical properties as the system's phase space trajectory [6–8,27,28]. Most nonlinear analysis of experimental data is done on trajectories reconstructed in this manner. In the development presented here, we consider the case of a single-channel time series. The procedure generalizes immediately to multivariate data.

In an m-dimensional space, the correlation integral $C_m(r)$ is the fraction of all distances between distinct pairs of points on the trajectory that do not exceed the value r [18,19]:

$$C_m(r)=\frac{2}{N_v(N_v-1)}\sum_{j=i+B+1}^{N_v}\sum_{i=1}^{N_v-B-1}\Theta(r-|y_i^{(m)}-y_j^{(m)}|) . \tag{2}$$

Here, $\Theta(\ )$ is the Heaviside function which has the value 1 if its argument is non-negative and is zero otherwise, N_v is the number of embedding vectors used, and $|y_i^{(m)}-y_j^{(m)}|$ is the distance between $y_i^{(m)}$ and $y_j^{(m)}$. The sum is taken only for those i's and j's that are separated in time by more than B sampling times to avoid artifactual correlations among consecutively sampled points on the attractor [29]. (Hereafter, we refer to B as the "blind".) It is seen that $C_m(r)$ is monotone increasing with r and, under this normalization, has a maximum value of 1. The correlation integral is what statisticians call the cumulative distribution function or the sample distribution function (see, e.g., Refs. [25,26]) of the interpoint distances. Grassberger and Procaccia have shown that the correlation integral satisfies the scaling relationship

$$\lim_{r\to 0}\lim_{m\to\infty} C_m(r)\propto r^{d_2}\exp(K_2 m\tau) , \tag{3}$$

where d_2 is the correlation dimension and K_2 the order-2 Kolmogorov entropy [18,19].

The correlation dimension is usually evaluated by taking the slope of the double-logarithmic plot of $C_m(r)$ vs r in that interval of r values where the right-hand side of Eq. (3) is satisfied, the so-called scaling region. Instead of taking the high-m limit, Takens [7] shows that taking $m > 2d_2$ may be sufficient. Sauer, Yorke, and Cadagli [8] show that $m > d_2$ is good enough.

K_2 is calculated by taking differences of log-log plots of $C_m(r)$ vs r for different values of m. Values of d_2 and K_2 obtained in this manner have been used in numerous studies to identify, classify, or diagnose the behavior of many complex systems (see, e.g., Refs. [1–4]).

More recently, they have also been used to compare the properties of test data with those of their surrogates, sequences of random numbers designed to have some statistical and/or spectral properties that are identical to those of the original data [22,23]. Generalizations of Eq. (2) to order-q correlation integrals, where q in principle ranges from $-\infty$ to $+\infty$, have been used to estimate order-q dimensions and entropies [20,21], which provide a more comprehensive description of the reconstructed trajectory or of the attractor on which it moves.

A. Computational details

The embedding window T_w, the time interval spanned by each embedding vector, is given by

$$T_w=(m-1)L\tau\ , \tag{4}$$

where, as above, m is the embedding dimension, τ is the time interval between successive elements of the time series, and L, the lag, is the number of elements of the time series between successive components of an embedding vector. It is known [30,36] that, for dimension calculations using sufficiently high embedding dimensions, it is the embedding window, and not the embedding dimension m and the lag L separately, that is important. Although there is no unanimity on how to choose an optimal embedding window (see, e.g., Refs. [30–35]), using one that is based on the autocorrelation function or higher-order correlation functions of the data has led to consistently reasonable results. In the following, we use the time to first minimum of the autocorrelation function as a window, primarily because it is easily computable. We used the same quantity as the "blind," B.

Although it is the embedding window that is important in dimension calculations, there is no reason to expect that this remains true for the calculation of other quantities. Indeed, Kaplan and Glass [37,38] have shown that for calculating a measure of deterministic structure which they introduced it is the lag rather than the embedding window that is important. To obtain a lag that is determined by the data, we make use of a criterion proposed by Schuster [32,33], one that has lately also been used by a number of other investigators (e.g., [31,35]). Schuster's criterion is based on the observation that if a set of dimension d is successively embedded in spaces of dimensions m and $m+1$ then, if $d>m$, nearest neighbors in m dimensions may not remain nearest neighbors in $m+1$ dimensions, and nearest-neighbor distances may increase as one goes from the smaller embedding space to the larger. If a sphere is projected onto its equatorial plane, the north and south poles coincide in the projection, becoming "false nearest neighbors." Increasing the embedding dimension increases the distance between false near neighbors. Using too small a lag has a similar effect. If the lag is too small, all points cluster around the main diagonal of the embedding space, and nearest-neighbor distances are artificially small. As the

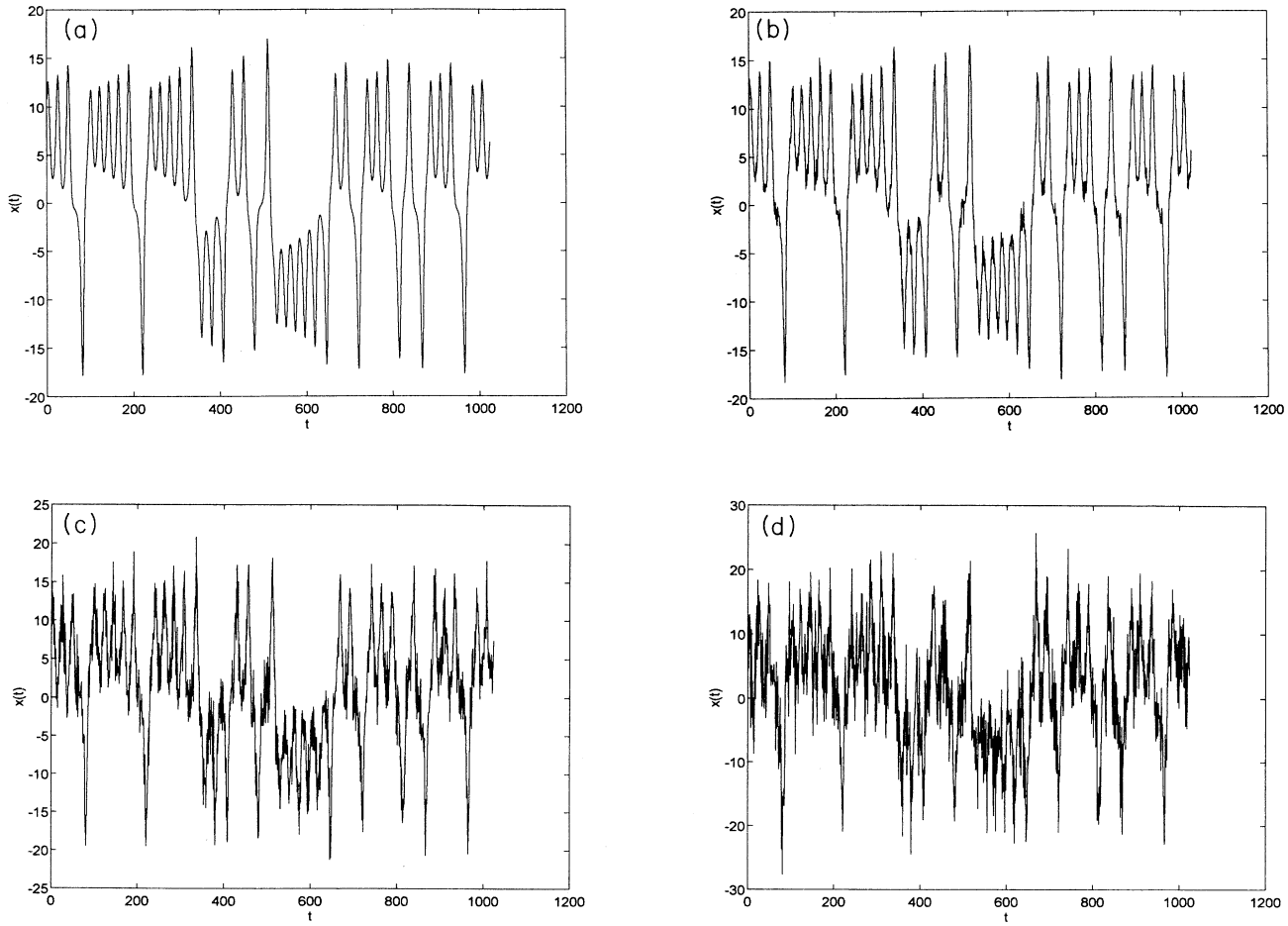

FIG. 1. Time series for the x variable of the Lorenz equations, Eq. (5). (a) No added noise. (b) The signal in (a) after the addition of white Gaussian noise to give a signal-to-noise ratio (SNR) of 20 dB, (c) SNR = 10 dB, and (d) SNR = 5 dB.

lag is increased, interpoint distances increase. To quantify these notions, we have used a variant of a false nearest-neighbor procedure introduced by Schuster and his colleagues [32,33] to help determine appropriate values of L and m.

B. The correlation dimension

The limiting scaling behavior given in Eq. (3) and similar behavior of higher-order correlation integrals underlies the calculation of generalized dimensions and entropies. The necessity of taking limits as r becomes small and m large imposes severe limitations on the reliability of the calculated quantities. It is small values of r that are most sensitive to noise and experimental uncertainties. And as the embedding dimension m is increased, nearest-neighbor distances on the embedded attractor increase, raising even more the lowest level to which the small r limit can be taken.

Limitations on the reliability with which dimensions and entropies can be calculated from experimental data makes it inappropriate to use them when precise values of these quantities are required. This is especially so when the intent is to get an absolute characterization of the attractor being studied. In the rest of this section we illustrate the nature of some of these problems by giving computational examples of what happens (1) when a signal becomes increasingly noise corrupted, (2) when trying to compare time series from systems with similar dimensions, and (3) when using insufficient amounts of data to study signals from moderately high-dimensional systems.

1. Additive noise

Figure 1(a) shows a segment of a series of x values calculated from the Lorentz equations [39], integrated in steps of 0.01,

$$(dx/dt, dy/dt, dz/dt) = (-10(x-y), x(28-z)-y, xy-\tfrac{8}{3}z) . \quad (5)$$

Figures 1(b)–1(d) show the same time series after Gaussian white noise has been added to give signal-to-noise ratios (SNR's) of 20, 10, and 5 dB, respectively (the SNR in decibels (dB) $=10\log_{10}$[(signal variance)/(noise variance)]). Figure 2(a) shows plots of $\log_e C_m(r)$ vs $\log_e(r/r_{\max})$ for all four cases, $r_{\max}$ being the largest interpoint distance on each attractor, while Fig. 2(b) shows plots of the slopes vs $\log_e(r/r_{\max})$ for each of the graphs in Fig. 2(a). In each calculation, 2048 points were used and the lag and embedding window were chosen as described in the previous section.

The least steep graph in Fig. 2(a), which corresponds to the graph with the long, flat plateau in Fig. 2(b), was calculated from the uncorrupted or "clean" data. The increasingly steep graphs at larger values of $\log_e(r/r_{\max})$ were calculated from the 20, 10, and 5 dB data, respectively. The plateau value of the slope in Fig. 2(b), approximately 2.0 for the clean data, is the estimate of the correlation dimension obtained from this calculation. This value characterizes the attractor only in the scaling region, the region in which the graph shows a plateau. As noise is added, the plateau region shrinks, making it increasingly difficult to claim that a meaningful calculation of dimension has been achieved. It also becomes increasingly difficult to assert that the resulting graphs still characterize signals that came from the same underlying system.

Figure 2 provides a dramatic demonstration of how disastrously noise affects the calculation of the correlation dimension. Interestingly, most of the effects of noise on the Lorenz attractor, even at the SNR $=5$ dB level, is restricted to the smallest 5% of interpoint distances. In Fig. 2(a), this corresponds to $\log_e C_m(r) \leq -3.0$ or $\log_e(r/r_{\max}) \leq -1.6$, but since $\log_e(r/r_{\max})$ is used as the

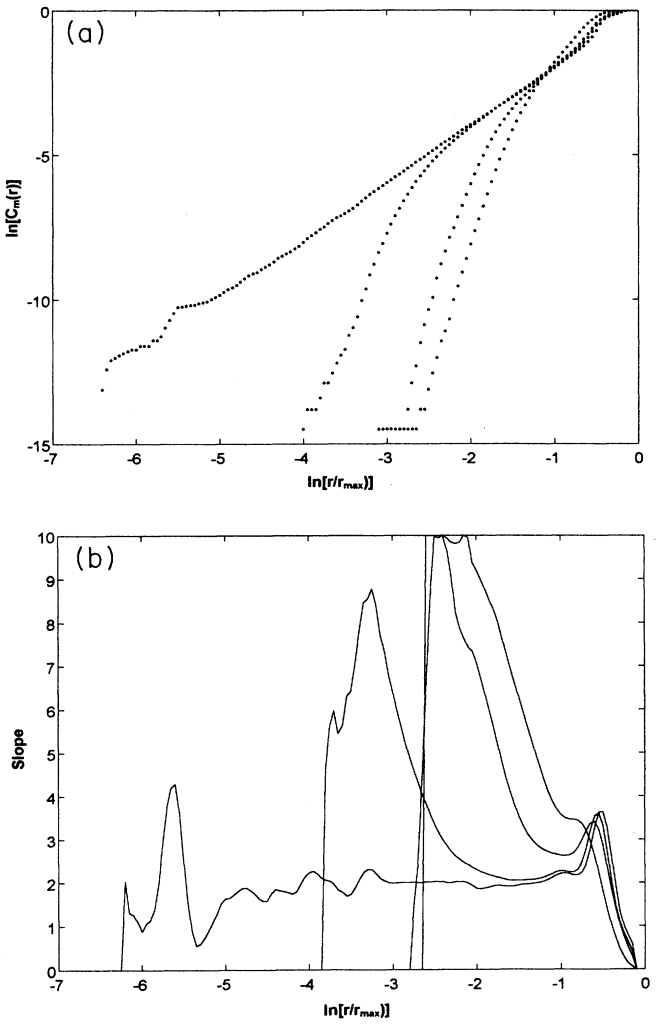

FIG. 2. (a) $\ln[C_m(r)]$ vs $\ln(r/r_{\max})$ for the signals shown in Fig. 1. Distances are scaled to $r_{\max}$, the largest interpoint distance on the attractor; $m=11$, $L=4$ for all cases. The least steep graph is that of the noise-free data. The increasingly steep graphs at larger values of $\ln(r/r_{\max})$ are of the data sets with SNR's of 20, 10, and 5 dB, respectively. (b) The slopes of the graphs in (a) obtained by making least squares fits of successive nine-point segments of the graphs. 95% of all interpoint distances are in the region $\ln(r/r_{\max}) \geq -1.6$.

ordinate in Fig. 2 the effect of this small fraction of distances is grossly exaggerated. Noise corrupts at small length scales. However, these are the length scales that are essential to an estimate of dimension which, as indicated in Eq. (3), is obtained in the $r \to 0$ limit.

2. Similar dimensions

There are similar limitations on the usefulness of the correlation dimension to distinguish systems with nearly equal dimensions, even when using relatively clean data. Such is the case, for instance, with the attractors of the Lorenz equations, Eq. (6), the Rössler equation [40], integrated in steps of 0.05,

$$(dx/dt, dy/dt, dz/dt) = (-y-z, x+0.2y, 0.4+xz-5.7z) , \quad (6)$$

and the Mackey-Glass equation [41]

$$dx(t)/dt = \frac{0.2x(t-\delta)}{1+[x(t-\delta)]^{10}} - 0.1x(t) , \quad (7)$$

integrated in steps of 0.5, with delay $\delta = 17$. All three have correlation dimensions that are approximately 2.0 [29,42]. That is, the differences in their correlation dimensions are in the second decimal place so that these must be calculated with an accuracy of the order of 1% if they are to be useful in distinguishing any one attractor from the others. Accuracies of this quality when using the Grassberger-Procaccia algorithm impose data requirements that are effectively unobtainable with experimental data.

Figure 3 shows the slope of the $\log_e C_m(r)$ vs $\log_e(r/r_{\max})$ plots for the three cases. These graphs show, first of all, the difficulty of establishing the location of a scaling region and, once this has been done, of obtaining an estimate of the average slope for the region with the desired accuracy and precision. It is clear that merely quoting estimated values of the correlation dimensions with their corresponding uncertainties will not suffice to distinguish these attractors.

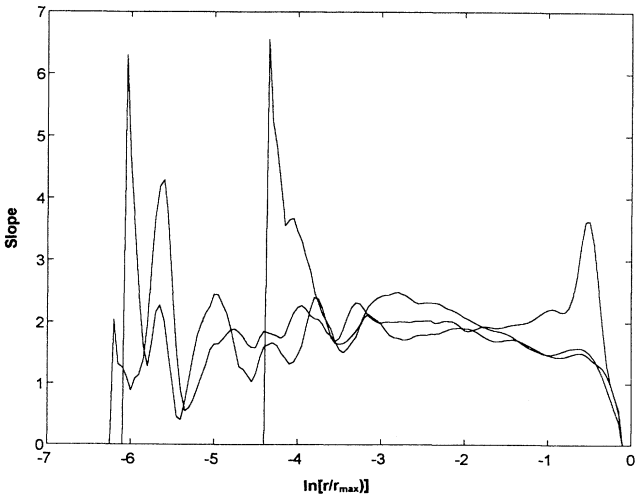

FIG. 3. Slopes of the $\ln[C_m(r)]$ vs $\ln(r/r_{\max})$ graphs of the Lorenz [Eq. (5); $m=11$, $L=4$], Rössler [Eq. (6); $m=11$, $L=5$], and Mackey-Glass (delay $\delta=17$; $m=11$, $L=5$) [Eq. (7)] systems. All three have correlation dimensions of approximately 2.0.

3. High dimensions

There are some rather stringent limits on the number of data points needed to determine dimensions reliably. A limit due to Eckmann and Ruelle [43] requires $(1/\rho)^{d_2/2}$ data points to resolve an attractor of correlation dimension d_2, where ρ is the ratio of the length scale being characterized to the maximum length on the attractor. For $\rho=0.1$, this gives a rather generous $10^{d_2/2}$. Nerenberg and Essex [44] propose a more stringent $2^{d_2}(d_2+1)^{d_2}$, while Smith [45] proposes an even more stringent 42^{d_2} (however, see Ref. [46]). A commonly used rule of thumb that has not been given much rigorous backing is 10^{d_2}. It may be possible to estimate the value of the correlation dimension using fewer data than are required by the more stringent of these requirements once it is established that the correlation integral satisfies the scaling property given in Eq. (3). However, it may take considerably more to establish that this is so. If this cannot be done, that is, if it cannot be shown that the correlation integral has a non-negligible scaling region, then it cannot be claimed that the data set is characterizable by a correlation dimension. Figures 4(a)–4(d) illustrate this. Figure 4(a) shows the $\log_e C_m(r)$ vs $\log_e(r/r_{\max})$ plot for a time series generated by the Mackey-Glass equation, Eq. (7), with a delay, $\delta=150$, which has a high dimensional attractor. Figure 4(b) shows the slope of this graph vs $\log_e(r/r_{\max})$. Figures 4(c) and 4(d) show the corresponding plots for filtered random noise (cf. Ref. [23]). In both cases, 4096 data points were used. This is half of the number of points used in Ref. [23] to exaggerate the problems that arise when an insufficient number of data are used. None of the graphs shows a convincing scaling region. Assigning a correlation dimension to either system on the basis of Fig. 4(a)–4(d) is not entirely defensible. Nevertheless, there are discernible differences in the graphs for the two systems.

C. Surrogate data

The difficulty of using dimensions and similar measures to distinguish between deterministic and stochastic or noise-dominated signals have led a number of investigators to propose the use of surrogate data [22,23]. These are randomized sequences some of whose statistical and/or spectral properties are similar to those of the time series being studied. Measures calculated using the original data set are compared to those obtained using its surrogates. If the measures so obtained are not significantly different, then one concludes that the original data are not distinguishable from correlated noise, which is what the surrogate data are.

Following Theiler *et al.* [22] we consider two types of surrogates. The first, referred to as algorithm I, or type I, or *phase-randomized* [5] surrogates, is linearly correlated

noise with the same power spectrum (and hence the same autocorrelation function) as the original data set. It is obtained by (a) calculating the Fourier transform of the original data, (b) randomizing the phases of this transform, and then (c) taking the inverse of the phase-randomized transform.

Algorithm II, or type II, or *Gaussian-scaled* [5] surrogates address the null hypothesis that the original set is linearly correlated noise that has been transformed by a static, monotone nonlinearity. It has the same distribution, and therefore the same statistical moments to all orders, as the original data. A comparison of the original data set and its Gaussian-scaled surrogates using the usual statistical moments such as mean, variance, kurtosis, etc., would therefore find no differences.

A Gaussian-scaled surrogate is obtained as follows: (a) A Gaussian-distributed set of random numbers that has the same rank structure as the original data is constructed; (b) a random-phase surrogate of this Gaussian distributed set is constructed; (c) the original data set is shuffled so that it has the same rank structure as the random-phase surrogate constructed in step 2. Since Gaussian-scaled surrogates are shuffles of the original data, they have exactly the same statistical measures as the original, provided these measures do not depend on the way the data are ordered in time.

The quantities used for comparing a data set to it surrogates may be any one of the nonlinear dynamical measures used to analyze time series. Theiler *et al.* [22] suggest creating several realizations of each type of surrogate. A measure, the correlation dimension, say, for each of these surrogate sets is obtained and then one calculates the quantity

$$S=\frac{|d_{\text{data}}-\langle d_{\text{surr}}\rangle|}{\sigma}\ , \tag{8}$$

where d_{data} pertains to the original data, and $\langle d_{\text{surr}}\rangle$ and σ are the mean and standard deviation, respectively, for the set of surrogates. The value of S quantifies the difference between the raw data and its surrogates. Several alternative statistical procedures for evaluating results obtained with surrogates are compared in Ref. [5].

This procedure is effective when the measure used for

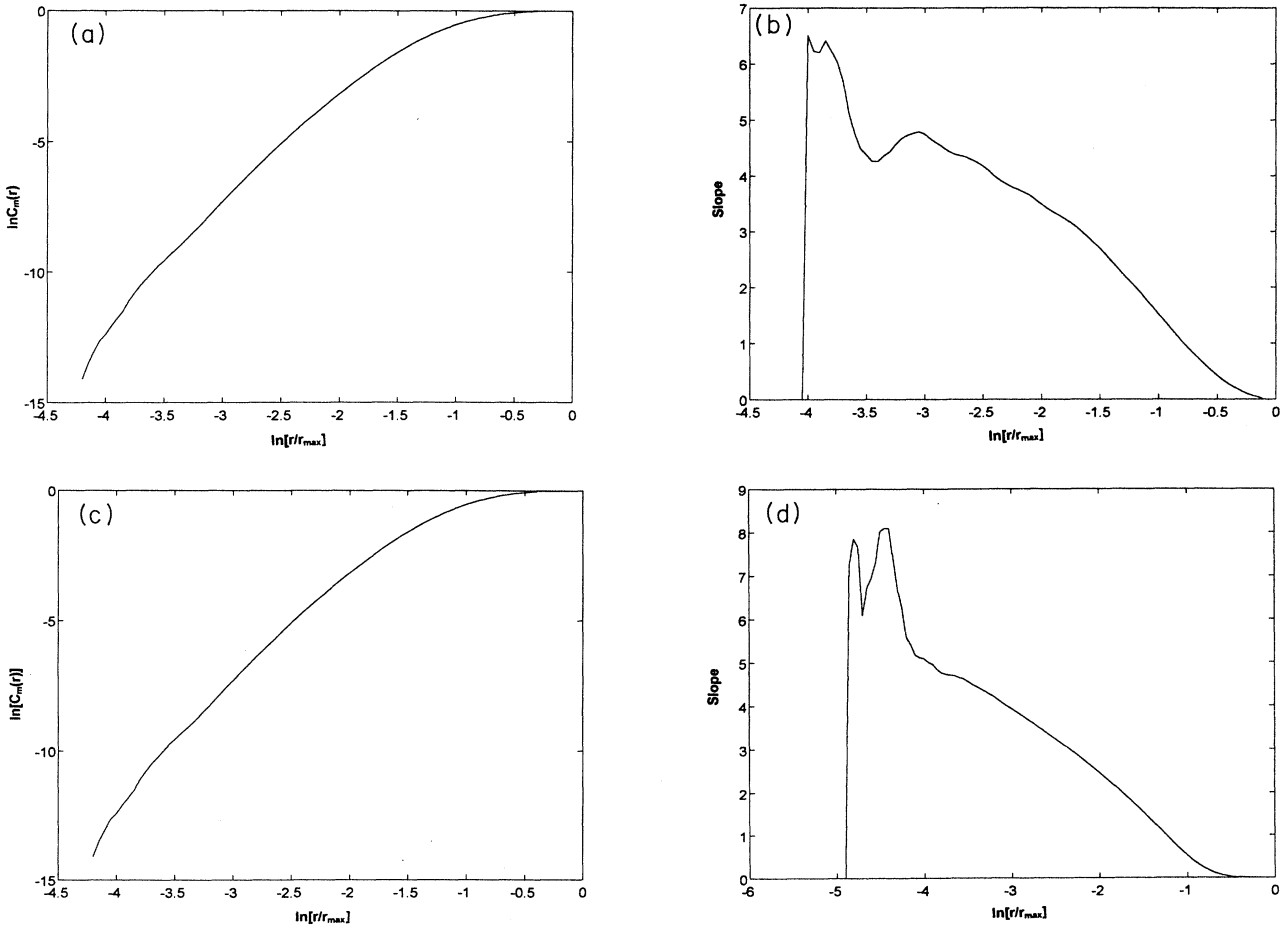

FIG. 4. (a) $\ln[C_m(r)]$ vs $\ln(r/r_{\max})$ for a 4096-point output of the Mackey-Glass equation [Eq. (7)] with delay $\delta=150$; $m=21$, $L=6$. (b) Slope of (a). The corresponding graphs for filtered random noise with $m=11$, $L=1$ are shown in (c) and (d). None of the graphs shows a scaling region.

the comparison can be calculated reliably. As seen in the previous sections, however, there are situations where the conditions for the proper calculation of these measures cannot be met.

III. COMPARING CORRELATION INTEGRALS

The examples given in the previous section illustrate the difficulty of obtaining precise values of the correlation dimension to describe limited amounts of noise-corrupted data. Calculations of order-q dimensions and entropies [20] are beset by similar problems. These problems arise, in part, from the need to establish a scaling region of reasonable length, and the fact that calculating dimensions involves estimating slopes numerically, a not very robust numerical procedure. Calculating Kolmogorov entropies is similarly afflicted as it involves taking the differences of equally uncertain quantities. The correlation integral, on the other hand, only involves counting, once interpoint distances have been calculated. Being a finite sum, it is a relatively robust object. It is less sensitive to perturbations than dimensions and entropies, a property that becomes a virtue rather than a limitation when the object is to identify, or classify, or diagnose, a time series of limited length that may have been perturbed by noise and other experimental uncertainties.

The examples above also show that, even when the conditions for a proper calculation of dimensions or entropies are not met, making it impossible to claim a dimension, the correlation integrals of similar attractors still have some similarities, and those of different attractors have some differences. These can be exploited to estimate how similar, or how different, the systems are. We propose to exploit these by using the Kolmogorov-Smirnov test to compare correlation integrals.

A. The Kolmogorov-Smirnov test [25,26]

The following is a brief account of the Kolmogorov-Smirnov test for the case when two sets of measurements are being compared. For a more comprehensive treatment, see, e.g., Ref. [25]; for computational details, see Ref. [26].

Let $\{\eta_1,\eta_2,\ldots,\eta_{N1}\}$ and $\{\xi_1,\xi_2,\ldots,\xi_{N2}\}$ be two sets of measurements (or calculations) of the same quantity. One wishes to test the null hypothesis that the two sets of measurements come from the same distribution. For this purpose, Kolmogorov makes use of the sample distribution functions, or cumulative distribution functions, $S_1(\eta)$ defined by

$$S_1(\eta)=\frac{1}{N_1}\sum_{i=1}^{N_1}\Theta(\eta-\eta_i) \tag{9}$$

and $S_2(\xi)$, which is similarly defined. $S_1(\eta)$ is the fraction of all the measurements that do not exceed η, and $S_2(\xi)$ is the similar quantity for ξ. As noted earlier, these are similar to the correlation integral, Eq. (2). Kolmogorov's statistic D is the largest difference between the two cumulative distribution functions for the same value of the argument,

$$D=\sup_{\eta}|S_1(\eta)-S_2(\eta)| . \tag{10}$$

If the null hypothesis is true, that the two sets of measurements come from the same distribution, then the probability for D to exceed the observed value is given by

$$Q_{\rm KS}(z)=2\sum_{r=1}^{\infty}(-1)^{r-1}\exp(-2r^2z^2) , \tag{11}$$

where

$$z=D\sqrt{N_1N_2/(N_1+N_2)} . \tag{12}$$

Algorithms for the calculation of this probability may be found in Ref. [26].

B. Caveats

The Kolmogorov-Smirnov test compares sequences without regard to the values of variables on which these sequences depend. When the sequences being compared are values of correlation integrals, the values of the interpoint distance r used for calculating the correlation integral become almost irrelevant. This makes it unnecessary to use the same normalizations or the same units of length in calculating the correlation integrals. It is, however, important (i) that each of the correlation integrals have a sufficiently large number of values [N_1 and N_2 in Eq. (10)] and (ii) that the two sequences have the same lower and upper bounds. But these values must correspond to actual distances on the attractor.

(i) The effect of the number of points used to specify each correlation integral is apparent in Eqs. (11) and (12). For $N_1\approx N_2\approx N$, the parameter z defined in Eq. (12) is of the order of $N^{1/2}D$, and is the argument of Gaussian functions that determine the probability [Eq. (11)]. It is clear that the sensitivity with which the test can distinguish two correlation integrals from each other is rather sensitive to the number of points used in the comparison.

(ii) The choice of that segment of the correlation integral used for the comparison is similarly crucial. Attractors with similar correlation dimensions have correlation integrals that behave similarly in their respective scaling regions (cf. Fig. 3). Comparing only values in the scaling region is thus unlikely to distinguish these attractors. The correlation integrals of similar attractors may have meaningful differences at large distances, but correlation integrals saturate—they approach unity at large distances. On the other hand, the smallest distances may be noise dominated and are therefore unlikely to contain much pertinent information.

In the following, we compare correlation integral sequences in the range $0.01\le C_m(r)\le 0.99$ [or $-4.6\le\log_e C_m(r)\le -0.01$], with r in steps of $\Delta r=0.001\sigma$, where σ is the standard deviation of the raw data. This includes almost all of the interpoint distances on the attractor, but avoids the smallest, most noise-corrupted ones, as well as the largest, where the correlation integrals begin to saturate. The choice of Δr, on the other hand, guarantees that this range of $C(r)$'s involves a few thousand entries in the $C(r)$'s being compared. Calculated probabilities can depend on the size of

Δr and the range of $C(r)$ values used. It is therefore important to specify these quantities whenever the Kolmogorov-Smirnov test is used to compare attractors. Limited computational experiments suggest, however, that for a small enough value of Δr the final results are robust against considerable variations in the range of $C(r)$ that is used. For example, in comparing the Lorenz and Rössler attractors (see Sec. IV below), essentially the same results are obtained as the range of $C(r)$ is varied for $\{0.01 \leq C(r) \leq 0.99\}$ to $\{0.3 \leq C(r) \leq 0.7\}$.

C. Lags and windows reconsidered

In the previous discussion, attention has been directed to the sensitivity of dynamical measures, particularly the correlation integral, to embedding parameters m and L. In our application of the Kolmogorov-Smirnov test, the sensitivity of $S(C_m(r))$ is of importance. This sensitivity is investigated in the calculations presented in this section. We present evidence, that, with lags of the order obtained with Schuster's criterion, the shape of $S(C_m(r))$ approaches a limit as the embedding dimension and the number of embedding vectors are increased.

Figure 5(a) examines the sensitivity of $S(C_m(r))$ to the number of points in the embedding space. The graphs represent two sets of calculations, one using 20 to 200 embedded vectors (in steps of 20), the other 200 to 4700 embedded vectors (in steps of 500). Both sets of calculations used a lag of 10 in 50 dimensions. The lowest graph corresponds to 20 vectors, the highest to 40 vectors. There are negligible differences in $S(C_m(r))$ if more than 160 vectors are used.

Figures 5(b) and 5(c) examine the effect on $S(C_m(r))$ of varying the embedding parameters. Figure 5(b) shows $S(C_m(r))$'s calculated using 1024 vectors in 50 dimensions with lags of $1,2,\ldots,20$. The two topmost graphs are for lags 1 and 2, the lowest are for lags 19 and 20, those with intermediate values of the lag cluster in between. In most of the calculations we present here, we used a lag of 10, which is compatible with that suggested by Schuster's criterion.

Figure 5(c) shows $S(C_m(r))$'s obtained using 1024 vectors with lags of 10 in embedding spaces with dimensions $5,10,\ldots,100$. The topmost graph is for a five-dimensional space. As the embedding dimension increases, the graphs approach the bottom of the band. Graphs for embedding dimensions 35–100 are practically coincident.

For a given number of embedding vectors, however, the graph of $S(C_m(r))$ becomes severely distorted at extremely large windows. This distortion is in the form of a steplike structure in $S(C_m(r))$ which occurs when the embedding dimension is so large that distances between neighboring embedding vectors exceed the scale Δr used in calculating the correlation integral.

The limiting form of $S(C_m(r))$ discussed above is reached at values of the embedding window that are too large to be used for the calculation of correlation dimensions. With windows of this size, the correlation integral shows no scaling region, making it impossible to conclude, using calculations with these values of the window, that the data can be characterized by a correlation di-

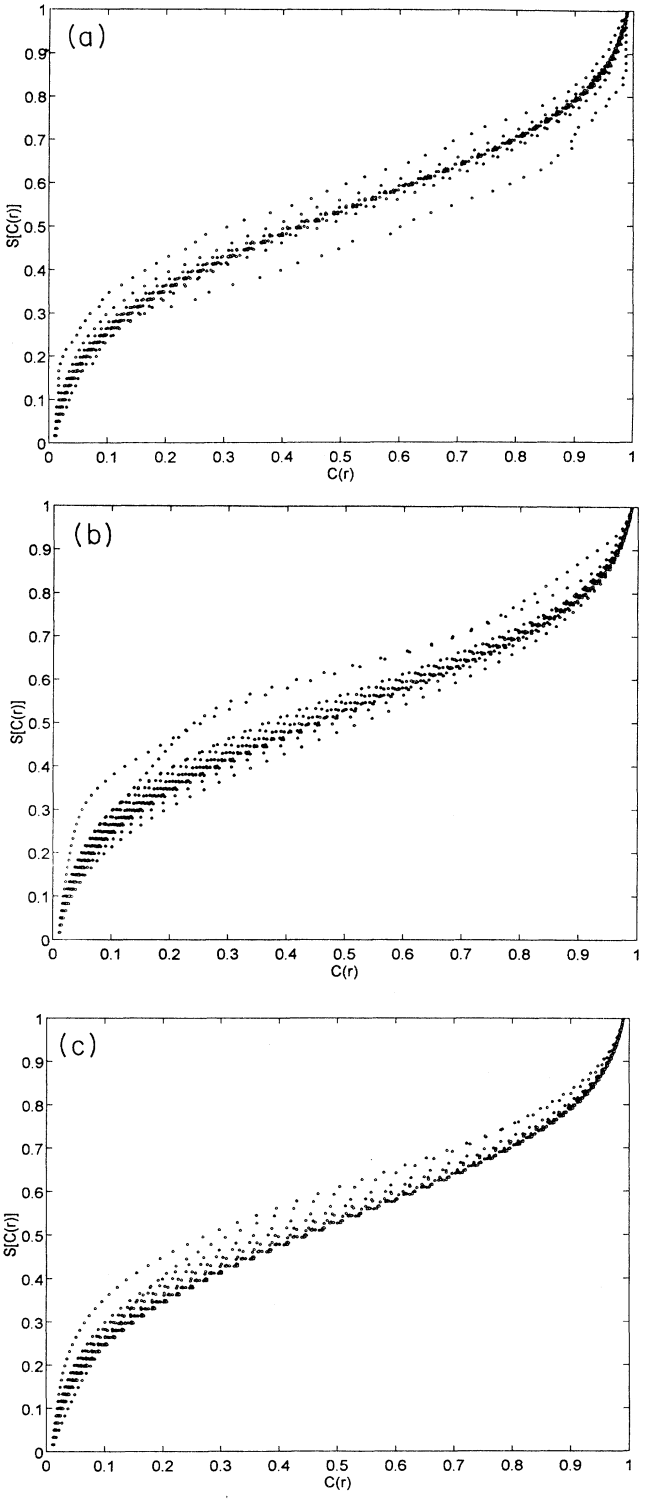

FIG. 5. Cumulative distribution function $S(C_m(r))$ of the correlation integral of the Lorenz attractor using the following combinations of parameters. (a) Lag = 10; embedding dimension = 50; number of embedding vectors = 20,40,...,200,700,...,4700. (b) Lag = 1,2,...,20; Embedding dimension = 50; number of embedding vectors = 1024. (c) Lag = 10; embedding dimension = 5,10,...,100; number of embedding vectors = 1024.

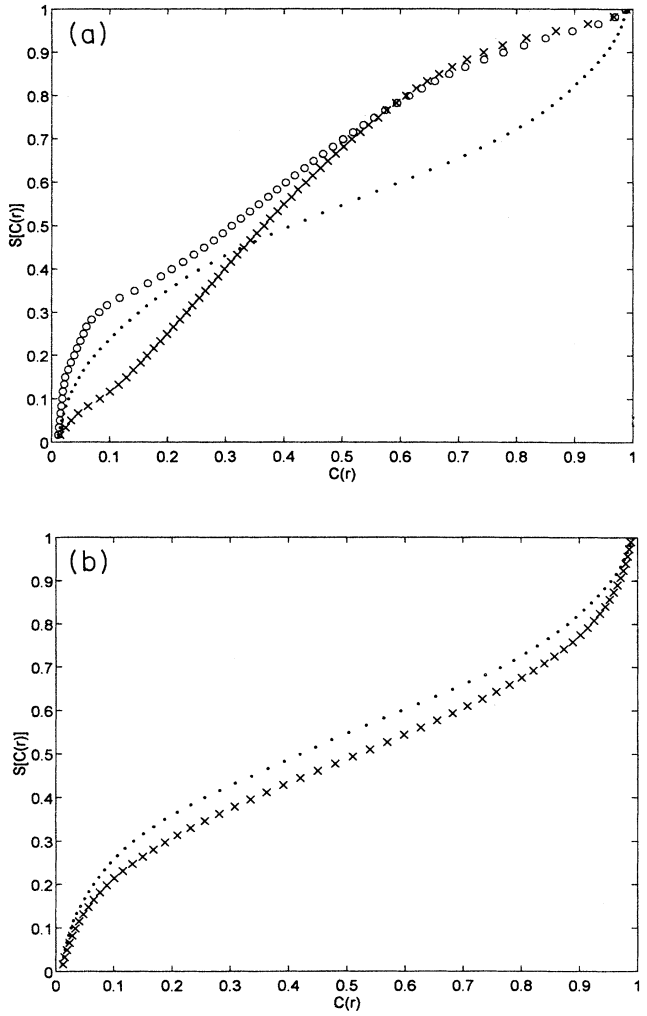

FIG. 6. (a) $S(C(r))$ vs $C(r)$ for the Lorenz (●), Rössler (×), and Mackey-Glass ($\delta=17$) (○) attractors. All were calculated using 1024 embedding vectors with a lag of 10 in 50 dimensions. (b) $S(C(r))$ vs $C(r)$ for the Mackey-Glass ($\delta=150$) (●) and filtered noise data (×) using 4096 embedding vectors with a lag of 10 in 50 dimensions.

mension. Nevertheless, as we show below, the limiting forms of the $S(C_m(r))$'s can be sufficiently distinctive as to lead to meaningful comparisons using the Kolmogorov-Smirnov test.

Figure 6(a) shows $S(C_m(r))$ vs $C_m(r)$ for the Lorenz, Rössler, and Mackey-Glass ($\delta=17$) attractors. All were calculated using 1024 embedding vectors with a lag of 10 in a 50-dimensional embedding space. Figure 6(b) shows $S(C_m(r))$ vs $C_m(r)$ graphs for the Mackey-Glass ($\delta=150$) and filtered noise data using 4096 embedding vectors with a lag of 10 in a 50-dimensional embedding space. Both sets of graphs show that graphs corresponding to the different attractors are clearly distinguishable. This distinguishability is quantified using the Kolmogorov-Smirnov test in the next section.

IV. COMPARING SIMILAR ATTRACTORS

Table I shows results of comparing the correlation integrals of the "clean" Lorenz attractor with noise-corrupted Lorenz attractors having SNR's of 5, 10, 15, and 20 dB as well as comparisons among the various noise-corrupted data sets. In each entry in the table, the first line gives the median while the second gives the mean and standard deviation of ten comparisons in which, for each comparison, noise-corrupted data of the appropriate SNR are generated anew from the uncorrupted data. In each case, 1024 embedding vectors were constructed using a lag of 10 in a 50-dimensional embedding space.

Since we are concerned with probabilities which are not likely to be symmetrically distributed, the median is a better descriptor than the mean. In cases when there are a lot of values near zero, for instance, these small values contribute negligibly to the mean, which can be unduly influenced by a few large outliers. The variance, however, still provides some indication of the spread of the values being considered.

Table I shows that the test is able to assert that, even with as small a signal-to-noise ratio as 15 dB, the noise-contaminated attractor is still unambiguously the Lorenz attractor. At 10 dB, the effects of noise intrude. Howev-

TABLE I. Clean vs noise-corrupted Lorenz data. The Kolmogorov-Smirnov probability Q_{KS}, that the correlation integral of the data file identified by the column and that of the file identified by the row come from the same distribution. The first line of each entry is the median obtained in ten independent determinations. The second line gives the mean and standard deviation. In each calculation, noise-corrupted data sets were independently generated from the uncorrupted data.

	Median and mean ± standard deviation			
SNR	20 dB	15 dB	10 dB	5 dB
Uncorrupted	0.97	0.78	0.047	0.015
	0.95±0.06	0.63±0.40	0.14±0.28	0.08±0.17
20 dB		0.96	0.16	7.7×10^{-4}
		0.84±0.19	0.19±0.16	0.12±0.25
15 dB			0.89	0.001
			0.62±0.41	0.25±0.32
10 dB				0.10
				0.30±0.29

er, the test itself makes it possible to evaluate how "noise-like" the 5 and 10 dB attractors are (see below). It is interesting, but perhaps not very surprising, that one can conclude to a higher degree of confidence that the noise-corrupted signals come from the same distribution (15 vs 20 dB or 10 vs 15 dB) than is possible in a comparison of the noisy signals with the uncorrupted signal.

When the Lorenz, Mackey-Glass ($\delta=17$), and Rössler attractors are compared, again using 1024 vectors with a lag of 10 in a 50-dimensional embedding space, the resulting Kolmogorov-Smirnov probability that any two of these attractors are the same is zero to within the resolution of the single-precision calculation used. A similar comparison of the Mackey-Glass ($\delta=150$) and filtered random noise using 4096 vectors with a lag of 10 in a 50-dimensional space gave a probability of the null hypothesis that is less than 10^{-8}.

Table II compares a number of data sets with their phase-randomized and Gaussian-scaled surrogates. The data sets used are (1) the "clean" Lorenz attractor, (2) the noise-contaminated Lorenz attractor with SNR $=5$ dB, (3) the noise-contaminated Lorenz attractor with SNR $=10$ dB, (4) the Mackey-Glass attractor with a delay, $\delta=150$, (5) filtered random noise. The table entries were obtained by comparing the data with 100 surrogates of each type. Again, both the median and the mean $\pm$ standard deviation are shown.

Table II shows that in the case of the Lorenz attractor with a SNR of 5 dB there is some ambiguity in distinguishing it from its surrogates, although the values of the mean, median, and standard deviation are compatible with a zero probability that it is the same as its surrogates. The rest are less ambiguous; the figures for the Mackey-Glass ($\delta=150$) and the filtered random noise leave no doubt whatever that the former is not linearly correlated noise but that the latter is indistinguishable from noise.

TABLE II. Comparisons with surrogate data. Kolmogorov-Smirnov probabilities Q_{KS} that the correlation integral of the data file listed in column 1 comes from the same distribution as the correlation integral of its phase-randomized surrogates (column 2) or of its Gaussian-scaled surrogates (column 3). The first line of each entry is the median while the second line gives the mean and standard deviation of calculations using 100 surrogates. Calculations using Lorenz data used 1024 embedding vectors, those using Mackey-Glass data and filtered random noise used 4096.

	Median and mean $\pm$ standard deviation	
Data file	Phase randomized	Gaussian scaled
Lorenz	7.6×10^{-5}	9.8×10^{-5}
(uncorrupted)	0.04 ± 0.14	0.09 ± 0.24
Lorenz	0.0016	0.058
(SNB $=10$ dB)	0.15 ± 0.27	0.26 ± 0.34
Lorenz	0.21	0.28
(SNB $=5$ dB)	0.31 ± 0.31	0.38 ± 0.35
Mackey-Glass	0.00	0.00
($\delta=150$)	0.00 ± 0.00	$<10^{-5}$
Filtered random	0.93	0.98
	0.70 ± 0.37	0.81 ± 0.29

V. SUMMARY AND DISCUSSION

Most nonlinear dynamical measures used to characterize complex dynamical systems require information about the small scale structure of the system's attractor. These structures are easily obliterated by noise and, in the case of higher-dimensional attractors, cannot be made apparent without large amounts of data. In situations where these measures are not computable or in the more general situation when the need is to compare attractors rather than to characterize individual attractors, comparing correlation integrals using the Kolmogorov-Smirnov test is potentially useful. This procedure is less sensitive to noise than the usual nonlinear measures that require fine-grained information, and is usable even when these measures cannot be computed. In addition, it is relatively insensitive to embedding lag or embedding window, making it unnecessary to spend much technical effort in seeking optimal values of these parameters.

Application of the test to compare the Lorenz attractor with various noise-corrupted versions shows that it is able to identify the Lorenz attractor with near certainty even when the signal-to-noise ratio is as low as 15 dB. The test is less able to identify the attractor when the SNR is 10 dB or less, but by using it in conjunction with surrogate data it is possible to establish with considerable confidence that the data set is not the kind of correlated noise represented by the surrogates.

It successfully distinguishes attractors with nearly the same dimension, even in the case of the Mackey-Glass ($\delta=150$) attractor and filtered random noise. Both have correlation dimensions in excess of 4 which could not be reliably calculated using relatively small data sets. When used with surrogates, it shows with almost unit probability that the filtered random data set is indeed, random and that the correlation integral of the Mackey-Glass ($\delta=150$) is clearly different from those of its surrogates.

These results suggest that use of the Kolmogorov-Smirnov test to compare correlation integrals is a potentially useful additional tool for the analysis of complex dynamical signals in situations where relatively high levels of noise are unavoidable or when there are stringent lower limits on sampling intervals or upper limits on the total measurement times. All of these conditions exist in the study of physiological systems. Neural signals, for instance, are intrinsically noisy; the neural membrane effectively acts as a low-pass filter and severely attenuates signals with frequencies in excess of 15 Hz, and the nervous system is notoriously nonstationary (see, e.g., Ref. [47] and references quoted therein). Neurophysiological signals are inevitably noisy, and measurements yield either short, or long but nonstationary data sets. As seen above, these conspire against a proper application of the usual fine-grained measures of nonlinear behavior. For these and similar situations, coarse-grained tools such as those proposed in Ref. [24] and that proposed here may well be more appropriate and useful.

ACKNOWLEDGMENTS

A.M.A. would like to acknowledge partial support from NIH Grant No. NS32983 and the Navy-ASEE (American Society for Engineering Education) Sabbatical Support Program, and is grateful for the hospitality of the Naval Air Warfare Center, Aircraft Division, Warminster. P.E.R. would like to acknowledge partial support from NIH Grants No. NS19716 and No. NS32983 and the University of Western Australia. He also wishes to thank A. I. Mees and K. Judd for illuminating discussions.

[1] *Dimensions and Entropies in Chaotic Systems,* edited by G. Mayer-Kress (Springer-Verlag, Berlin, 1986).
[2] *Measures of Complexity and Chaos,* edited by N. B. Abraham, A. M. Albano, A. Passamante, and P. E. Rapp, (Plenum, New York, 1989).
[3] *The Ubiquity of Chaos,* edited by S. Krasner (AAAS, Washington, 1990).
[4] *Interpretation of Time Series from Nonlinear Systems,* edited by P. G. Drazin and G. P. King (North-Holland, Amsterdam, 1992).
[5] P. E. Rapp, A. M. Albano, I. D. Zimmerman, and M. A. Jimenez-Montaño, Phys. Lett. A **192**, 27 (1994).
[6] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[7] F. Takens, in *Dynamical Systems and Turbulence (Warwick, 1980),* edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, New York, 1981).
[8] R. Mañé, in *Dynamical Systems and Turbulence (Warwick, 1980),* edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, New York, 1981).
[9] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[10] J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986).
[11] S. Sato, M. Sano, and Y. Sawada, Prog. Theor. Phys. **77**, 1 (1987).
[12] R. Stoop and P. F. Meier, J. Opt. Soc. Am. B **5**, 1037 (1988).
[13] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).
[14] H. G. Solari and R. Gilmore, Phys. Rev. A **37**, 3096 (1988).
[15] H. G. Solari and R. Gilmore, Phys. Rev. A **38**, 1556 (1988).
[16] N. B. Tufillaro, H. G. Solari, and R. Gilmore, Phys. Rev. A **41**, 5717 (1990).
[17] P. Melvin and N. B. Tufillaro, Phys. Rev. A **44**, R3419 (1991).
[18] D. T. Kaplan, Physica D **73**, 38 (1994).
[19] P. Grassberger and I. Procaccia, Phys. Rev. Lett **50**, 346 (1983).
[20] H. G. E. Hentschel and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).
[21] K. Pawelzik and H. G. Schuster, Phys. Rev. A **35**, 481 (1987); **36**, 4529 (1987).
[22] J. Theiler, J. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer, in *Nonlinear Modeling and Forecasting,* edited by M. Casdagli and S. Eubank (Addison-Wesley, Reading, MA, 1991); Physica D **58**, 77 (1992).
[23] P. E. Rapp, A. M. Albano, T. I. Schmah, and L. A. Farwell, Phys. Rev. E **47**, 2289 (1993).
[24] P. E. Rapp, G. Goldberg, A. M. Albano, M. B. Janicki, D. Murphy, E. Niemeyer, and M. A. Jiménez-Montaño, Int. J. Bifurc. Chaos **3**, 525 (1993).
[25] M.G. Kendall and A. Stewart, *The Advanced Theory of Statistics,* 3rd ed. (Hafner, New York, 1973), Vol. 2.
[26] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1986).
[27] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[28] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[29] J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[30] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp, Phys. Rev. A **38**, 3017 (1988).
[31] A. M. Albano, A. Passamante, and M. E. Farrell, Physica D **54**, 85 (1988).
[32] H. G. Schuster, in *Measures of Complexity and Chaos [2].*
[33] W. Liebert, K. Pawelzik, and H. G. Schuster, Europhys. Lett. **14**, 521 (1991).
[34] Th. Buzug and G. Pfister, Physica D **58**, 127 (1992).
[35] H. D. I. Abarbanel and M. M. Schuschik, in *Complexity and Chaos,* edited by N. B. Abraham, A. M. Albano, A. Passamante, P. E. Rapp, and R. Gilmore (World Scientific, Singapore, 1993).
[36] J. G. Caputo, B. Malraison, and P. Atten, in *Dimensions and Entropies in Chaotic Systems,* edited by. G. Mayer-Kress (Springer-Verlag, Berlin, 1986).
[37] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).
[38] D. T. Kaplan, in *Complexity and Chaos* [35].
[39] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[40] O. Rössler, Phys. Lett. **57A**, 397 (1976).
[41] M. Mackey and L. Glass, Science **197**, 287 (1977).
[42] J. D. Farmer, Physica D **4**, 366 (1982).
[43] J.-P. Eckmann and D. Ruelle, Physica D **56**, 185 (1992).
[44] M. A. H. Nerenberg and C. Essex, Phys. Rev. A **42**, 7065 (1990).
[45] L. A. Smith, Phys. Lett. A **133**, 283 (1988).
[46] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurc. Chaos, **1**, 521 (1991).
[47] P. E. Rapp, The Biologist **40**, 89 (1993).