Statistics for mathematical properties of maps between time series embeddings

Louis M. Pecora, Thomas L. Carroll, and James F. Heagy
Code 6341, Naval Research Laboratory, Washington, D.C. 20375
(Received 5 April 1995)

We develop a set of statistics which are intended to characterize in terms of probabilities or confidence levels whether two data sets are related by a mapping with certain mathematical properties. Given these statistics we can ask how confident we can be that the mapping is continuous, injective, differentiable, or has a differentiable inverse. The intended use is for experimental or numerical situations in which multiple time series are generated and one wants to know what relation exists among them, but the mapping between them is unknown or intractable. Examples of applications are testing filtered chaotic data for continuity and differentiability, testing two data sets for synchronization (in the most general sense), testing one data set for determinism forward and backward in time, and determining when transformations on two- or three-dimensional images are well behaved (diffeomorphisms). We test the statistics on several of these cases and show that they are useful for characterizing relations between data sets and for shedding light on phenomena which occur when data are transformed, for example, a dimension increase on filtering a chaotic data set.

PACS number(s): 05.45.+b, 47.52.+j, 02.40.−k, 02.50.−r

I. INTRODUCTION

The dynamics of physical systems can be represented concisely by geometric objects in an n-dimensional physical phase space. These are trajectories, attractors, basins of attraction, unstable manifolds, etc. For some time special techniques, centered around time-series reconstructions, have allowed scientists to connect experimental results to those geometrical objects. This technique consists of forming a mapping between a time series of N measurements $(h_1,h_2,h_3,\ldots,h_N)$ and vectors in some simple d-dimensional space, viz. $\mathbf{v}^i=(h_i,h_{i+1},h_{i+2},\ldots,h_{i+d-1})$. These vectors then form points on a trajectory and the d-dimensional space becomes a representation of the actual physical phase space of the system.

The main theorems connecting these two realms of measurement and physical phase space are Takens' theorem [1], its precursors (e.g., [2]), and its sequels (e.g., [3]). The essence of these mathematical demonstrations is that the trajectory formed from the time series is diffeomorphically related to the actual phase-space trajectory of the dynamical system. Diffeomorphically means that the map between the original, physical phase space and the time-series reconstructions has no tears (is continuous), does not glue together geometrically separate points (is injective), and is smooth both from the phase space to the reconstruction (differentiable) and in the reverse direction (differentiable inverse). In addition, it is an embedding in that it preserves the dimension of the manifold on which the trajectory exists (the Jacobian of the map has a rank equal to the manifold's dimension in the original phase space).

Some or all of these properties are important for the preservation of a link between the actual system dynamics and a faithful reconstruction of the physical system's motion. This link is crucial to making confident conclusions about experimental results. There are several situations in which such a link is desirable, but tests for reliability are absent.

The general question we are investigating here is, given two time series, when can we say with some confidence that there is a map relating them which is a diffeomorphism or which at least has some of the properties of a diffeomorphism, such as continuity or differentiability? In this case we know data points in the domain and associated data points in the range of the function, but we do not know the function directly. This is a rather general problem transcending time-series analysis. We make some comments on this point in the conclusions.

We have found several statistical approaches to properties such as continuity, injectivity, and differentiability which provide quantitative tests for each property. Specifically, we have focused on statistical tests to determine when one can be confident that two data sets have a particular property relating them (e.g., one is a continuous function of the other). We introduce these statistics here and show several common applications including one which sheds light on the question of filtering chaotic data. We first motivate the search for statistical approaches to such properties and review some recent progress.

Also because we are staying close to the mathematical meaning of the properties of functions, we use the word "embedding" only to mean a diffeomorphism that preserves rank. In much of the literature the word embedding is also used loosely to describe the reconstruction of a phase-space trajectory by a delay technique. For the latter we will use the word reconstruction. We want to avoid confusing the two uses of embedding here.

II. MOTIVATION

A. Filtering

Although filtering of time-series data is often practiced with great liberty, the work of Badii *et al.* [4] has shown that certain filters can add to the dynamics of the experimental system, distorting the reconstruction of the attractor and increasing its dimension. Other filters [5] do not affect the reconstruction as much, in some case not at all. The process of filtering a data set can be written as a transformation on the attractor; the mapping is between the original attractor and the filtered version. The question becomes, when is this transformation continuous and differentiable? That link, if established, would guarantee the preservation of dimension [6].

B. General synchronization

Dynamical systems, even chaotic ones, can be in synchrony [7–19]. Simple synchronization occurs in compound systems when two dynamical variables asymptotically converge to the same dynamical values: they vary in time, but are always equal at any time. Afraimovich *et al.* have suggested a more general definition of synchronization [8]. This latter type occurs when two variables in a physical system are not necessarily equal, but are in a one-to-one, smooth relation with each other. In more mathematical terminology they are diffeomorphically related. This can occur, for example, when nearly identical systems are coupled, but the systems have slightly different parameters [8,9,20–22]. Experimentally, this question is of great interest in situations where multiple measurements are taken simultaneously on the system and one is interested in discerning whether any of the signals are in synchrony with any of the others (e.g., in brain/neural activity during decision-making tests [23]).

C. Determinism

The question of whether a time series even represents a deterministic system is of fundamental interest. In this case the mapping is between one time series and another which is just a time-shifted version of the first. For example, certain colored noise systems can have many properties of deterministic systems, like a finite fractal dimension [24]. A basic property one expects from a deterministic system is that of continuity forward in time. That is, points very close in phase space should map forward to points still close in phase space. Beyond this "lowest" form of determinism one could ask that the forward mapping be smooth. In that case we would like to extend the test for continuity to a test for (forward) differentiability. We can also ask if the dynamics are invertible (a test for injectivity of the forward mapping) and smoothly invertible (differentiability of the backward mapping). Each test would help us to determine what type of system we are measuring and how best to model it.

D. Recent work

Much work has been done on gauging whether the initial reconstruction of a time series trajectory is a good one. Good here means the reconstruction has some properties of or relating to Taken's embedding theorem.

The problem of determining the correct embedding dimension has been attacked from two directions. One is the use of false nearest neighbors first suggested by Schuster *et al.* [25] and later Liebert *et al.* [26], and refined by Kennel *et al.* [27], in which one reconstructs the times series in increasingly larger dimensions until there cease to be pairs of points which are nearest neighbors in one dimension, but are not nearest neighbors in the next larger dimension. This is essentially a test for continuity; points close in the original manifold should not be mapped to distant points in the image, which is the reconstruction in this case. Kennel *et al.* propose a simple statistic and some guidelines on using it to help determine when one has false nearest neighbors.

Another approach is to test for the smoothness of the reconstruction. This test and an associated statistic was originally presented by Kaplan and Glass [28] with recent variations by Salvino and Cawley [29,30]. Essentially, a local vector field is approximated on a grid over the reconstructed attractor. The statistic attempts to measure the smoothness and the significance of the local vector field. This is essentially a test for differentiability of the reconstruction, with a useful statistic to test how meaningful the result is.

Wayland *et al.* [31] developed a simple test for determinism which measures the variance of measured error in translation vectors for nearby phase-space points to their images at future times. This gives a good indication of how far into the future one can extrapolate until the relation between nearby points becomes equivalent to white noise. Their translation error quantity is like a crude continuity test, although they do not associate a statistical or probabilistic quantity with it.

Other work tests for the best time-interval size to choose in generating a time series. This is an important issue, but is not central to the present paper. See Refs. [32–36] and the references there in for more information.

The issue of determinism has been approached using surrogate data by Theiler [37,38]. Here one creates another data set (the surrogate) which has some statistical properties (e.g., autocorrelations) which are identical to the actual time series. The surrogate, however, is not a deterministic series. For example, it can be generated from a Fourier series which has the same coefficient amplitudes as the original series, but has its phases randomized. This provides a good statistical null hypothesis [39], however, the relation to analysis-calculus properties and their statistics is unclear.

In his recent approach to finding a statistic for determinism, Kaplan [40] has come closest to deriving statistics for mathematical proprieties of time series. As we noted above, the "lowest level" on which one could define forward determinism is to check for continuity in the mapping between a reconstruction point at one time and its image at a later time. Kaplan's approach is to check

the relation between the number of points at the present time in a small set (defined by δ) and the number of points at a later time in an image set (defined by ε). These numbers are then compared to those expected for various random distributions which form the null hypotheses. Applications to some simple maps and some vector fields show that the statistic holds promise as a gauge of determinism. Throughout Kaplan often implicitly assumes more than continuity. Near-neighbor points are often assumed to have a linear relation to their images. This assumes differentiability. It is not always clear what level of determinism Kaplan's statistic is testing, although that should be easy to clear up in applications. We note that this test comes closest of all to connecting a mathematical analysis definition (that of continuity by δ-ε distances) with statistics. We will return to this point later and use it as the foundation for our statistics, although our approach differs somewhat from Kaplan's.

Another recent approach to a statistic for a mathematical property (diffeomorphism) is that of Rul'kov *et al.* [41]. They were interested in determining when two systems were in generalized synchronization, as determined from their respective time series. One can "line up" the two time series and check properties in the mapping from one to the other. Here, as in determinism, we can think of testing at two levels. One is just to test for a one-to-one property in the mapping from one reconstruction to another (a homeomorphism). Another is to test for the smoothness in this mapping and its inverse (a diffeomorphism). The test Rul'kov *et al.* actually developed is a mutual false nearest-neighbor statistic modeled on previous definitions of false nearest neighbors [27]. Calculation of this statistic requires the average over the two reconstructions of the quotient of two ratios. One ratio is the nearest-neighbor distance on the first reconstruction to the distance between the corresponding points on the second reconstruction. The other ratio is just the reverse; it is the nearest-neighbor distance on the second reconstruction to the distance between the corresponding points on the first reconstruction. If there is a smooth one-to-one mapping between the time series, the quotient of these two ratios should be close to unity. Here, again, the tacit assumption is that of differentiability (just a one-to-one continuous mapping would not necessarily give unity for the quotient). The statistic appears to be a robust one at least for some simple cases and (importantly) for time series from a real circuit. However, faithfulness to actual mathematical definitions of continuity or differentiability is not clear. We will show later that this statistic is a simple approximation to local statistical correlations which we employ for differentiability testing.

E. Objectives

Many of the above methods implicitly and occasionally explicitly assume certain analytical properties. In the following we will develop statistical tests which are *explicitly* related to certain analysis properties. We attempt to derive the statistics by closely following the original analysis definition of the property (e.g., the δ-ε definition of continuity). In this way compound statistics can be easily derived to characterize a compound mathematical property (e.g., continuously differentiable, C^1).

One of our aims is to make explicit use of the statistical tool of the null hypothesis. This is an assumption that is usually contrary to what one is trying to show. The object will be to test if we can reject the assumption and thereby lend confidence to what we wish to show. The reason for using such a hypothesis is that it helps to quantify our statistic, defined below, when we do not have *a priori* a probability distribution for the particular system we are studying. Not all tests developed for chaotic time series use this approach, but several of the more fruitful ones do [30,37,38,40,42].

In all the following we will try to characterize the *entire* reconstruction. To do this we will average our statistical quantities over all points or over a random collection of points in the time series. We feel this generates an average statistic that can be used to gauge how well other average quantities such as fractal dimension or Lyapunov exponents are preserved from one data set to the other. In doing this we are tacitly assuming some sort of ergodicity. Certainly one could also focus on particular points or regions of the attractor and apply the same statistics.

To generate the actual statistics we work primarily with the null hypotheses that the two data sets we are testing are randomly related, with no correlations. Other hypotheses are certainly possible (e.g., that the time series are correlated noise [40]). Similarly, we suspect that there may be other approaches to our objectives, refinements to our results, and certainly other analysis properties to characterize statistically.

III. ANALYSIS STATISTICS

We are interested in characterizing, in a statistical sense, the mathematical property of embedding from one time series (the first) to another (the second). See Fig. 1(a) for a diagram of the mapping. We emphasize here that we are in the position of knowing data points in the domain and range of a function, but we do not explicitly know the function in the sense of knowing the rules for generating a range point given a domain point.

Characterizing an embedding breaks down to characterizing several elementary mathematical properties of a mapping; continuity, injectivity, differentiability, and rank invariance; we sometimes refer to these as analysis properties. An embedding can heuristically be viewed in three steps which serve to introduce the nomenclature. First an embedding is a homeomorphism from the first time series onto the second. A mapping is homeomorphic if it is continuous, injective, and has continuous inverse. In terms of time series this means [see Fig. 1(b)] there are no points in the first time series' phase space which are arbitrarily close, but which are mapped to distant points on the second time series. Injectivity, roughly, is just the opposite: no arbitrarily close points in the second time series are mapped by the inverse to distant points back in the first time series [see Fig. 1(c)]. Continuity guarantees that there are no rips or tears in the

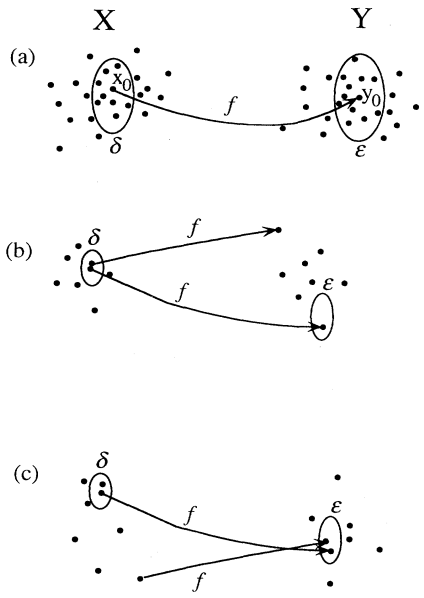

FIG. 1. (a) Diagram of mapping f: $X \rightarrow Y$, (b) diagram of discontinuity, (c) diagram of noninjectivity.

mapping. Injectivity (and inverse continuity) guarantee that the mapping does not glue together separate parts of the first time series. Points that are close in one time series are *always* close in the other.

Second, an embedding is differentiable. Differentiability is a smoothness criterion. In one direction [Fig. 1(b)] we require the mapping to have a derivative. This means the mapping can be approximated locally by a linear operator. In the other direction [Fig. 1(c)] we want the inverse mapping to have this quality. This puts even more restrictions on the map which will also lead to a statistic beyond continuity.

Finally, an embedding must preserve rank. What this means is that if we start with an object on some surface that is of a certain dimension, we want to map to an object on another surface that has the same dimension. This is guaranteed by requiring that the rank of the Jacobian of the mapping be equal to the original surface dimension. This requirement if often ignored in work on embeddings, but it is an important part of the definition (see Ref. [43]). We do not focus any attention on a statistical test for the rank, but we do note that some attention must be paid to rank in the differentiability test below.

A. Continuity

We start with a basic concept of continuity of a mapping f from a space X to another space Y. We assume there is a metric $\| \ \|$ on each space. In practice we use the Euclidean metric. The function f is continuous at a point $\mathbf{x}_0 \in X$ if $\forall \varepsilon > 0 \exists \delta > 0$ such that $\|\mathbf{x}-\mathbf{x}_0\| < \delta \Longrightarrow \|f(\mathbf{x})-f(\mathbf{x}_0)\| < \varepsilon$. Simply put, if we pick an open ε-sized set around $f(\mathbf{x}_0)$, then there is some small enough δ-sized set around $\mathbf{x}_0$ from which *all* of the points are mapped by f into the ε set. This guarantees that nearby points are not mapped to distant points as in Fig. 1(b).

A statistical version of the standard analysis statement of continuity when we only know some points in the domain and range of f, but not f itself, simply amounts to counting points in δ and ε sets. The statistical version can be generated in two steps. First we construct an algorithm to select points that, given δ and ε, are consistent with the mathematical definition of continuity. Second we apply an appropriate probability distribution consistent with the null hypothesis we have chosen. Note that these two steps are independent. Given points selected by the algorithm we can use different null hypotheses and thereby generate different probability distributions for our statistic. These two steps are used to generate other statistics, too, for example, the differentiability statistic below.

For the first step, the algorithm for finding points, start with ε and δ values. Find all the points $\mathbf{x}_i$ which are within a distance δ of $\mathbf{x}_0$. Now check to see if *all* the images $\mathbf{y}_i = f(\mathbf{x}_i)$ of those points are within a distance ε of $\mathbf{y}_0 = f(\mathbf{x}_0)$. If not, decrease δ by some factor, find a new set of $\mathbf{x}_i$'s, and check that the new images are within ε of $\mathbf{y}_0$. We repeat this process until we either have all $\mathbf{y}_i$'s in the ε set or δ has been decreased so far that we cannot find any $\mathbf{x}_i$ points, other than $\mathbf{x}_0$. All we are doing here is requiring the points to fulfill the definition of continuity above.

For the second step we introduce a null hypothesis which we use to generate a probability for our simple statistic for continuity. Let there be N points in the reconstructions, let n_δ be the number of $\mathbf{x}_i$ points found in the δ set, and n_ε be the total number of $\mathbf{y}_i$ points found in the ε set. We exclude $\mathbf{x}_0$ and $\mathbf{y}_0$ since they are always present by construction and their presences should not influence the statistic. Also note that, in general, $n_\varepsilon > n_\delta$ since points from other parts of the attractor outside the δ set may be mapped into the ε set, too. That latter fact does not affect continuity. Our null hypothesis is that the $\mathbf{y}$ points are randomly and independently distributed over the Y space reconstruction with respect to the $\mathbf{x}$ points. For this situation the probability p of one of the n_δ δ points mapping into the ε set is just the ratio n_ε/N. Here we assume that n_ε/N is a good estimate for the actual probability. The probability of all n_δ $\mathbf{x}$ points mapping into the ε set is then $p^{n_\delta} = (n_\varepsilon/N)^{n_\delta}$.

If this probability is low enough, then we can reject our null hypothesis; the $\mathbf{x}$ and $\mathbf{y}$ points are not randomly related. Note that this does not mean we have proven they are related in any particular way. Rejection of the null hypothesis is not an automatic acceptance of the alternate hypothesis [44,45] that the mapping is continuous. Rather, it gives us confidence that the effect we are seeing, in this case n_δ points all mapped into the ε set, it not caused by some random coincidence.

We express this confidence by quantifying what we meant above by having a probability that is "low enough." First note that we are really dealing with a bi-

nomial distribution [46] $b(m;n_\delta,p)$, which gives the probability of finding m points out of n_δ inside the ε set if the probability for finding one point in the ε set is p. In our case our event is in the "tail" of the distribution, $m=n_\delta$, and so the probability of this happening is p^{n_δ}. The likelihood of this happening we define as the ratio of the probabilities $p^{n_\delta}/p_{\max}$. The quantity $p_{\max}$ is the maximum in the binomial distribution which occurs, usually, at some intermediate $m<n_\delta$ value. In other words, it is not enough to have the probability p^{n_δ} of the event be small, but it must be small compared to the probability of the most likely event, which here occurs with probability $p_{\max}$.

Then we define the confidence as the continuity statistic $\Theta_{C^0}(\varepsilon,j)=1-p^{n_\delta}/p_{\max}$, where j is the index of the point at which we are testing for continuity ($\mathbf{x}_0=\mathbf{x}_j$). The subscript C^0 shows that the statistic is for continuity (differentiability is not assumed here) and the dependence on ε is made explicit. When $\Theta_{C^0}\approx 1$ we are confident, *vis à vis* the null hypothesis, that we have a continuous function. When $\Theta_{C^0}\approx 0$ we are *not* confident that we have continuity since we cannot reject the null hypothesis that the event happened by accident.

We note several things about our statistic in relation to time-series applications. First we typically have several thousand points in a time series and for small ε values we would expect n_ε to be at least an order of magnitude smaller. This implies that finding only a few δ points that all map into the ε set would be all that is required to reject an assumption of random relation between the **x**'s and **y**'s. This is in agreement with Kaplan [40] who found that for determinism tests, close points mapping forward in time to close points, was a rare event for a sequence of random points. In fact we often find that in calculating $\Theta_{C^0}(\varepsilon,j)$ at points on an attractor we get "binary" results: either $\Theta_{C^0}(\varepsilon,j)\approx 1$ or 0. Hence, this property is generic for continuity of any mapping and is not restricted to tests for determinism.

Second, these results give no guidance on choosing ε, except that if the confidence in continuity is low below some ε_c value, then one must work at a lower resolution (larger ε value) in the Y space. In working below ε_c we cannot be sure if the mapping will have the properties of a continuous function. Working above ε_c similarly does not guarantee that we are in a small enough neighborhood where we can be sure there are no effects like curvature. Estimation of how small an ε we need must come from other tests. These are important considerations when, as in an experiment, we do not know the form of the dynamics and/or the mapping between the two time series.

In our applications we will examine the average of $\Theta_{C^0}(\varepsilon,j)$ over the attractor or over a set of randomly chosen points on the attractor. We write

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

where we calculate $\Theta_{C^0}(\varepsilon,j)$ at n_p points. When no confusion arises we drop the explicit dependence on ε.

B. Inverse continuity

The statistic for inverse continuity of the mapping f is actually easy to generate. This statistic must cover the following mathematical statement: at a point $\mathbf{y}_0\in Y$ if $\forall\,\delta>0\exists\varepsilon>0$ such that $\|\mathbf{y}-\mathbf{y}_0\|<\varepsilon \Longrightarrow \|f^{-1}(\mathbf{y})-f^{-1}(\mathbf{y}_0)\|<\delta$ the map f has a continuous inverse. See Fig. 1(c) for a graphical description of when this definition fails. Note that this is just the reverse of the continuity definition. Hence we can formulate a statistic $\Theta_{I^0}(\varepsilon,j)$ just like $\Theta_{C^0}(\varepsilon,j)$ above, where by I^0 we imply the continuous inverse function statistic with no differentiability assumption. Then we write $\Theta_{I^0}(\delta,j)=1-p^{n_\varepsilon}/p_{\max}$, where $p=n_\delta/N$. Note that in these formulas $p_{\max}$, n_δ, and n_ε are *not* the same values as for the continuity statistic, even when the set sizes are the same. Analogous to the continuity statistic the requirement is that all n_ε points map into the δ set and the δ set can, in general, contain points mapped from regions of the attractor other than the ε set. In general, $n_\delta>n_\varepsilon$ for inverse continuity.

As for continuity we have confidence in the injectivity and inverse continuity when $\Theta_{I^0}(\varepsilon,j)\approx 1$ and little confidence when the statistic is nearly zero. Other comments on the continuity statistic translate to the inverse continuity statistic with appropriate interchange of **x**'s, **y**'s, δ's, and ε's. Finally, we have the average statistic

$$\Theta_{I^0}(\varepsilon)=\frac{1}{n_p}\sum_{j=1}^{n_p}\Theta_{I^0}(\varepsilon,j)\ , \tag{2}$$

which we will use to characterize the entire mapping relating the two time series.

C. Injectivity and homeomorphism

Injectivity, whether a function is one to one, is determined from the Θ_{C^0} and Θ_{I^0} statistics. The relation between the two reconstructions which we have been calling f is not guaranteed to be a function from X to Y in the strictest mathematical sense. The mapping f is a function providing that there is only one point in the range of each domain point. For example, $f(x)=x^2$ is a function, but $f(x)=\sqrt{x}$ is not, unless we put some restrictions on the range. Our continuity statistic Θ_{C^0} actually simultaneously tests for "functionality" since by proving continuity we also are showing that the (less demanding) requirement of being a function is fulfilled. The inverse continuity statistic simultaneously tests whether f^{-1} is a function (only one point in the domain of f for each range point), as well as whether it is continuous. Combining Θ_{C^0} and Θ_{I^0} provides a statistical test of injectivity. This combination can be expressed as the new compound statistic $\Theta_{C^0}(\Theta_{I^0})$. In fact, if the product $\Theta_{C^0}(\Theta_{I^0})$ is near 1, we can be confident that we are dealing with a mapping that is a homeomorphism.

We do not make use of this statistic here, but this serves to show that other mathematical properties can have statistics associated with them and that some can be related to more basic statistics such as Θ_{C^0} or Θ_{I^0}.

D. Differentiability

For differentiability we again start with a mathematical definition. A mapping f: $X \rightarrow Y$ is differentiable at $\mathbf{x}_0$ if there exists a linear operator A such that $\forall\ \varepsilon > 0 \exists \delta > 0$ for which $\|\mathbf{x}-\mathbf{x}_0\| < \delta \Longrightarrow \|f(\mathbf{x}_0)+A(\mathbf{x}-\mathbf{x}_0)-f(\mathbf{x})\| < \varepsilon\|\mathbf{x}-\mathbf{x}_0\|$. This means f is well approximated locally by A and is therefore smooth. The question is what statistical connection can we make to this analysis concept? The problem here is that we need to establish the existence of a linear operator and estimate how well it maps points locally.

We have chosen to use the well-known statistic of multivariate correlation and use a null hypothesis which generates a probability for a confidence statistic analogous to that for the continuity at $\mathbf{x}_0$. We give here a heuristic reason for this choice. See Refs. [44,45] for more details and proofs.

It is known that the method of least squares yields the best estimate of a linear operator that satisfies the equation

$$A\mathbf{a}=\mathbf{b} \tag{3}$$

for a given set of $(\mathbf{a}_i,\mathbf{b}_i)$ pairs of measurements [44] when the data have a normal distribution. For our $(\mathbf{a}_i,\mathbf{b}_i)$ pairs will use the zero-mean variables $\Delta\mathbf{x}_i=(\mathbf{x}_i-\bar{\mathbf{x}})$ and $\Delta\mathbf{y}_i=(\mathbf{y}_i-\bar{\mathbf{y}})$, where $\bar{\mathbf{x}}$ is the mean of the n_δ vectors found for the continuity statistic and $\bar{\mathbf{y}}$ is the mean of their images. The least-squares solution is given by

$$A=\mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\ , \tag{4}$$

where $\mathbf{X}$ is the matrix whose columns are the $\mathbf{x}$ vectors, $\mathbf{X}=(\Delta\mathbf{x}_1,\Delta\mathbf{x}_2,\ldots,\Delta\mathbf{x}_n)$, and $\mathbf{Y}$ is the matrix whose columns are the $\mathbf{y}$ vectors, $\mathbf{Y}=(\Delta\mathbf{y}_1,\Delta\mathbf{y}_2,\ldots,\Delta\mathbf{y}_n)$. For the inverse of A we get the estimate

$$A^{-1}=\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\ . \tag{5}$$

If we have good estimates of A and A^{-1}, we should get $AA^{-1}=1$ (the unit matrix). Putting this together with (4) and (5) we have a correlation statistic r:

$$r^2=\mathrm{tr}[\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}]/d_a\ , \tag{6}$$

where tr means the trace and d_a is the dimension of the attractor (more on this later). This is typically the statistic used in multivariate analysis to estimate the amount of linear correlation between two data sets [44,45]. For example if $r^2=0.6$, then we can say that 60% of the relation between the $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs is a linear one.

Now we can ask how probable this correlation is for our null hypothesis. In our case we want to establish that such a correlation will not likely occur by accident. We take our null hypothesis to be that the $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs are uncorrelated. The correlation for uncorrelated data is 0. We can derive a probability for finding a certain correlation >0 for our null hypothesis. The derivation is straightforward, but tedious. It is based on estimates for the distribution of covariance coefficients [44,47]. We show this derivation in Appendix A. The result is that the approximate probability for finding a correlation r^2 in spaces of d-dimension vector pairs on a d_a-dimensional subspace (the attractor) that are assumed uncorrelated is

$$p \sim e^{-1/2(n_\delta-d_a-1)^2r^2d_a}\ . \tag{7}$$

As before, the process of determining a likelihood or confidence level is a two-step process.

First we define our algorithm for finding points $\mathbf{x}_j$ which satisfy the definition of differentiability for the map (locally at $\mathbf{x}_0$) from X to Y. We choose the ε value, which in this case now determines the error we will allow in the local linear estimate of our function. This ε has a different meaning from that used in continuity. We pick a δ value and find associated $\mathbf{x}_i$ points within δ of $\mathbf{x}_0$. This gives us a set of local $\Delta\mathbf{x}$ and corresponding $\Delta\mathbf{y}$ pairs. We use these to find a least-squares approximation to A, the local linear map. Now we check whether $\|\Delta\mathbf{y}-A\Delta\mathbf{x}\| < \varepsilon_s\Delta\mathbf{x}$ as the definition of differentiability requires. We use a scaled $\varepsilon_s=\varepsilon\sigma_{\Delta\mathbf{y}}/\sigma_{\Delta\mathbf{x}}$, where $\sigma_{\Delta\mathbf{y}}$=standard deviation of the $\Delta\mathbf{y}$ vectors and $\sigma_{\Delta\mathbf{x}}$=standard deviation of the $\Delta\mathbf{x}$ vectors. Using ε_s is necessary to eliminate scale differences between X and Y vectors. If all $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs satisfy the inequality we move on the calculation of the statistic. If not, we choose a smaller δ and try again with a new, smaller set of $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs. We continue this algorithm until the inequality is satisfied or we run out of points.

The second step is to calculate the correlation r^2 and the associated probability p for obtaining r^2 given by Eq. (7) which is determined by our null hypothesis of no correlation between $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs. Our differentiability statistic at the point $\mathbf{x}_j$ becomes $\Theta_{C^1}(\varepsilon,j)=1-p$. If we find no $\mathbf{x}$ points in the δ set, then we choose $\Theta_{C^1}(\varepsilon,j)$ to be zero. The average statistic, as before, becomes

$$\Theta_{C^1}(\varepsilon)=\frac{1}{n_p}\sum_{j=1}^{n_p}\Theta_{C^1}(\varepsilon,j)\ . \tag{8}$$

Note that we could also take into account the actual value of the correlation since it, too, has a value between 0 and 1 and define $\Theta_{C^1}(\varepsilon,j)=r^2e^{-1/2(n_\delta-d_a-1)^2r^2d_a}$. For now we keep the above definition, but this serves to point out that there are often many ways to generate a useful statistic in a given situation.

For the differentiability of the inverse mapping we follow the approach for the inverse continuity statistic and use the above algorithm to calculate r^2 for that situation where the roles of X and Y and δ and ε are reversed. It is important to note that the $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs for differentiability and inverse differentiability are *not* necessarily the same. For example, we can have a continuous function which is not injective (e.g., $y=x^2$) and we will find many $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ pairs for the continuity and therefore differentiability statistic, but few or no pairs for the

associated injectivity and therefore inverse differentiability statistic. Hence, we must calculate r^2 *separately* for each case. As before we switch the roles of X and Y and δ and ε to define an inverse differentiability statistic $\Theta_{I^1}(\varepsilon,j)=1-e^{-1/2(n_\varepsilon-d_a-1)^2r^2d_a}$, so that when $\Theta_{I^1}(\varepsilon,j)\approx 1$ we are confident that we have a differentiable function at the jth point. We further define an average statistic for the entire attractor as

$$\Theta_{I^1}(\varepsilon)=\frac{1}{n_p}\sum_{j=1}^{n_p}\Theta_{I^1}(\varepsilon,j)\ . \tag{9}$$

We remark here that what we have mostly done is to take standard statistical approaches and apply them locally to the time-series trajectory reconstructions. Instead of applying statistical measures to the entire set of data we apply them only to a subset which we choose using geometric or analysis considerations.

E. Rank

In the above we have not made clear what value one uses for d_a, the dimension. The choice of d_a = the dimension of X and Y is not usually correct. To see this consider that we can embed an attractor that exists on a two-dimensional manifold in a five-dimensional space, for example, the Henon system, which we use below. Then the dimension of the derivative correlation matrices, locally, should be 2, not 5. Also, we would have to use the singular-value decomposition (SVD) of the matrices $\mathbf{X}\mathbf{X}^T$ and $\mathbf{Y}\mathbf{Y}^T$ to determine their inverses.

This observation brings up a crucial issue. If one knows the dimension ahead of time, then one can choose d_a and use SVD to cut off the less relevant dimensions in the correlation calculation. Even if one does not know d_a, care must be taken in calculating quantities like $(\mathbf{X}\mathbf{X}^T)^{-1}$. We should still use SVD and pick a cutoff appropriate to our time-series data, for example, the noise level. The significant directions in the phase space have a dimension which we assign to d_a. This type of situation has also been studied in the context of noise reduction and attractor reconstruction [48–52].

Note that for large values for ε we will include large portions of the attractor which may be curved or folded in our ε set causing d_a to equal the dimension of Y. As we decrease ε we find that d_a decreases and levels off. Only at the smallest sets where all the statistics decrease rapidly does d_a also drop to 1 (or zero when no points are found). We use the upper and mid-range d_a's in our calculation of differentiability statistics.

Cawley and Hsu have examined this problem in other contexts [49]. More research needs to be done on finding the best estimate of the embedded manifold dimension. Since this is a separate manifold property one could, perhaps, design a null hypothesis to estimate the probability of obtaining certain values of d_a given the $\mathbf{x}$ and $\mathbf{y}$ pairs or, equivalently, the matrices $\mathbf{X}\mathbf{X}^T$ and $\mathbf{Y}\mathbf{Y}^T$. We have not explored this possibility yet.

IV. APPLICATIONS

Below we show the results of our approach to some test situations: filtering of Henon data, synchronization in chaotic, coupled systems, and tests for determinism. We show typical results here. More detailed conclusions on each set of applications will be published elsewhere. Throughout we emphasize mostly the continuity and inverse continuity statistics [$\Theta_{C^0}(\varepsilon)$ and $\Theta_{I^0}(\varepsilon)$], although we do display some results for the differentiability statistics. In all our tests and applications we scale ε and δ to the standard deviation of the attractor which serves to give a good feeling for the relative ε and δ set sizes (which are all that matter) when examining graphs of the statistics. By standard deviation we mean the estimate computed from the second moment of the distances from the mean of the reconstructed vectors.

A. Test results

The very process of reconstructing a time-series trajectory is itself something that can be studied by the present approach. We examined the reconstruction of a two-dimensional (2D) system, the Henon system ($a=1.4$, $b=0.3$), in a 5D space using the first coordinate of the Henon map as the time-series variable. Hence, we are examining the mapping f: $(x_i,y_i)\rightarrow(x_i,x_{i+1},\ldots,x_{i+4})$ which is a direct experimental test of these statistics *vis à vis* Taken's embedding theorem. In this case we know the manifold dimension $d=2$ and we choose this for our differentiability tests. We take 200 random $\mathbf{x}_j$ points on the attractors to calculate our average statistics. We worked solely in single-precision computer calculations.

Figure 2 shows the Henon attractor along with several ε sets of various sizes. We see that set sizes 0.05 and

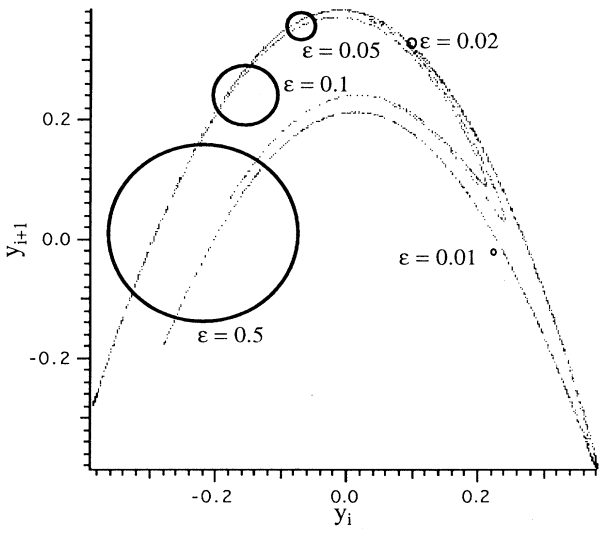

FIG. 2. Two-dimensional time-delay plot of the Henon x time series with various ε set sizes shown.

below should capture the local features of the attractor quite well.

Figure 3(a) shows the continuity statistic for time series of various lengths. As we expect, the continuity drops when we go to small enough ε values. This results from a simple lack of points in the small sets, so the null hypothesis cannot be rejected for many random centers around the attractor. However, when more points are added to the time series we see that the confidence in continuity goes up at the smaller ε set sizes. This is a hallmark of these statistics. When the mapping is truly continuous, injective, or whatever, the respective statistic continually improves, approaching 1.0 asymptotically. Only as we approach the single-precision level (approximately $10^{-6}\varepsilon$ set sizes) might we experience a saturation. In an experiment we would expect saturation at the noise level as tests below bear out.

Figure 3(b) shows the differentiability statistic for various length time series. The inverse continuity and inverse differentiability statistics follow very similar patterns. If these data were from an experiment, we would conclude that we have, with high confidence at high resolution, a true embedding.

As a check on the differentiability statistics we examined a 3D linear transformation on the Lorenz system. A Lorenz system with parameters $\sigma=10$, $b=8/3$, and $\rho=60$ was integrated for 64 000 points using a fourth-order Runge-Kutta routine with a 0.02 time step. A linear transformation

$$A=\begin{bmatrix} 1.0 & 0.1 & 0.01 \\ 0.1 & 1.0 & 0.1 \\ 0.01 & 0.1 & 1.0 \end{bmatrix} \tag{10}$$

was applied to the full time series of the Lorenz system $(x_i,y_i,z_i)^T$, $i=1$ to 64 000 to generate a new 3D time

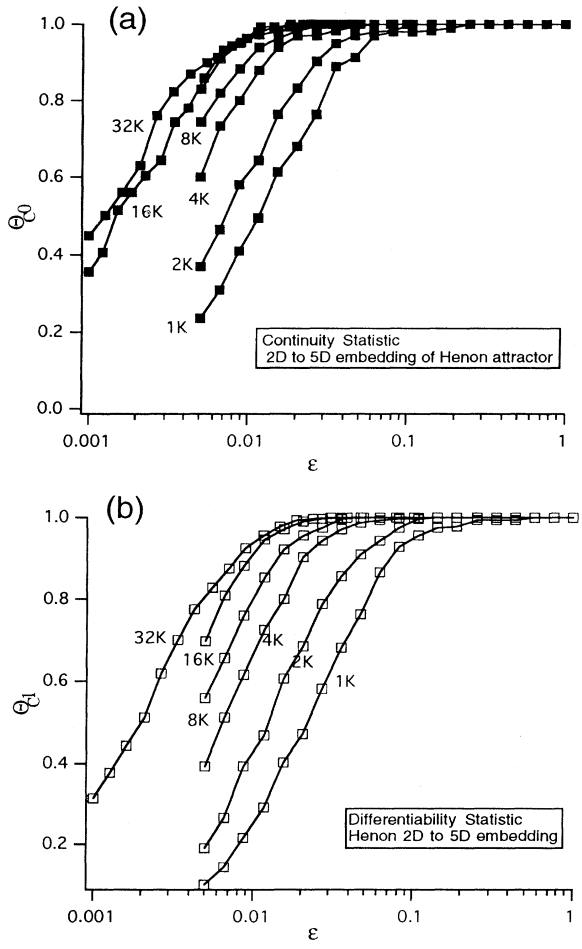

FIG. 3. (a) The continuity statistic for the 5D reconstruction of a Henon map, (b) the injectivity statistic for the same reconstruction. Individual curves are labeled by the number of points in the time series used to generate them.

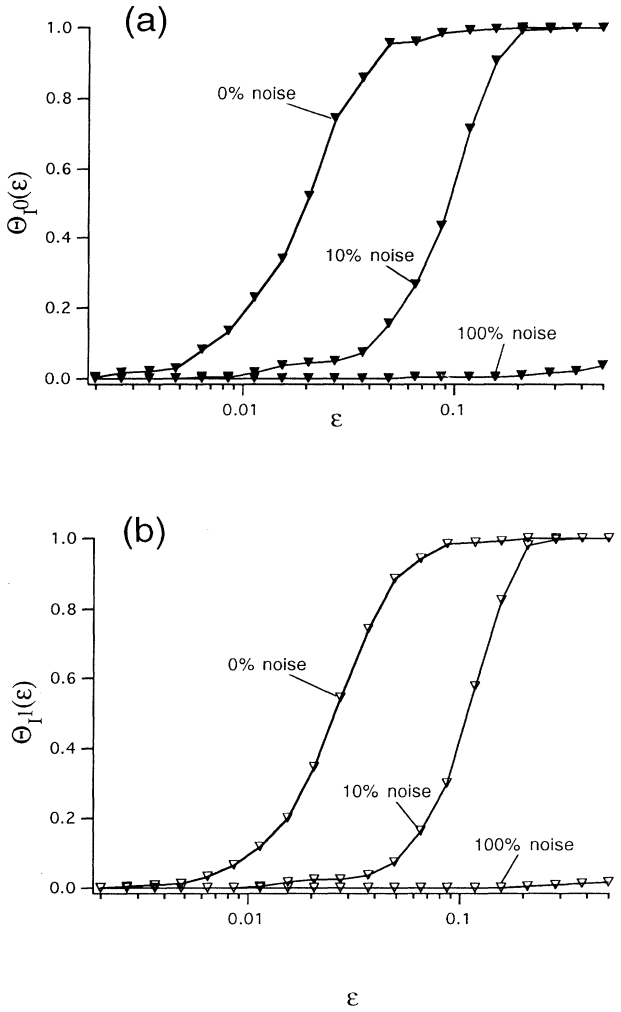

FIG. 4. (a) Injectivity for Lorenz linear tests of clean, 10%, and 100% noise, (b) inverse differentiability for the same.

series. For both time series the dimension is equal to 3. Both continuity and inverse continuity gave high confidence levels to small ε set sizes (e.g., 0.01 for 16 000 time-series points), but more importantly, the correlations were ≈ 1.0 as were the differentiability statistics down to that resolution, too. The latter result is expected for linearly correlated data [44,45]. Figure 4(a) shows the inverse continuity statistic and Fig. 4(b) shows the inverse differentiability statistic for the mapping between the original Lorenz series and the linearly transformed one.

We then added noise from a random number algorithm [53] to the linearly transformed Lorenz data. Figure 4(a) shows the injectivity and Fig. 4(b) shows the inverse differentiability statistics for these data sets as compared to the clean set mentioned above. The effect of 10% noise is the cause of the "cutoff" in the statistics at $\varepsilon = 10\%$ of the standard deviation. Adding more points to the time series does not cause this cutoff to move to smaller ε values, unlike in the clean data. The continuity and differentiability statistics follow the same patterns. A noise level of 100% removes all confidence in all the statistics. We expect this since this situation is almost ex-

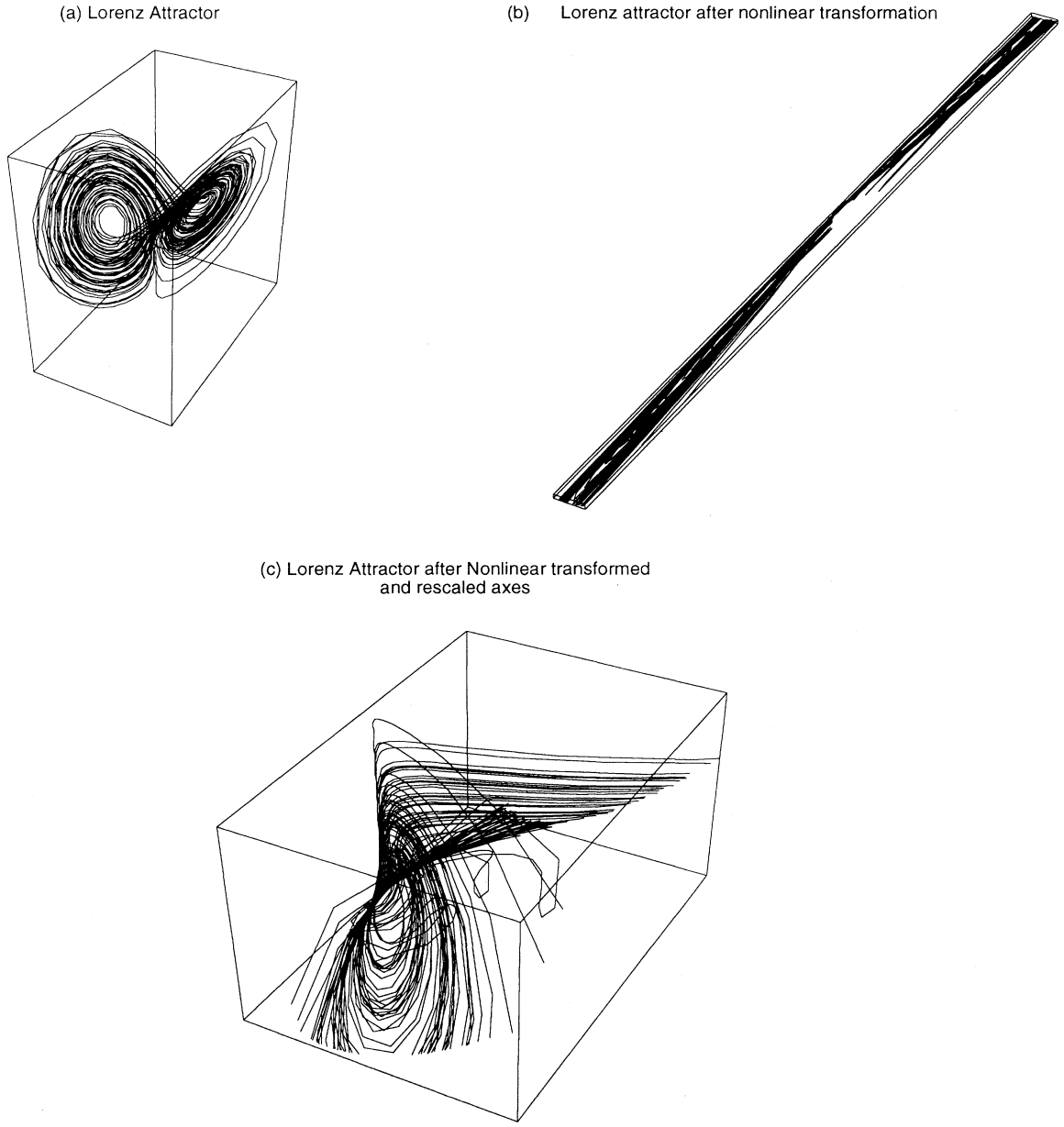

FIG. 5. (a) Original Lorenz attractor with 0.05 ε set, (b) nonlinearly transformed Lorenz attractor, (c) scaled, nonlinearly transformed Lorenz attractor with (x,y,z) scaling factors of (20,1,100), respectively.

actly our null hypothesis.

We next transformed the 3D Lorenz data using a nonlinear transformation given by

$$x' = \alpha x + \beta y^2 ,$$
$$y' = \alpha y - \beta yz , \tag{11}$$
$$z' = \alpha z + 100\beta \cos\left[\frac{y}{25}\right] ,$$

with $\alpha = 0.0$ and $\beta = 1.0$. Figure 5(a) shows the original Lorenz attractor, 5(b) shows the unscaled transformed attractor, and 5(c) shows the rescaled transformed attractor. We can see that the transformation (11) severely distorts the trajectory. Figure 6 shows the continuity statistic for this case. The results imply that the continuity of the nonlinear transformation is quite good despite distortions. Note that the statistics do not saturate. They improve as we add points.

Finally, we tested an extreme version of a nonlinear transformation. We took a 1D time series from a Lorenz system, the x component. We squared it and examined the mapping from the 4D reconstruction using the original time series to the 4D reconstruction using the squared values. Since $y = x^2$ is a continuous function, we do see a good continuity statistic (Fig. 7, solid lines with boxes) that improves as we lengthen the time series until for the longer series we can be $>90\%$ confident in continuity down to ε values of ~ 0.1. However, the same function is not injective. Figure 7 (dashed lines with triangles) shows the injectivity which falls quickly for ε values less than 0.5 even for long time series.

We conclude here that the statistics are performing as expected for known mappings. We now use them on applications where they help shed light on the relation between time series generated by methods other than simple functions.

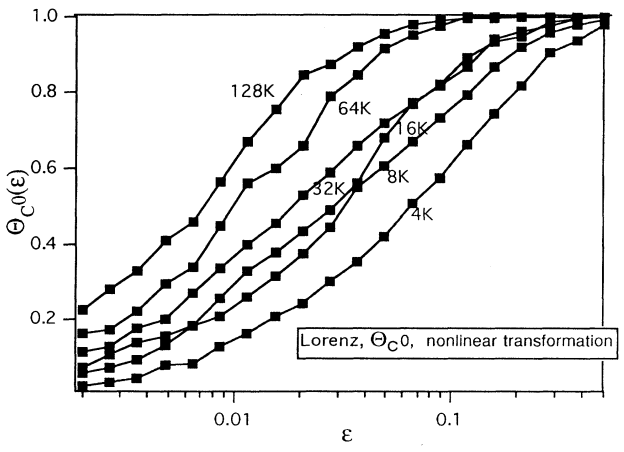

FIG. 6. Continuity statistic for nonlinearity transformed 3D Lorenz time series.

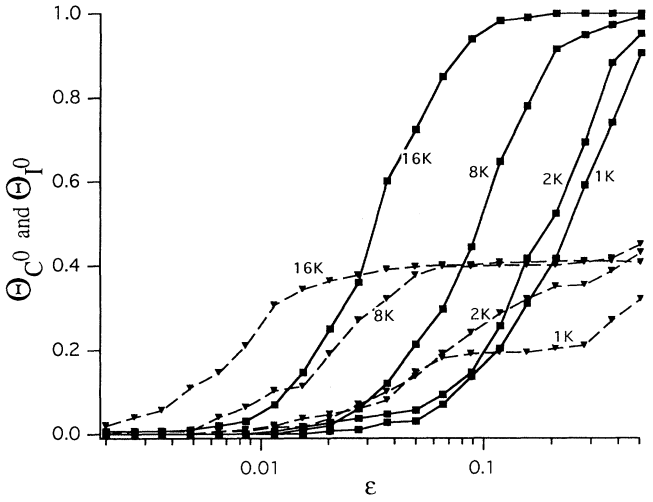

FIG. 7. Continuity (solid lines) and injectivity (dashed lines) statistics for square transformation of Lorenz x time series.

B. Filtering

Filtering time series can be hazardous, as shown by Badii *et al.* [4]. The use of a dynamical filter, like a linear, time-invariant (LTI) filter, can cause the fractal dimension of the resulting reconstruction based on the filtered data to be larger than the actual fractal dimension of the physical system. On the other hand, Mitschke [5] has shown that other filters, like an acausal filter, may have little or no effect on the fractal dimension. If the filter induces a smooth, continuous transformation [6] we know the fractal dimension will be invariant. If not, we cannot be certain of the fractal dimension we calculate from the filtered time series. Below we use our statistical measures to shed light on the differences between the LTI and acausal filters.

The act of filtering a time series with a convolutional filter can be viewed as a mapping between the original and the filtered time series. We can make this explicit.

The vectors in X are constructed by the usual time-delay method:

$$\mathbf{x}(t) = (h(t), h(t+\tau), \ldots, h(t+[d-1]\tau)) \in X ,$$

where $h(t)$ is the time series of measurement. If $g(t)$ is the time series obtained by a filtering of $h(t)$, then the vectors in Y are given by $\mathbf{y}(t) = (g(t), g(t+\tau), \ldots, g(t+[d-1]\tau))$. We write $\mathbf{y}(t) = f(\mathbf{x}(t))$. We show the details of this relationship in Appendix B.

We can now ask the questions as to when f is continuous, injective, differentiable, or diffeomorphic. Broomhead *et al.* [54] have shown that finite-impulse-response (FIR) filters, essentially those with finite extent in time, in principle induce a diffeomorphism between the original resolution and the filtered reconstruction. Presumably, LTI and acausal filters, which have infinite extent, might also induce diffeomorphisms (and thereby guarantee the

preservation of fractal dimension) if the filter falls off fast enough. Just how fast is given by Isabelle *et al.* [55]: exponentially faster than the smallest Lyapunov exponent into the past and faster than minus the largest Lyapunov exponent into the future. However, we introduce a caveat here in that we need to invert the filter to get an inverse mapping, which is an ill-posed problem for an infinite filter [56], and surely will be one for all practical purposes for an FIR which has a long time domain. However, if we are only searching for continuity to preserve fractal dimension, then we may still be in luck, since we do not need the filter inverse for this property. We examine several mathematical properties of filtered time series next.

We started with a time series from the Henon map (parameters $a=1.4$, $b=0.3$). We filtered this series through a set of LTI or acausal filters. Each filter was parametrized by η which defines the action of the filter in the Fourier domain. The LTI filter has a time-domain form $\propto e^{-\eta t}$, which has a Fourier transform given by $i\omega/(\eta^2+\omega^2)$. The LTI filter changes the phases of the Fourier series of the original time series. The acausal filter does not have a simple form in the time domain, but it does have a Fourier domain form given by $1/\sqrt{\eta^2+\omega^2}$. The acausal filter does not change phases of the original time series. Both filters have the same $1/\omega$ amplitude fall off at high frequency. Note that the "amount of filtering" increases for both filters as η decreases.

According to the Kaplan-Yorke conjecture [57], as implemented for LTI filters by Badii *et al.* [4], we calculate that the dimension of the LTI filtered data should begin increasing as η falls below ~ 1.58. For the acausal filter there is, at present, no corresponding estimate. Figure 8 shows the results of the fractal dimension for the attractors reconstructed from the filtered time series. The dimensions were determined using the method of [58] for a data reconstruction in five dimensions. We see that both filters increase the dimension, but the LTI filter causes the largest increase at any particular η value.

Figure 9 shows the continuity of the mapping from the original 5D reconstruction to the reconstruction of the filtered data as a function of η at a typical resolution (ε value). We chose an ε value of 0.04 in the filtered reconstruction where the standard fractal dimension plots are linear. We see immediately that we begin to lose continuity in both LTI and acausal cases near $\eta=1.1$. However, the LTI filtered reconstruction consistently has a lower continuity statistic (Θ_{C^0}) than the acausal data. This immediately sheds some light on why acausal filters appear to be "safer" for filtering chaotic data than LTI filters [5]—they preserve the continuous differentiability property better than LTI filters.

If we look further at the continuity statistic versus ε in Fig. 10 for two η values, one ($\eta=1.2$) just below the 1.58 threshold and one ($\eta=0.4$) at a point where both data sets are highly filtered, we see some interesting results. First, we note that $\Theta_{C^0}(\varepsilon)$ for $\eta=1.2$ improves as we add points to the time series as shown in Fig. 10(a). This indicates that, as we might suspect, the discontinuities caused by either filter are still small. However, as shown in Fig. 10(b) for the LTI filter at $\eta=0.4$, $\Theta_{C^0}(\varepsilon)$ saturates at values of low statistical confidence as more points are added to the series. This implies that the mapping caused by the LTI filter is highly discontinuous. Although this does not imply that the fractal dimension *will* change, we should not be surprised when that happens. The values of $\Theta_{C^0}(\varepsilon)$ for $\eta=0.4$ for the acausal filter do not saturate so dramatically, but clearly do not improve to high confidence levels as they do for larger η values or for the known continuous examples in the preceding section. Hence, the fractal dimensions calculated for these time series are also suspect. This is in line with the smaller change in fractal dimension for the acausal filter at this η value.

It is also interesting to examine the inverse continuity statistic $\Theta_{I^0}(\varepsilon)$ for these filters. Figure 11 shows this

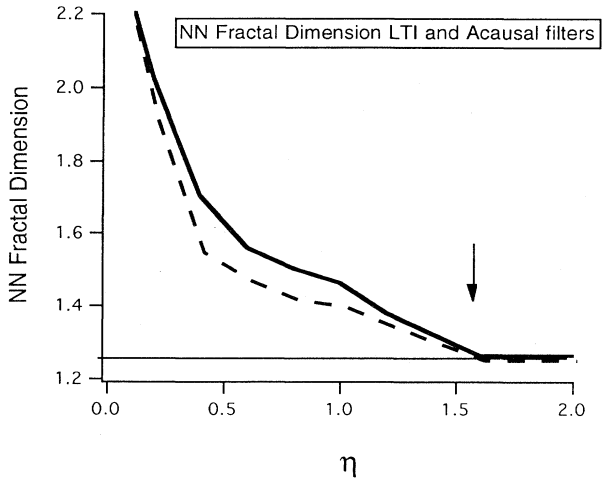

FIG. 8. Dimensions of attractor reconstructed from filtered Henon time series for LTI and acausal filter.

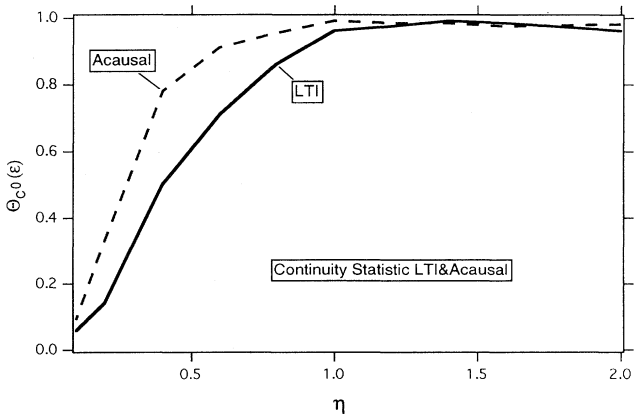

FIG. 9. Continuity statistic at $\varepsilon=0.04$ vs η for LTI and acausal filtered Henon data.

statistic at $\varepsilon=0.04$ for LTI and acausal data as a function of η. This case is just the reverse of the continuity case in that the acausal filter is consistently less injective than the LTI filter. This might affect any properties that depend on injectivity or continuity of the inverse function. However, it is not relevant for fractal dimension since only continuity matters [6].

Generally, for these tests the differentiability statistics follow the trends in the continuity and injectivity statistics, so we can conclude that the LTI also leads to time series in which, for small enough η, we cannot be confident that the mapping from the original series to the filtered series is C^1. We will cover more details on the embedding of filtered data and our statistics elsewhere [59].

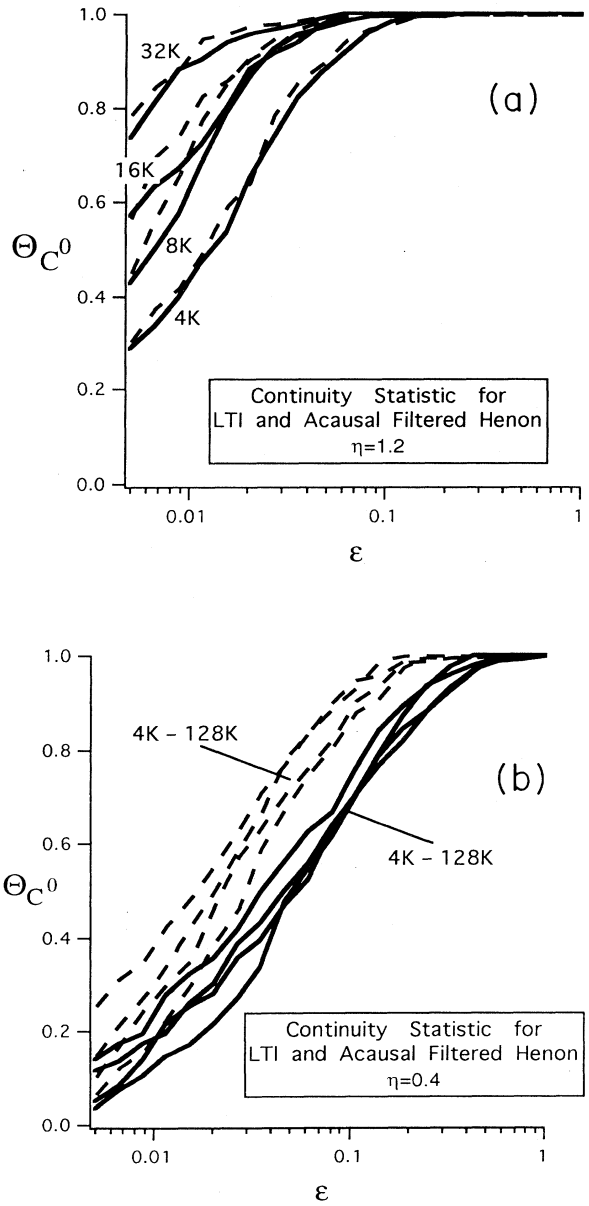

FIG. 10. Continuity statistic vs ε for several time-series lengths (4000 to 128 000 points) for LTI and acausal filtered Henon data, (a) for slight filtering ($\eta=1.2$), (b) for heavy filtering ($\eta=0.4$).

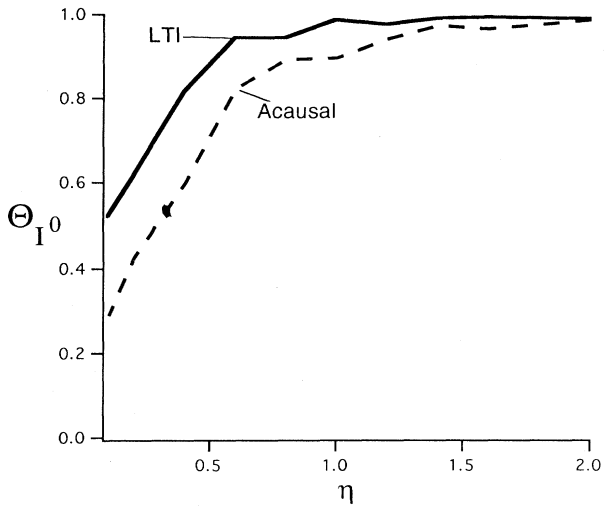

FIG. 11. Injectivity statistic at $\varepsilon=0.04$ vs η for LTI and acausal filtered Henon data.

C. Synchronization

A great deal of interest has been generated from the discoveries of various scenarios in which chaotic systems can synchronize. These scenarios include mutually coupled systems [8,10,13,14,18,19,60–62], one-way driving of systems [9,11,20,22,60,63–66], and control schemes to synchronize chaos [67,68]. In almost all of these studies synchronization is assumed to be the situation in which the variables of one system equal, exactly, those in another.

However, Afraimovich *et al.* [8] considered a more general situation in which the variables in one system are in a smooth, one-to-one correspondence with those in another system. This is another way of saying that there is a diffeomorphism between the two systems [8]. This might occur when we have two chaotic systems coupled strongly enough to cause synchronization, but whose parameters differ slightly. Exact synchronization is not possible [8,21,22], but there may still be a unique relation between the phase-space positions of each system. This situation is usually labeled as *generalized synchronization* [41].

Since generalized synchronization would be expected in any experiment (we can never get two dynamical systems with exactly equal parameters), we would like a way to determine when two systems display such behavior. This usually means we want to compare two times series and conclude whether or not they are synchronized in a general way.

Recently, a method based on the idea of false, nearest neighbors [27,69] was proposed. It concerns the search for mutual, false, nearest neighbors. This test is performed using two time series, just like our X and Y sets in this paper. One calculates the quantity $P(d)$ which is a quotient as explained in the Introduction. If the systems are in general synchronization one would expect that nearest neighbors would be mapped into nearest neighbors and that the ratio $P(d)\approx 1$. This is a heuristic restatement of the Afraimovich diffeomorphism criteria [8]. Rulkov *et al.* show that for some model systems and for a real set of synchronized, chaotic, electronic circuits $P(d)$ does indeed approach 1. For unsynchronized systems $P(d)$ is often much greater than 1, since in that case nearest-neighbor pairs in X usually map into distant neighbors in Y.

From our vantage point we can see that the Rulkov method is a simple version of our differentiability statistic, the local correlation. The basic difference is that our correlation will go to zero if the systems are not synchronized. We can also test for synchronization by using only the continuity and injectivity statistics. In this case we would really be testing for a homeomorphism between the X and Y times series.

The advantage of the Rulkov $P(d)$ quantity is that it is relatively easy to implement and is fast computationally. The advantage of our statistical analysis approach is that we can measure several general synchronization quantities (homeomorphism or diffeomorphism) and associate a confidence level with them. Additionally, since the statistics depend on the level of resolution desired (ε), we can state at what resolution we can still have confidence that both systems are synchronized. Both approaches should be complimentary and useful in several situations.

We tested for general synchronization in a pair of Lorenz systems which were dissipatively coupled through their y components:

$$\begin{aligned}
&\dot{x}_1=\sigma_1(y_1-x_1),\quad \dot{x}_2=\sigma_2(y_2-x_2)\ ,\\
&\dot{y}_1=-x_1z_1+r_1x_1-y_1+c(y_2-y_1)\ ,\\
&\dot{y}_2=-x_2z_2+r_2x_2-y_2+c(y_1-y_2)\ ,\\
&\dot{z}_1=x_1y_1-b_1z_1,\quad \dot{z}_2=x_2y_2-b_2z_2\ ,
\end{aligned}\tag{12}$$

and which were integrated with a fourth-order Runge-Kutta algorithm with a time step of 0.02. A time-series pair was generated for each different coupling and $\{\sigma,b,r\}$ parameter set by taking the x component from each system every seven integration steps (time step equal to 0.14). Another time-series pair was generated by taking the z component from each system at time intervals of 0.14. For both pairs 16 000 data points were used in the analysis.

In order to give some feel for the size of the ε sets used in the analysis, Fig. 12 shows the attractor in two dimensions and several typical ε sizes. Note that sizes below about 0.1 standard deviations are small and cover only local areas of the attractor. Although our reconstructions are in five dimensions to guarantee full unfolding of the attractor [69], the ε sets should be similarly sized.

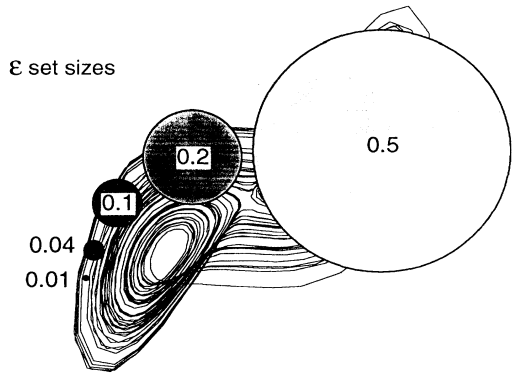

FIG. 12. Plot of the Lorenz attractor with ε sets of typical sizes in fractions of the standard deviation of the data is shown.

For the x component sets we integrated Eqs. (12) for (i) a fully synchronized identical system case $\{\sigma_i,b_i,r_i\}=\{10,8/3,60\}$ for $i=1,2$ and $c=60$, (ii) partially synchronized systems with slightly different parameters $\{\sigma_1,b_1,r_1\}=\{10,8/3,62\}$ and $\{\sigma_2,b_2,r_2\}=\{10,8/3,60\}$ near the threshold of synchronization $c=2.8$ where intermittent behavior is seen—the systems synchronize, then burst out of synchrony occasionally [70,71], (iii) strongly coupled systems $c=60$ with a large parameter difference $\{\sigma_1,b_1,r_1\}=\{10,8/3,136\}$ and

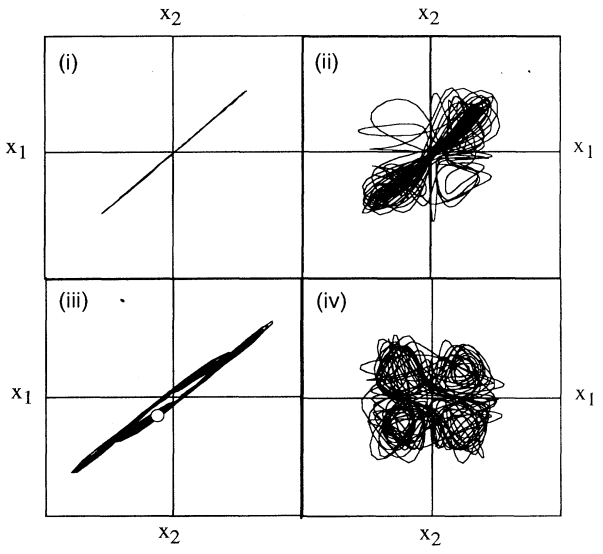

FIG. 13. Phase plots of x components from two Lorenz systems with various couplings and parameter settings (see text for exact values). (i) Highly coupled, matching parameters, (ii) weakly coupled, near sync threshold, slight parameter mismatch, (iii) strongly coupled with large parameter mismatch, and (iv) uncoupled, matching parameters. In (c) the circular hole shows an ε set of size 0.06.

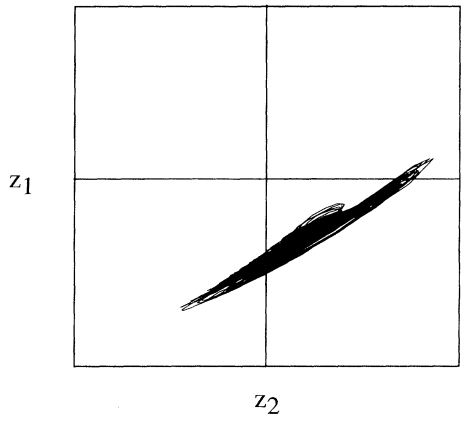

FIG. 14. similar to Fig. 13, except for the z component. The plot is for case (iii) in the text which is also (iii) in Fig. 13.

$\{\sigma_2, b_2, r_2\} = \{10, 8/3, 60\}$, and (iv) completely unsynchronized systems $\{\sigma_1, b_1, r_1\} = \{10, 8/3, 62\}$ and $\{\sigma_2, b_2, r_2\} = \{10, 8/3, 60\}$ and $c = 0$. For the z component we duplicated run (iii) above.

Figure 13 shows the behavior of the x component of the systems when plotted against each other for runs (i) to (iv). Figure 14 shows the behavior of the z components for run (iii). Note that the z component figure is quite different from the corresponding x component figure.

Figure 15 shows the continuity statistic for the four x component cases (i) to (iv). The other statistics follow these trends very closely, so we will only focus on this one. The first case (i) may seem trivial, but it serves to define limits on the statistics. Note that $\Theta_{C^0}(\varepsilon)$ eventually falls off sharply below an ε of about 0.03. We saw this effect in the above tests as a result of having a finite number of points in the time series. In other words, given the number of points (and in an experiment, a noise level) this curve is the best we could achieve in putting a confidence level on whether the two x time series are related by a continuous function. Similar statements apply to the other statistics. In this case ε sizes below 0.1 are quite small on this attractor and we can be confident that we are looking at local effects. Finally, and most importantly, in this case we know the system in great detail, but in an experiment, where we rarely known the system in any detail, it would indeed be correct to say that we can only be confident about the continuity and subsequently the synchronization for ε sets above ~ 0.03.

For case (ii), the intermittency causes the systems to lose synchrony occasionally. In fact for ε sets with good statistics ($\varepsilon > 0.1$) we can say that the system is in sync only about 70–80 % of the time. This might be a good way to measure asynchronous intermittent bursts since the statistics will not discount points off the diagonal like a simple test might, so long as the systems are truly synchronized. In many experiments, getting data strictly on the diagonal is not possible, yet the systems may still be in sync; these statistics will detect that case.

Case (iii) is the most interesting. Despite the distortion caused by the large difference in the r parameter between the systems we see that the systems are still in sync to a good confidence level ($\sim 80\%$ at $\varepsilon \approx 0.06$). This corresponds to a small ε set and is shown in Fig. 13(c). At this resolution we are only sampling points completely off the diagonal.

Case (iii) for the z component is shown in Fig. 16. We

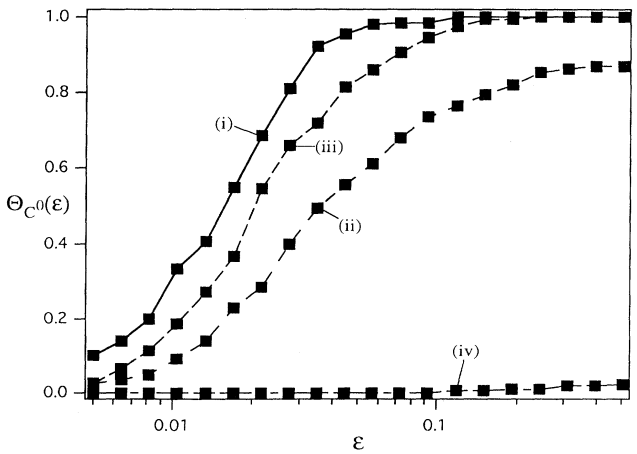

FIG. 15. Continuity statistic $\Theta_{C^0}(\varepsilon)$ for general synchronization tests. (i) Exactly synchronized case, (ii) partially synchronized case, (iii) general synchronization case, (iv) completely unsynchronized case.

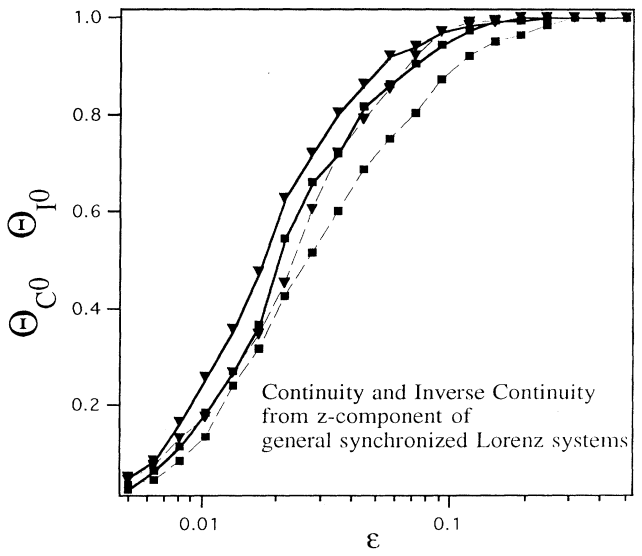

FIG. 16. Continuity statistic $\Theta_{C^0}(\varepsilon)$ (squares) and inverse continuity statistic $\Theta_{I^0}(\varepsilon)$ (triangles) and for general synchronization tests in the Lorenz system using z component time series for 14 000 points (dashed lines) and 30 000 points (solid lines), case (iii) system in text.

show $\Theta_{C^0}(\varepsilon)$ and $\Theta_{I^0}(\varepsilon)$ for 14 000 points as in the x component cases and for 30 000 points. Increasing the number of points increases the confidence levels of the statistic. This suggests that true synchronization is taking place down to $\varepsilon \approx 0.05$ at 80% confidence.

Case (iv) shown in Fig. 15 shows what we expect from our previous tests. The phase-space points of two uncoupled chaotic systems are randomly related to each other.

We applied the same tests to synchronized and unsynchronized circuits. The details of the circuits are given in [7,10]. The circuits consist of a drive and a response, which are modeled by [7]

$$\begin{aligned} \frac{dx}{dt} &= -\alpha(\Gamma x + \beta y + \lambda z) , \\ \frac{dy}{dt} &= -\alpha(-x - \gamma y + 0.02y) , \\ \frac{dz}{dt} &= -\alpha[-g(x) + z] , \end{aligned} \tag{13}$$

for the drive and

$$\begin{aligned} \frac{dx'}{dt} &= -\alpha(\Gamma x' + \beta y + \lambda z') , \\ \frac{dy'}{dt} &= -\alpha(-x' - \gamma y + 0.02y') , \\ \frac{dz'}{dt} &= -\alpha[-g(x') + z'] , \end{aligned} \tag{14}$$

for the response, where

$$g(x) = \begin{cases} 0 & \text{if } x \le 3 \\ \mu(x-3) & \text{if } x > 3 \end{cases}$$

and $\alpha = 10^4$ s, $\Gamma = 0.05$, $\beta = 0.5$, $\lambda = 1.0$, $\gamma = 0.133$, and $\mu = 15$. The y signal from the drive (13) drives the response (14).

We extracted two time series from the drive-response circuits for two scenarios, one where the circuits were well matched and we expect good synchronization and one where the γ parameter was changed by $\sim 20\%$ to a value of 0.160 and we expect degraded synchronization. The data were sampled at 0.1-ms intervals. We collected 32 000 points. Figure 17 shows the plot of the x variable in the response versus its counterpart in the drive for (a) the matched and (b) the unmatched systems. Conclusions from Fig. 17 would be obvious, that the unmatched system is not as well synchronized. We might even be able to quantify this by measuring the "width" of the diagonal band of points in Fig. 17(b) versus Fig. 17(a). However, the statistics $\Theta_{C^0}(\varepsilon)$ and $\Theta_{I^0}(\varepsilon)$ show an interesting occurrence not discernible from Fig. 17.

We calculated $\Theta_{C^0}(\varepsilon)$ and $\Theta_{I^0}(\varepsilon)$ by reconstructing the

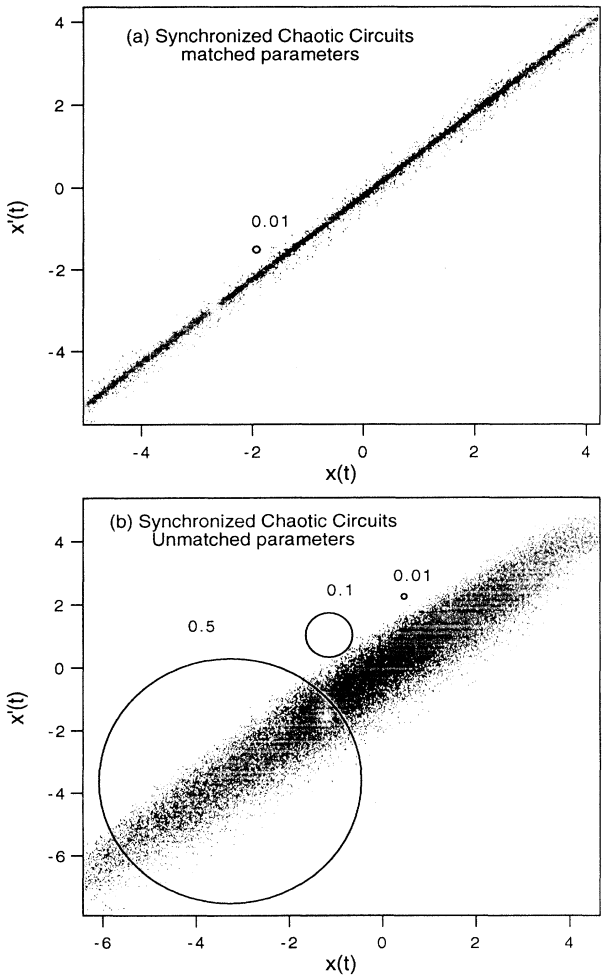

FIG. 17. Phase plots of x voltage from (a) well-matched and (b) poorly matched synchronized, chaotic circuits. Typical ε sets are shown for reference.

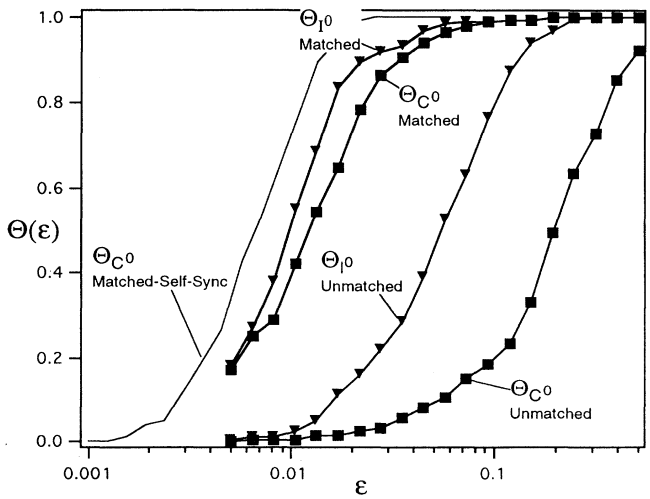

FIG. 18. Continuity and injectivity statistics for well-matched and poorly matched synchronized, chaotic circuits along with the continuity statistic calculated for the drive circuit which is mapped to itself (the identity map). The latter is the best case possible and is used as a comparison for statistics for the other mappings between reconstructions.

two time series (one from the drive and one from the response) each in a six-dimensional space, taking every third point from the time series to avoid in-sample correlations. Figure 18 shows the results for the matched and unmatched data sets as well as the results from a pair of identical data sets (the same drive set used twice). The latter serves to set the limits on the best statistics we can hope for since it represents the identity mapping, which is perfectly continuous and injective. The matched circuit has nearly equal continuity and injectivity statistics, but the confidence levels are lower than for the identical data set case. This is probably due to some inevitable slight mismatch which must occur in any experiment and to a small noise level which is around the 0.003 ε level. The surprise is that the statistics are not the same for the mismatched case. Both confidence levels are lower, but the continuity statistic is at a low level even for large ε sets (see Fig. 17 for a guide to ε sizes). The injectivity statistic does not fall off until well below $\varepsilon=0.1$. We conclude that the association of points on the drive attractor with points on the response attractor is not unique. That is, points in the same δ neighborhood of the drive are often mapped to widely separated points in the response. However, the reverse situation is that points in δ neighborhoods of the response are almost always mapped to points on the drive within $\varepsilon=0.1$ of each other. Response points are highly correlated with associated drive points, but not vice versa. This is even slightly evident in the values of $\Theta_{C^0}(\varepsilon)$ and $\Theta_{I^0}(\varepsilon)$ for the matched case. It is not at all evident in Fig. 16. Heuristically, we can say that even for a mismatched case the response of this system is still highly "slaved" to the drive.

D. Determinism

For our last test of analysis statistics we examined the problem of detecting deterministic behavior, both forward and backward in time, for a Lorenz series generated as in the general synchronization tests from the x component. We integrated the data the same way as for the general synchronization study and sampled the data every seven integration steps (sample interval=0.14) for a 16 000 point time series.

Forward determinism can be measured by $\Theta_{C^0}(\varepsilon)$ for simple determinism and by $\Theta_{C^1}(\varepsilon)$ for smooth determinism. Similarly, backward (reversible) determinism can be measured by either $\Theta_{I^0}(\varepsilon)$ or $\Theta_{I^1}(\varepsilon)$. We chose 1, 2, 4, 8, 16, 32, and 64 points into the future and calculated all four statistics. Figure 19 shows the relationship of the points on the attractor to the starting point. The point size in the figure is scaled to the ε set size (0.0719) used to calculate the statistic. Points far into the future (numbers 16 and 32) are connected to point 0 (the starting point) by trajectories that wind several times around the attractor.

Figure 20 shows the continuity and injectivity statistic along with their differentiability counterparts. For relatively short times (up to the fourth sample equal to 16 sample time steps ≤ 0.56) we can be 90% confident that the x times series represents a deterministic process. By the 16th step the system has gone around the attractor

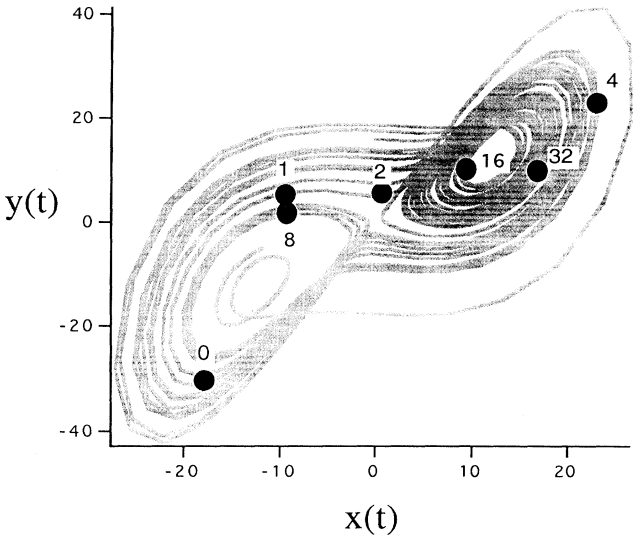

FIG. 19. Points on the Lorenz attractor for which determinism statistics were calculated. The point size is representative of the ε set size used which was 0.0719. Points 16 and 32 are reached by traversing several circulations of the attractor.

approximately three times. The maximum Lyapunov exponent for this system is ~ 1.2 so that the ε set will be expanded approximately 2.3 times in one direction. This is apparently enough to cause all the points in the original set at time step 0 to fall outside of the ε set at the 16th point. Adding more points to the time series will cause the statistic's fall off to move further into the future.

The inverse continuity statistic falls off much faster than the continuity statistic. Presumably this is because the smallest Lyapunov exponent, which dominates the time-reversed behavior, it so negative (or positive for reverse time). For these parameters the smallest exponent

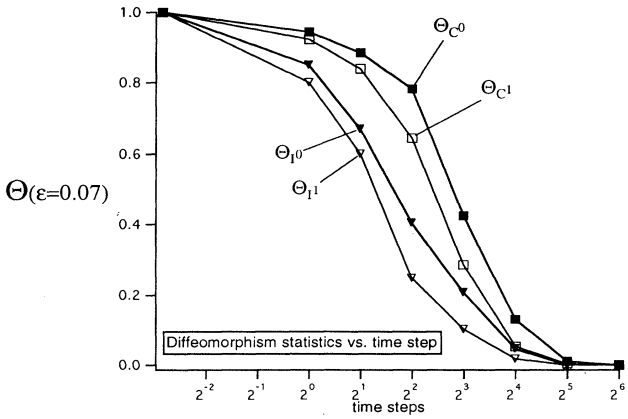

FIG. 20. The statistics for continuity, injectivity, and their differentiability counterparts for the x component time series from the Lorenz system shown in Fig. 18.

≈ -22. hence, we can have a 90% confidence level that the dynamics are invertible only to the first point (0.14 in time from the 0th point). Again, adding more data to the time series will cause this fall off to move to larger time steps, so we would conclude that we are examining an invertible dynamical process.

Finally, both differentiability statistics follow their counterparts closely. We conclude that the dynamics are smooth both forward and backward in time.

An interesting observation from all this is that the fall offs in all statistics are sigmoidal with a roughly exponential behavior at intermediate times. The fall off in predictability for a time series from a chaotic system is also signoidal [72] with roughly exponential behavior at intermediate times. For the statistics the fall off is caused by the requirement of going to smaller and smaller δ sets which eventually have no points at all mapped to the (forward-time) ε set. The positive Lyapunov exponent causes this "loss" of points as a function of forward time. The same loss of points will cause poor statistics for predictability. We conjecture that fall off of statistics and predictability are related, although more work needs to be done on this topic.

V. CONCLUSION

By adhering to the original mathematical definitions we have shown that we can devise reliable statistics which can be applied to pairs of time series from experiments or numerical studies to determine their relation to each other. These are most useful in situations where the relationship between the series is unknown or intractable (e.g., the acausal filter). There are most likely several approaches to devising such statistics, limited only by one's ingenuity in devising null hypotheses or in combining them in compound forms, e.g., to get a continuously differentiability statistic $=\Theta_{C^0}(\varepsilon)\Theta_{C^1}(\varepsilon)$. Our choices here were offered for simplicity and generality. Specific choices for particular circumstances (e.g., colored noise [40]) might be more appropriate.

The advantage of deriving such statistics is that they apply for fundamental analysis concepts and hence are very general. This is evident in the variety of tests we used: filtering, determinism, and synchronization. Many data-analysis questions are often restatements of more fundamental mathematical questions. Deriving statistics for the fundamental concepts yields tools that have broad application. We have only touched on a few here and we will present more detailed results elsewhere.

With regard to the continuity statistic we note that there may exist mathematical functions which are discontinuous, say on a set of measure zero, but whose X and Y samplings will, with probability 1, miss those discontinuities hence yielding good continuity statistics. We do not want to enter here into a philosophical argument over whether such functions are or are not truly observable *in toto,* but we note that from a practical standpoint it is unlikely that such discontinuities would affect real measurements and data analysis.

One application that is not necessarily time-series related that we have not mentioned is the comparison of two- and three-dimensional images. This has important medical application. Images of "text-book anatomy" can be kept in a computer database and used for comparison with images from a patient. The problem is that no two people are alike and the patient's images must be contorted, usually by diffeomorphic transformations to match those of the database. This system is being developed by Greander and Miller [73,74]. One question that comes up is when is the transformation of the patient's images a *reliable* diffeomorphism and when is it not. If it is not, it may indicate a pathology and so the reliability measure becomes a diagnostic tool [74]. Statistics similar to those suggested here might fill the roll of that diagnostic tool.

Another, more general consideration is that in any multivariate statistical setting one often wants to test whether there is any relation between several measured variables and another independent measurement (for example, does the state of the economy depend on unemployment, the interest rate, and the inflation rate). Presently, this is done with tools like correlations. The continuity statistic we present here is actually a more fundamental test for a relationship. At the very least we are usually interested in whether future measurements of several variables with similar values can predict the same relationship. That is just a colloquial statement of continuity. At this time we have not found evidence of this approach in statistics.

The use of statistics appears to be on the rise in the analysis of time series of dynamical systems and attractors. Simultaneously, the use of geometric or mathematical concepts is also increasing (including extraction of differential-geometric quantities from data [75]). Both fields are necessary for understanding experimental results in dynamical systems. The marriage of classical analysis (calculus, differential topology) and statistics suggests that a new topic perhaps called topological statistics is being born.

ACKNOWLEDGMENTS

We would like to acknowledge the help of Tim Sauer in understanding embeddings, maps that preserve fractal dimensions, and in guidance on topological proofs about homeomorphisms on compact objects, Jim Yorke in understanding the subtleties of embeddings, James Theiler for discussions on filtered data and embeddings, Charles Kincaid for help and guidance in developing statistical approaches to topological and analysis concepts, Mike Miller for insight on diffeomorphisms in 2D and 3D images, Steve Schiff and Robert Dodier for critical readings of the manuscript, and Larry Sirovich for suggesting statistical correlations as a solution to the original problem and thereby stimulating the search for null hypotheses.

APPENDIX A: DERIVATION OF LIKELIHOOD FOR CORRELATION

In deriving a likelihood for continuity we considered the probability of obtaining n_δ points in the ε-sized set from a random mapping. This was rather straightforward in that we simply used the "last term" of the bino-

mial distribution, p^{n_δ}. In obtaining a likelihood for calculating a correlation (or its trace) we are faced with the problem of finding the probability of deviations in second moments. Second moments are what the correlation depends on. This problem has an approximate solution based on the central limit theorem [44,47]. We use this approximate solution to generate a distribution for our null hypothesis of no correlation between the $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ vectors. For simplicity in this section we simply write $\mathbf{x}$ and $\mathbf{y}$ for $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$. There will be no confusion.

First, we transform to canonical coordinates [44,45] $\mathbf{z}^i=(\mathbf{X}\mathbf{X}^T)^{-1/2}\mathbf{x}^i$, for $i=1$ to n_δ and $\mathbf{X}\mathbf{X}^T$ is defined in Sec. III C. Similarly, $\mathbf{w}^i=(\mathbf{Y}\mathbf{Y}^T)^{-1/2}\mathbf{y}^i$. We define a compound data set by combining the $\mathbf{z}$ and $\mathbf{w}$ d-dimensional vectors into a $2d$ vector $\mathbf{u}=(\mathbf{z},\mathbf{w})^T$. In approximating a probability distribution for the second moments of the compound data set the moments themselves $E[u_iu_j]$ are treated like a set of $d(2d+1)$ independent random variables. In the end we will be interested in the distribution of the $E[z_iw_j]$ variables since only these relate directly back to our statistics. Note that u_i is the ith component of the measured $\mathbf{u}$ vector, not the ith measurement, i.e., we would approximate $E[u_iu_j]$ by

$$E[u_iu_j]=\frac{1}{n_d}\sum_{k=1}^{n_\delta}u_i^ku_j^k\ , \tag{A1}$$

where n_d is the number of degrees of freedom (see below).

Using the central limit theorem Layland [47] has shown the distributions of $E[u_iu_j]$ have a Gaussian form. This means in order to get the approximate distribution we need the second moments of these second moments, i.e., their covariances. Following Refs. [44,47,76] the covariance of the second moments becomes

$$E[E[u_iu_j]E[u_lu_m]]=E[u_iu_ju_lu_m]-E[u_iu_j]E[u_lu_m]\ . \tag{A2}$$

In order to evaluate these moments we note the following. The hypothesis of no correlation between $\mathbf{x}$ and $\mathbf{y}$ pairs and the canonical form of the variables imply that

$$E[z_lw_m]=0\ , \tag{A3}$$

$$E[z_iz_j]=E[w_iw_j]=\frac{\delta_{ij}}{n_\delta-d_a-1}\ , \tag{A4}$$

$$E[z_iz_jz_lw_m]=E[w_iw_jw_lz_m]=0\ , \tag{A5}$$

$$E[z_iz_jw_lw_m]=E[z_iz_j]E[w_iw_j]\ . \tag{A6}$$

We use $n_\delta-d_a-1$, rather than just n_δ in the denominator to determine an estimate of the residue. This is because we are determining d_a parameters (a row of the matrix A) and we really only have $n_\delta-d_a-1$ degrees of freedom [46]. In simpler terms we need to correct for the fact that in a d_a-dimensional space almost any random d_a+1 pairs of vectors will be perfectly correlated since the number of parameters in the fit (the A matrix) equals the number of variables, $\mathbf{x}$ and $\mathbf{y}$. Hence for the first d_a+1 vectors a good correlation *cannot* cause a low probability and thereby a rejection of the null hypothesis.

These relations show that the covariances in (A2) can be written in a block-matrix form of dimensions $d(2d+1)\times d(2d+1)$. Let Γ be the matrix of the covariances of the second moments [Eq. (A2)], then

$$\Gamma=\begin{bmatrix}\Gamma_1 & 0\\ 0 & \Gamma_2\end{bmatrix}\ , \tag{A7}$$

where Γ_1 is the submatrix of covariances for $E[w_iw_j]$ and $E[z_iz_j]$ variables, and Γ_2 is the submatrix of covariances for $E[z_iw_j]$ variables. The "off-diagonal" terms for these two submatrices are zero since they amount to calculating the moments which have three $z(w)$ factors and one $w(z)$ factor which by (A5) will vanish.

The approximate distribution for the second moments becomes [44]

$$e^{-(1/2)\mathbf{v}^T\Gamma^{-1}\mathbf{v}}=e^{-(1/2)\mathbf{v}^T\Gamma_1^{-1}\mathbf{v}}e^{-(1/2)\mathbf{v}^T\Gamma_2^{-1}\mathbf{v}}\ , \tag{A8}$$

where we have written $\mathbf{v}$ as generic for the appropriate vector of variables $E[u_iu_j]$ in each term. Equation (A8) shows that the probability distribution factors because of the block form of Γ.

We are interested only in the second term containing Γ_2. This leaves an exponent which when transformed back to the original $\mathbf{x}$ and $\mathbf{y}$ variables using (A1) becomes our correlation trace statistic for differentiability:

$$\begin{aligned}\mathbf{v}^T\mathbf{v}&=\sum_{i,j=1}^{n_\delta}[(\mathbf{X}\mathbf{X}^T)^{-(1/2)}\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-(1/2)}]_{ij}\\&\quad\times[(\mathbf{Y}\mathbf{Y}^T)^{-(1/2)}\mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-(1/2)}]_{ji}\\&=\mathrm{tr}[(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}\mathbf{X}^T]=r^2d_a\ ,\end{aligned} \tag{A9}$$

where r^2 is the correlation in Eq. (6). Including the factors from Γ^{-1} the final distribution becomes

$$p=e^{-(1/2)(n_\delta-d_a-1)^2r^2d_a}\ . \tag{A10}$$

Note that this falls off rapidly with the number of points in the δ set and with increasing dimension. This is not hard to understand when one considers that we are asking the probability that noncorrelated data series "accidentally" generate a correlation of r^2. If $r^2\sim1.0$, then we are asking that n vectors accidentally arrange themselves to have a high correlation among most pairs of vectors. This requires roughly $(n_\delta-d_a-1)^2$ combinations of d_a components to be highly correlated. This is a much more stringent criteria than we have for the continuity statistic where we only require that points *individually* fall in a certain set, but other than that their deviations from the mean of the set have no correlation with each other.

APPENDIX B: DERIVATION OF FILTER MAPPING

Let $h(t)$ be a measured signal (a function of the physical phase space point). In the following we will work with continuous signals for simplicity, but the scheme fol-

lows through for discrete measurements as are done in experiments. Then the original reconstruction can be written as

$$\mathbf{x}(t)=(h(t),h(t+\tau),\ldots,h(t+[d-1]\tau))\in X\ ,$$

where d is the embedding dimension. We must remember that the flow of the physical system plays a crucial role in writing time-lag reconstructions [1,3] and it does so in the filtering version, too. Let Φ_t be the flow on the physical system represented in its phase space by $\mathbf{s}(t)$. Then we get $\mathbf{s}$ at a later time by application of Φ: $\mathbf{s}(t+\tau)=\Phi_\tau(\mathbf{s}(t))$. Our measurement h is then a function of $\mathbf{s}$ and Φ: $h(t+\tau)=h(\mathbf{s}(t+\tau))=h(\Phi_\tau(\mathbf{s}(t)))$.

If R is the convolutional filter, the filtered signal $g(t)$ is given by

$$\begin{aligned} g(t)&=\int R(t-t')h(t')dt' \\ &=\int R(t-t')h(\Phi_{t'-t}(\mathbf{s}(t))dt' \end{aligned} \tag{B1}$$

and the reconstruction from the filtered time series is

$$\mathbf{y}(t)=(\boldsymbol{g}(t),g(t+\tau),\ldots,g(t+[d-1]\tau))\in Y\ .$$

Now using the property $\Phi_t\Phi_\tau=\Phi_{t+\tau}$ we can write the filtered time-series reconstruction as a map from a point in the physical phase space to a point in the final reconstructed phase space. The intermediate mapping between X and Y is given by

$$\begin{aligned} \mathbf{y}(t)&=f(\mathbf{x}(t))=\int R(t-t')\mathbf{x}(t')dt' \\ &=\int R(t-t')\Psi_{t'-t}(\mathbf{x}(t))dt'\ , \end{aligned} \tag{B2}$$

where

$$\Psi_{t'-t}(\mathbf{x}(t))=(\ldots,h(\Phi_{t'-t+i}(\mathbf{s}(t))),\ldots) \tag{B3}$$

for $i=0,\ldots,d-1$.

Although most of this seems like formal maneuvering, it is necessary to show that we indeed have a point-to-point mapping between all spaces: the physical space, the **X** space, and the **Y** space.

[1] F. Takens, *Detecting Strange Attractors in Turbulence,* in *Dynamical Systems and Turbulence, Warwick, 1980,* edited by D. Rand and L.-S. Young (Springer, Berlin, 1981), p. 366.

[2] R. Mané, *Embeddings,* in *Dynamical Systems and Turbulence,* edited by R. Rand and L. S. Young (Springer, Berlin, 1981), Vol. 898, p. 230.

[3] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **64**, 579 (1991).

[4] R. Badii *et al.*, Phys. Rev. Lett. **60**, 979 (1988).

[5] F. Mitschke, Phys. Rev. A **41**, 1169 (1990).

[6] T. Sauer and J. A. Yorke (unpublished).

[7] T. Carroll, Am. J. Phys. **63**, 377 (1995).

[8] V. S. Afraimovich, N. N. Verichev, and M. I. Rabinovich, Inv. VUZ. Rasiofiz. RPQAEC **29**, 795 (1986).

[9] Thomas L. Carroll and Louis M. Pecora, IEEE Trans. CAS. **38**, 453 (1991).

[10] J. F. Heagy, T. L. Carroll, and L. M. Pecora, Phys. Rev. E **50**, 1874 (1994).

[11] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).

[12] C. Pérez-Villar *et al.*, Int. J. Bifurc. Chaos **3**, 1067 (1993).

[13] A. S. Pikovski, Radiophys. Quantum Electron. **27**, 390 (1984).

[14] N. F. Rul'kov *et al.*, Int. J. Bifurc. Chaos **2**, 669 (1992).

[15] E. E. Shnol, PMM U.S.S.R. **51**, 9 (1987).

[16] W. Singer, Ann. Rev. Physiol. **55**, 349 (1993).

[17] N. N. Verichev and A. G. Maksimov, Sov. Radio Phys. **32**, 713 (1990).

[18] A. R. Volkovskii and N. F. Rul'kov, Sov. Tech. Phys. Lett. **15**, 249 (1989).

[19] H. G. Winful and L. Rahman, Phys. Rev. Lett. **65**, 1575 (1990).

[20] T. L. Carroll and L. M. Pecora, Physica D **67**, 126 (1993).

[21] Louis M. Pecora and Thomas L. Carroll, Phys. Rev. A **44**, 2374 (1991).

[22] N. F. Rulkov and A. R. Volkovskii (unpublished).

[23] S. Bressler, R. Coppola, and R. Nakamura (unpublished).

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

[25] H. G. Schuster *et al.*, in *Measures of Complexity and Chaos,* edited by N. B. Abraham *et al.* (Plenum, New York, 1989), pp. 349–358.

[26] W. Liebert, K. Pawelzik, and H. G. Schuster, Europhys. Lett. **14**, 521 (1991).

[27] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).

[28] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).

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

[30] L. W. Salvino and R. Cawley, *Applications of a Statistical Test for "Smooth" Dynamics in Embedded Time Series,* in *Proceedings of the 2nd Annual Office of Naval Research/Naval Undersea Warfare Center Technical Conference on Nonlinear Dynamics and Full Spectrum Processing, Mystic,* edited by R. A. Katz, AIP Conf. Proc. No. 296 (American Institute of Physics, New York, 1993).

[31] R. Wayland *et al.*, Phys. Rev. Lett. **70**, 580 (1993).

[32] A. M. Albano, A. Passamante, and M. E. Farrell, Physica D **54**, 85 (1991).

[33] W. Liebert and H. G. Schuster, Phys. Lett. A **142**, 107 (1989).

[34] J. M. Martinerie *et al.*, Phys. Rev. A **45**, 7058 (1992).

[35] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, Physica D **73**, 82 (1994).

[36] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[37] J. Theiler *et al.*, *Using Surrogate Data to Detect Nonlinearity in Time Series,* in *Nonlinear Modeling and Forecasting,* edited by M. Casdagli and S. Eubank (Addison-Wesley, Reading MA, 1992), Vol. XII, pp. 163–188.

[38] J. Theiler *et al.*, Physica D **58**, 77 (1992).
[39] For nonstatisticians a null hypothesis can be thought of as a probabilistic version of reductio ad adsurdum. One makes an assumption and arrives at a conclusion which (hopefully) can be shown to be of low probability. This implies that the assumption is of low probability and it can be rejected on statistical grounds.
[40] D. T. Kaplan, Physica D **73**, 38 (1994).
[41] N. Rul'kov *et al.*, Phys. Rev. E **51**, 980 (1995).
[42] J. Theiler, Phys. Rev. A **41**, 3038 (1990).
[43] S. Sternbeg, *Lectures on Differential Geometry* (Chelsea, New York, 1983).
[44] G. A. F. Seber, *Multivariate Observations* (Wiley, New York, 1984).
[45] S. F. Arnold, *The Theory of Linear Models and Multivariate Analysis* (Wiley, New York, 1981).
[46] R. E. Walpole and R. H. Myers, *Probability and Statistics for Engineers and Scientists* (Macmillan, New York, 1972).
[47] M. W. J. Layland, Ann. Math. Stat. **43**, 123 (1972).
[48] D. Broomhead and G. P. King, Physica D **20**, 217 (1986).
[49] R. Cawley and G. Hsu, Phys. Rev. A **46**, 3057 (1992).
[50] E. Kostelich and J. A. Yorke, Phys. Rev. A **38**, 1649 (1988).
[51] T. Sauer, Physica D **58**, 193 (1992).
[52] D. Broomhead and G. P. King, J. Phys. A **20**, 563 (1987).
[53] W. H. Press *et al.*, *Numerical Recipes* (Cambridge University Press, New York, 1990).
[54] D. S. Broomhead, J. P. Huke, and M. R. Muldoon, J. R. Stat. Soc. B **54**, 373 (1992).
[55] S.. H. Isabelle, A. V. Oppenheim, and G. W. Wornell (unpublished).
[56] V. F. Turchin, V. P. Kozlov, and M. S. Malkevich, Sov. Phys. Uspekhi **13**, 681 (1971).
[57] J. L. Kaplan and J. A. Yorke, in *Functional Differential Equations and Approximations of Fixed Points,* edited by H.-O. Peitgen and H.-O. Walther (Springer-Verlag, Berlin, 1979), Vol. 730, p. 228.
[58] Y. Termonia and Z. Alexandrowicz, Phys. Rev. Lett. **51**, 1265 (1983).
[59] L. M. Pecora, T. L. Carroll, and J. F. Heagy, Phys. Rev. E (to be published).
[60] V. S. Anishchenko *et al.*, Int. J. Bifurc. Chaos **2**, 633 (1992).
[61] H. Fujisaka and T. Yamada, Prog. Theor. Phys. **69**, 32 (1983).
[62] J. M. Kowalski, G. L. Albert, and G. W. Gross, Phys. Rev. A **42**, 6260 (1990).
[63] A. R. Volkovskii and N. F. Rul'kov, Sov. Techn. Phys. Lett. **19**, 97 (1993).
[64] Louis M. Pecora and Thomas L. Carroll, *Driving Systems with Chaotic Signals,* in *Proceedings of the First Experimental Chaos Conference,* edited by S. Vohra *et al.* (World Scientific, Singapore, 1992).
[65] K. Cuomo and A. V. Oppenheim, Phys. Rev. Lett. **71**, 65 (1993).
[66] K. M. Cuomo, A. V. Oppenheim, and S. H. Srogatz, IEEE Trans. Circ. Syst. **40**, 626 (1993).
[67] Paul So, Edward Ott, and W. P. Dayawansa, Phys. Lett. A **176**, 421 (1993).
[68] K. Pyragas, Phys. Lett. A **181**, 203 (1993).
[69] M. B. Kennel and H. D. I. Abarbanel, Phys. Rev. E **47**, 3057 (1993).
[70] J. F. Heagy, T. L. Carroll, and L. M. Pecora, Phys. Rev. Lett. **73**, 3528 (1995).
[71] P. Ashwin, J. Buescu, and I. Stewart, Phys. Lett. A **193**, 126 (1994).
[72] Matthew B. Kennel, Henry D. I. Abarbanel, and J. J. Sidorowich (unpublished).
[73] Ulf Grenander and Michael I. Miller, J. R. Stat. Soc. B **56**, 349 (1994).
[74] M. I. Miller (private communication).
[75] M. R. Muldoon *et al.*, Physica D **65**, 1 (1993).
[76] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).

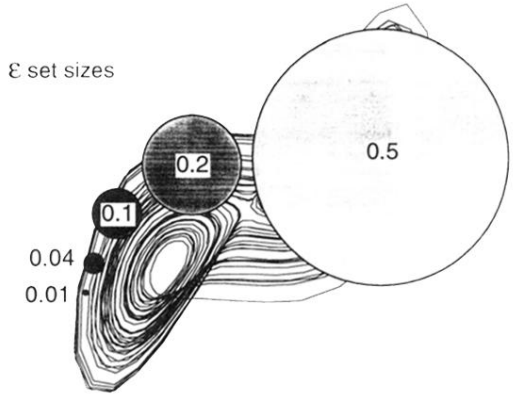

FIG. 12. Plot of the Lorenz attractor with ε sets of typical sizes in fractions of the standard deviation of the data is shown.

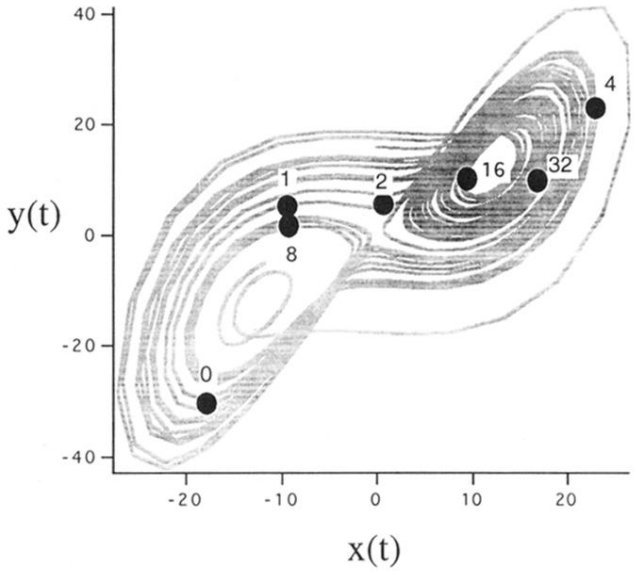

FIG. 19. Points on the Lorenz attractor for which determinism statistics were calculated. The point size is representative of the ε set size used which was 0.0719. Points 16 and 32 are reached by traversing several circulations of the attractor.