Time series and dependent variables

Robert Savit and Matthew Green

Physics Department, The University of Michigan, Ann Arbor, MI 48109, USA

We present a new method for analyzing time series which is designed to extract inherent deterministic dependencies in the series. The method is particularly suited to series with broad-band spectra such as chaotic series with or without noise. We derive quantities, $\delta_j(\varepsilon)$, based on conditional probabilities, whose magnitude, roughly speaking, is an indicator of the extent to which the kth element in the series is a deterministic function of the $(k-j)$th element to within a measurement uncertainty, ε. We apply our method to a number of deterministic time series generated by chaotic processes such as the tent, logistic and Hénon maps, as well as to sequences of quasi-random numbers. In all cases the δ_j correctly indicate the expected dependencies. We also show that the δ_j are robust to the addition of substantial noise in a deterministic process. In addition, we derive a predictability index which is a measure of the extent to which a time series is predictable given some tolerance, ϵ. Finally, we discuss the behavior of the δ_j as ε approaches zero.

1. Introduction

Time series which evidence characteristics of a broad-band spectrum are notoriously difficult to analyze. Traditional methods such as Fourier analysis or other linear transforms usually fail to offer many insights into the underlying structure of such a time series. Broad-band time series appear in many contexts including data signaling processes, biomedical applications, and economic systems and, unfortunately, are more the rule than the exception.

Crudely speaking, the broad-band nature of such series may be due to "noise", to deterministic processes of a chaotic, or near-chaotic nature, or to a combination of both. One important objective in the analysis of broad-band series is distinguishing among these alternatives. In this paper, we present a method for analyzing time series which allows one to distinguish between certain kinds of noise and certain deterministic processes. Our approach is based on the construction of conditional probabilities for the repetition of short sequential patterns of values in a time series. The conditional probabilities can be expressed in terms of the Grassberger–Procaccia correlation integrals [1] and contain information from the entire series. If the series is generated by a chaotic process, then our method samples all regions of the attractor. Using this method, we can determine, quantitatively, the extent to which a term in the time series is a function of previous terms in the series. As a by-product, we are able to determine, in the case of a chaotic system, the minimum embedding dimension necessary for a reasonable description of the dynamics. As we shall make clear below, this minimum embedding dimension generally depends on ε, the tolerance or uncertainty with which one wishes to observe the dynamics. Finally, we introduce a predictabil-

ity index which measures the degree to which a series is predictable (i.e., the extent to which a series is driven by reproducible deterministic dynamics) given values of previous elements in the series. To discuss our methods we will study several low-dimensional maps, both with and without noise.

The rest of this paper is organized as follows: In the next section we shall derive the conditional probabilities of interest in terms of the Grassberger–Procaccia correlation integrals. We shall also review the work of Brock, Dechert, Scheinkman, and collaborators [2, 3], and will show in what sense our work generalizes and extends their results. In section 3 we shall apply our method to several mathematical systems including the tent, logistic, and Hénon maps, as well as to sequences of numbers generated by a good random number generator. We shall show that our method correctly indicates what are the dependent variables in each of these cases. In section 4 we discuss some technical points which arise in systems with converging trajectories. In section 5 we shall briefly discuss the ε-dependence of our results. Section 6 consists of a summary and conclusion.

2. Tools for analyzing time series

Consider a time series of real numbers, $x(n)$. For most of this paper we assume that n is discrete. We will briefly comment on continuous time series in section 6. The sequence $x(n)$ may have a broad-band spectrum and still be generated by a relatively simple underlying deterministic process. Although it is not necessary for the application of our procedure, it may be useful to have in mind, as a paradigm, an underlying iterative map (either with or without noise) as the generator of the time series so that $x(n + 1) = F[\{x(j)\}] + \eta$, where F is some function which depends on some set of $x(j)$, $j < n + 1$, and η is some stationary delta-function correlated noise. Our method gives useful information in more

general contexts also, but this paradigm is helpful for directing the discussion.

2.1. Review of the BDS test

Let us suppose that we have a time series and we want to tell whether it is random, or whether it is generated by some underlying map. In this paper random will mean IID, that the numbers in the series are independent and drawn from a probability distribution which is identical for each number. A statistical test against the null hypothesis that a sequence of numbers is IID (the BDS test) has recently been devised by Brock, Dechert and Scheinkman [2]. Their test works as follows: Following Grassberger and Procaccia [1], one forms d-tuples from the time series which one considers to be vectors in a d-dimensional space:

$$v(i) = (x(i), x(i-1), \ldots, x(i-d+1))$$
$$= (v_1(i), v_2(i), \ldots, v_d(i)). \tag{1}$$

One then defines an indicator function $I_{ij}(\varepsilon)$, which is 1 if each Cartesian component of the vector $v(i)$ is within ε of the corresponding Cartesian component of the vector $v(j)$, and zero otherwise. Thus,

$$I_{ij}(\varepsilon) = \prod_{k=1}^{d} \Theta(\varepsilon - |v_k(i) - v_k(j)|) \tag{2}$$

and

$$C_d(\varepsilon) = \frac{1}{\nu} \sum_{\text{pairs}} I_{ij}(\varepsilon), \tag{3}$$

where ν is the number of pairs of vectors $(v(i), v(j))$. $C_d(\varepsilon)$ just measures the probability that two d-vectors are within ε of each other in all their Cartesian coordinates. The BDS statistic proceeds from the observation that for a sequence of numbers that are IID

$$N_d(\varepsilon) = C_d(\varepsilon) - [C_1(\varepsilon)]^d \cong 0 \tag{4}$$

in a statistical sense, for any d no matter how large. The contribution of Brock, Dechert and Scheinkman was to derive the correct normalization factor $K_d(\varepsilon)$ by which to divide (4) to be able to make a precise statistical statement. They showed that in the limit that the time series gets infinitely long,

$$\xi_d(\varepsilon) = N_d(\varepsilon)/K_d(\varepsilon) \qquad (5)$$

is a statistic normally distributed with a mean of zero and a standard deviation of one for any embedding dimension, d, if the sequence of numbers is IID. (The normalization factor, $K_d(\varepsilon)$, is not particularly illuminating, and so we do not present it here. It is derived in ref. (2).) Thus, (5) constitutes a test against the null hypothesis that the sequence of numbers is IID.

The BDS statistic has been extensively studied and has been shown to have power against many sequences generated by deterministic maps, but which pass standard tests of randomness based on autocorrelation functions. $\xi_d(\varepsilon)$ is a quantity that measures the "clumpiness" (other than that due to a non-flat probability distribution) in d embedding dimensions. Typically, when applied to a low-dimensional deterministic chaotic sequence, $\xi_d(\varepsilon)$ becomes significantly larger than one for values of d in which the attractor begins to show some non-trivial structure, and increases with d for larger values. For example, when applied to the tent map with ε chosen to be one-half the standard deviation of the sequence (about 0.25) $\xi_2(\varepsilon)$ is already about 300, $\xi_3(\varepsilon)$ is about 700, and $\xi_d(\varepsilon)$ continues to grow for larger values of d. For a noisy deterministic system with a higher-dimensional attractor, $\xi_d(\varepsilon)$ will typically be fairly small for low embedding dimensions and will begin to increase in dimensions in which the attractor begins to show some non-trivial structure.

2.2. Indicators for dependent variables

It is clear that the BDS statistic is very useful at ruling out IID behavior for a time series.

However, it is not clear what information is contained in the BDS statistic for larger dimensions. For example, since the tent map is generated by a single variable map, all the information about the tent map is already contained in two embedding dimensions. The fact that $\xi_d(\varepsilon)$ continues to grow for larger d just reflects the fact that the attractor of the tent map is always a one-dimensional structure in any embedding dimension, and as d increases it becomes less likely that a one-dimensional structure could have been produced by an IID process in d dimensions. On the other hand, maps such as the Hénon map in which there is explicit dependence on more than one (in this case two) previous variables also have values of $\xi_d(\varepsilon)$ that grow with d, but in this case there may be new information contained in higher-dimensional embeddings.

To see how to identify cases in which higher-dimensional embeddings provide new information, let us look again at the $C_d(\varepsilon)$'s. Consider the two subsequences of length d represented by two vectors, $v(i)$ and $v(j)$. It is easy to see that $C_d(\varepsilon)$ is just the probability that each member of one of the sequences of length d is within ε of the corresponding member of the other sequence. That is,

$$C_d(\varepsilon) = P(t_d, t_{d-1}, \ldots, t_1), \qquad (6)$$

where t_k stands for the statement

$$|v_{d-k+1}(i) - v_{d-k+1}(j)| \le \varepsilon.$$

Now, suppose we have two numbers from the time series, $x(i)$ and $x(j)$ such that $|x(i) - x(j)| \le \varepsilon$. What is the probability that $|x(i+1) - x(j+1)| \le \varepsilon$? This is just the conditional probability $P(t_2 | t_1)$. But, since $P(t_2, t_1) = P(t_2 | t_1)P(t_1)$, we have

$$P(t_2 | t_1) = C_2/C_1, \qquad (7)$$

where for notational simplicity we have dropped the ε. If the sequence of numbers is IID, then

$P(t_2|t_1) = P(t_2) = C_1$, and we have $C_2 = C_1^2$. Consider

$$\delta_1 = 1 - C_1^2/C_2. \tag{8}$$

If the sequence of numbers is IID, then in a statistical sense δ_1 will be zero. In fact, δ_1 is just proportional to the numerator of the first BDS statistic, $\xi_2(\varepsilon)$.

Although a significant deviation of δ_1 from zero indicates absence of IID behavior, the evaluation of (8) contains more precise information. δ_1 will differ appreciably from zero if $P(t_2|t_1) \neq P(t_2)$. This will be the case, for example, if the time series is generated by a deterministic system with a single lag dependence. If, on the other hand, the time series is generated by a map in which $x(i)$ depends only on terms in the series other than $x(i-1)$, then δ_1 will still be approximately zero.

Now consider possible dependencies of $x(i)$ on $x(i-2)$. Even if the sequence is generated by a map which has only explicit single lag dependence, there will still be an induced dependence of $x(i)$ on $x(i-2)$: If $x(i) = F(x(i-1))$, then $x(i) = F^{[2]}(x(i-2))$. (In fact, it is just this induced dependence that is measured by forming the semi-invariants from the C_d's [4].) We are interested, however, in any dependence of $x(i)$ on $x(i-2)$ in addition to this (trivial) induced dependence. Such additional dependence might be the result of additional explicit dependence on $x(i-2)$ in the fundamental map, or might result from the non-uniform curvature in a simple single lag map, as we shall discuss in the next section.

To form a measure of this additional dependence, note that

$$C_3 = P(t_3, t_2, t_1) = P(t_3|t_2, t_1) P(t_2, t_1) \tag{9}$$

so that

$$P(t_3|t_2, t_1) = C_3/C_2. \tag{10}$$

Now, if $x(i)$ depends intrinsically only on $x(i-1)$,

then $P(t_3|t_2, t_1) = P(t_3|t_2)$. In this case, assuming that the system is time translationally invariant and using (7), (10) becomes

$$C_3/C_2 = P(t_3|t_2, t_1) = P(t_3|t_2)$$
$$= C_2/C_1. \tag{11}$$

Define

$$\delta_2 = 1 - C_2^2/C_1 C_3. \tag{12}$$

From (11), if $P(t_1|t_2, t_1) = P(t_3|t_2)$, then $\delta_2 = 0$. The extent to which $\delta_2 \neq 0$ is a measure of the extent to which $P(t_3|t_2, t_1) \neq P(t_3|t_2)$, and thus is a measure of the extent to which there is intrinsic dependence of $x(i)$ on $x(i-2)$.

This argument can be continued, and in general one can define a quantity

$$\delta_j = 1 - C_j^2/C_{j-1} C_{j+1}, \tag{13}$$

which measures the degree of additional dependence of $x(i)$ on $x(i-j)$, aside from that induced by the explicit dependencies on $x(k)$ for $i - j < k < i$.

Two comments are in order about these results. First, the cognescenti will notice some formal similarity between the expressions (13), and the test for various degrees of Markovian behavior in a time series as developed by Kolmagorov [5]. The difference between our work and that of Kolmagorov is that the δ_j's express dependencies which are the result of averages over different regions of the map. The usual application of the formalism of Kolmagorov is to conditional probabilities for the occurrence of specific values of the variables, $x(n)$, of the time series. In our formalism, on the contrary, the conditional probabilities search the entire series and compare different substrings in the series for similar structure. Thus, the δ_j's provide less specific, average information. On the other hand, since the δ_j's represent global averages, we are able to use information from all regions of the attractor, and so by constructing the δ_j's we use the available information more

efficiently. Thus, our measures of conditional probability embody an analysis of the global properties of time series. Second, it should be clear that the δ_j's will be most effective at indicating dependencies which are generated by maps that are in some sense smooth. Our conditional probabilities measure the probability of continued closeness of sub-sequences of numbers in the time series. Such probabilities will pick up dependencies generated by smooth maps in which nearby points map, over short times, into points which are also relatively nearby. But these conditional probabilities may miss effects of maps which are sufficiently nonsmooth. This point will be discussed further in the next section.

2.3. Interpretation of the δ_j's and the predictability index

Suppose that we have computed a set of δ_j's for some time series. Clearly, the larger a given δ_j the greater the predictability of the time series. We now derive a measure of the additional predictability of the time series in terms of the δ_j's. First, let us be precise about what we mean by predictability. Imagine that we have a long time series $x(n), 1 \leq n \leq N$, and we wish to predict $x(N + 1)$. We will say that our prediction is accurate if we are able to come within $\pm \varepsilon$ of the correct value. The predictability of the series will increase as the probability of reproducible subsequences of strings increases.

If we use only the information about the probability distribution of values in the series, then we will be able to come within ε of the correct value of $x(N + 1)$ about C_1 of the time. For example, if the probability distribution is unimodal and peaked near the average of the x's, then, typically, we will guess right about C_1 of the time if we just guess the average. Suppose now that δ_1 is nonzero. The greatest degree of predictability we can have from dependence on the first previous number in the time series is if $P(t_2|t_1) = 1$. In that case it follows from (7) and (8) that $\delta_1 =$

$1 - C_1$. Thus a measure of the degree of predictability coming from one dependence is

$$S_1 = C_1/(1 - \delta_1). \tag{14}$$

If $\delta_1 = 0$ then $S_1 = C_1$ and, given only the value of the immediately preceding element, the series is no more predictable than random. On the other hand, if $\delta_1 = 1 - C_1$ then $S_1 = 1$, which means that the series is completely predictable given the value of the first preceding element[1]. Suppose now that both δ_1 and δ_2 are nonzero. If $P(t_3|t_2, t_1) = P(t_3|t_2)$ then there is no additional predicability from the two-lag, and $\delta_2 = 0$. If $P(t_3|t_2, t_1) = 1$ then the series is completely predictable given the two previous numbers, and $\delta_2 = 1 - C_1/(1 - \delta_1)$. Therefore (disregarding for the moment the possibility of negative δ_j's), the quantity

$$S_2 = C_1/(1 - \delta_1)(1 - \delta_2) \tag{15}$$

will range between C_1 if δ_1 and δ_2 are both zero and the series is no more predictable than random given information of the two previous numbers, and 1 if the series is completely predictable given information of the two previous numbers. This process continues, and in general we may define

$$S = C_1 \Big/ \prod_{j=1}^{\infty} (1 - \delta_j). \tag{16}$$

S is a predictability index which takes on values $C_1 \leq S \leq 1$. If the series is no more predictable than a random (IID) series would be, $S = C_1$. If, on the other hand, a number in the series is completely determined by previous values of the series then $S = 1$. In many cases δ_j will be sensibly zero for $j > U$. In these cases, therefore, the denominator of (15) can be well estimated by

$$\prod_{j=1}^{U} (1 - \delta_j).$$

---

[1] In exceptional circumstances some δ_j's can be negative. See section 4 for a further discussion.

Another way to describe S is to say that S_j measures the degree to which subsequences of length j are reproducible in the series to within $\pm \varepsilon$. Finally, the meaning of S can be made still clearer by rewriting the δ_j's in terms of conditional probabilities. Doing this, it is not difficult to see that

$$S = \lim_{j \to \infty} P\left(t_j | t_{j-1}, t_{j-2}, \ldots, t_1\right).$$

Thus, if $S = 1$, then the series is eventually completely predictable in the sense defined above.

2.4. Generalized deltas

The δ_j's discussed above are indicators of dependence on the jth lag in a time series. However, a little thought will reveal that the test for j-dependence as expressed in the calculation of the δ_j's essentially uses information about the time series embedded in exactly $j + 1$ dimensions. We can also ask whether the embedding in k dimensions shows evidence for j-lag dependence, with $k > j + 1$. As one might expect, this information is not independent of the information contained in the δ_j's. But, as we shall explain below, it is often useful to view the information about dependencies in this more general way.

With this in mind, let us suppose we have constructed k-tuples from a time series, thereby embedding the series in k dimensions. We wish to derive a set of indicators which are small in k dimensions if there is no dependence on lags further than $j - 1$. If $k > j + 1$ we can write

$$C_k = P(t_k, t_{k-1}, \ldots, t_1)$$
$$= P(t_k | t_{k-1}, \ldots, t_1) P(t_{k-1}, \ldots, t_1)$$
$$= P(t_k | t_{k-1}, \ldots, t_1)$$
$$\quad \times P(t_{k-1} | t_{k-2}, \ldots, t_1)$$
$$\quad \times P(t_{k-2}, \ldots, t_1)$$
$$= P(t_k | t_{k-1}, \ldots, t_1) P(t_{k-1} | t_{k-2}, \ldots, t_1) \cdots$$
$$\quad \times P(t_{j+1} | t_j, \ldots, t_1) P(t_j, \ldots, t_1). \quad (17)$$

If we assume dependence on lags no further than $j - 1$ steps back, then

$$P\left(t_{m+j} | t_{m+j-1}, \ldots, t_1\right)$$
$$= P\left(t_{m+j} | t_{m+j-1}, \ldots, t_{m+1}\right)$$

for $1 \leq m \leq k - j$. Assuming time translation invariance,

$$P\left(t_{m+j} | t_{m+j-1}, \ldots, t_{m+1}\right)$$
$$= P\left(t_j | t_{j-1}, \ldots, t_1\right)$$
$$= \frac{P\left(t_j, t_{j-1}, t_{j-2}, \ldots, t_1\right)}{P\left(t_{j-1}, t_{j-2}, \ldots, t_1\right)}$$
$$= \frac{C_j}{C_{j-1}}. \quad (18)$$

Therefore, if there is no dependence on lags further back than $j - 1$ steps, C_k can finally be written as

$$C_k = \left(C_j / C_{j-1}\right)^{k-j} C_j. \quad (19)$$

Using (19) it is straightforward to construct generalizations of the δ_j's which are sensitive to the assumption of j-dependence in k dimensions. These are defined as

$$\delta_j^{[k]} = \frac{C_k - \left(C_j / C_{j-1}\right)^{k-j} C_j}{C_k}$$
$$= 1 - \left(\frac{C_j}{C_{j-1}}\right)^{k-j} \frac{C_j}{C_k}. \quad (20)$$

With this definition $\delta_j = \delta_j^{[j+1]}$. As with the ordinary δ_j's, these generalized $d_{j+m}^{[k]}$'s are zero for $1 \leq m \leq k - j - 1$ if the assumption of no more than j-dependence is valid in k dimensions. The extent to which they are nonzero is a measure of the extent to which that assumption is not true. The utility of the generalized δ's lies in the fact that a time series may have a j-lag dependence which is weak and results in a δ_j which is nonzero,

but not too large. (This may be the case, for example, in a noisy deterministic system.) In such a case it may not be clear from the value of δ_j whether or not there is j-dependence. On the other hand, if there is j-dependence, then $\delta_j^{[k]}$ should grow with increasing k, while values of $\delta_j^{[k]}$ that are roughly independent of k are consistent with no significant j-dependence. Examining the k-dependence of $\delta_j^{[k]}$ gives us a longer lever arm with which to study the j-dependence of the time series[#2]. Finally we note again that since the δ_j and the $\delta_j^{[k]}$ are both expressed in terms of the C_j, the δ_j are a complete set of variables in terms of which the $\delta_j^{[k]}$ can be expressed. Specifically, it is easy to show that

$$\delta_j^{[k]} = 1 - \prod_{m=j}^{k-1} (1 - \delta_m)^{k-m}. \tag{21}$$

3. Application to mathematical examples

3.1. Random numbers

First we apply our method to quasi-IID sequences generated using a good random number generator. For such sequences, we expect all of the δ_j to be essentially zero indicating that there are no dependencies of one member of the sequence on any other. Table 1 lists the values of δ_j, $j = 1, \ldots, 5$ for two IID sequences, one generated using a uniform distribution and the other using a Gaussian probability distribution. Each sequence consists of 1000 points and our choice of tolerance, ε, is equal to one half of the standard deviation of the sequence, as it is for all of the sequences examined in this section. Note that the largest value of δ_j is about 0.05. This gives a measure of the size of the statistical fluctuations about zero, which should be considered insignifi-

Table 1
δ_j and predictability index for random sequences.

j	Uniform random numbers		Gaussian random numbers	
	δ_j	S_j	δ_j	S_j
0		0.265		0.264
1	-0.00296	0.264	0.0495	0.278
2	-0.00882	0.262	-0.00314	0.277
3	0.00809	0.264	0.0158	0.281
4	0.0671	0.283	0.00738	0.283
5	0.0228	0.290	-0.00344	0.282

cant in series of 1000 points. A value for one of the δ_j significantly greater than this can, with reasonable confidence, be taken as an indication of dependence. Notice also that some of the δ_j are negative. These negative values are not statistically significant in these series. However it is possible to obtain negative values of δ_j that are significant in certain systems. We shall discuss this in section 4.

In table 1 we also list values of S_j, the approximation to the predictability index taking into account δ_k, $k = 1, \ldots, j$. In both cases we see that S_j is approximately equal to C_1 independent of j, as we expect for a random sequence.

3.2. The tent map

The next simplest system to which to apply our method is the tent map defined by

$$\begin{aligned} x(n+1) &= 2x(n), & x(n) \leq 1/2 \\ &= -2x(n) + 2, & x(n) \geq 1/2 \\ &\equiv T[x(n)]. \end{aligned} \tag{22}$$

This map produces a chaotic sequence of numbers uniformly distributed on the unit interval. The autocorrelation functions of a sequence generated by this map are the same as those of a series of uniformly distributed IID numbers [6]. The first column in table 2 ($\sigma = 0$) lists the δ_j for $j = 1, \ldots, 7$ calculated for a sequence of 1000 tent map-generated points. The large value of δ_1 indicates that $P(t_2 t_1) \neq P(t_2)$ and so the sequence

---

[#2]In fact, $\delta_1^{[k]}$ are just proportional to the numerator of the BDS statistic (eqs. (4) and (5)), in $k+1$ embedding dimensions. The dependence of the BDS statistic on embedding dimension has been useful in examining the null hypothesis of IID behavior. See refs. [2, 3] for more discussion.

Table 2
δ_j and predictability index for sequences generated by the tent map with noise of strength σ. See eqs. (22) and (23).

j	$\sigma = 0.0$		$\sigma = 0.1$		$\sigma = 0.3$		$\sigma = 0.6$	
	δ_i	S_j	δ_j	S_j	δ_j	S_j	δ_j	S_j
0		0.267		0.268		0.269		0.274
1	0.521	0.557	0.508	0.544	0.472	0.509	0.285	0.383
2	0.0145	0.566	0.00595	0.548	0.0117	0.516	0.015	0.389
3	− 0.00632	0.562	− 0.00880	0.543	0.000916	0.516	− 0.00498	0.387
4	0.000467	0.562	− 0.00295	0.541	− 0.00196	0.515	0.00934	0.390
5	− 0.00424	0.560	− 0.00623	0.538	0.00440	0.517	− 0.00844	0.388
6	0.00426	0.562	− 0.00275	0.536	− 0.0201	0.507	− 0.0512	0.369
7	− 0.00870	0.557	0.0128	0.543	0.00976	0.512	0.0159	0.375

exhibits one dependence. It also tells us, a fortiori that the data are not IID, in sharp contrast to the results for the random sequence. The rest of the δ_j are small, indicating that there is no significant dependence greater than one, and that two dimensions are enough to reveal the dynamics of the attractor. These results are clearly consistent with eq. (22).

Table 2 also shows the results of applying our test to the tent map plus uniform IID noise. These sequences are generated according to the equation

$$x(n + 1) = (T[x(n)] + \sigma\eta) \mod 1, \qquad (23)$$

where η is a uniform random deviate between − 0.5 and 0.5 and σ is a parameter describing the strength of the noise. It is not difficult to see that as σ varies between 0 and 1 the sequence crosses over from a purely deterministic sequence ($\sigma = 0$) to a purely random one ($\sigma = 1$). Consistent with this, we note that as σ increases δ_1 decreases. On the other hand, δ_1 is still significant even for moderately large values of σ. Thus, our test is sensitive to the underlying dynamical dependence, even in the presence of significant noise. Of course, when $\sigma = 1$, δ_1 is statistically indistinguishable from zero.

In table 2 we also present S_j for the tent map with different amounts of noise. We note that (i) for all the maps listed here, S_1 is considerably larger than S_0 ($= C_1$), indicating additional infor-

mation and predictability from incorporating information about one-dependence, (ii) S_j is substantially unchanged for $j > 1$ and (iii) for a given j, S_j decreases with increasing noise. All of these effects are just what we expect. We note also that for the tent map without noise, S_1 is about $1/2$, whereas one might expect that S_1 should be one since for $j = 1$ we have incorporated all the dynamics of the tent map and there is no additional noise. However, since ε is nonzero we are not incorporating all the detailed information about the map into this calculation, nor are we seeking predictability to better than $\pm\varepsilon$. As ε goes to zero S_1 is expected to increase toward one in the absence of noise, as we shall discuss in section 5.

It is worthwhile to describe one other effect. Consider a sequence generated by taking every third iterate of the tent map, i.e.,

$$x(n + 1) = T^{[3]}[x(n)],$$

with $T(x)$ given in (22). It turns out that the δ_j for this sequence are all small. No indication of one-dependence is present as would be expected from a one-dependent deterministic map. Further, the predictability index $S \cong C_1$ despite the fact that the sequence is deterministic. This can be understood by noting that $T^{[3]}(x)$ oscillates rapidly between zero and one and in fact oscillates several times in an interval of length ε (which we set equal to one half the standard

deviation, as previously explained). Hence the sequence is not reproducible on this scale of ε because the underlying map produces sequences which are too rapidly divergent. In section 5 we shall discuss further the ε dependence of indicators.

3.3. The logistic map

Next we examine the logistic map at the point of fully developed chaos. This map, defined by the equation

$$x(n + 1) = 4x(n)[1 - x(n)], \tag{24}$$

also generates a chaotic sequence of numbers distributed on the unit interval. However, in this case the distribution is not uniform because of the nonuniform curvature in the quadratic map. The first column in table 3 lists the values of δ_j computed for a sequence of 1000 points generated from this map with ε again equal to half the standard deviation of the distribution. These results clearly indicate one-dependence, consistent with the form of the map (24). However, note that δ_2, although small, is just on the verge of being significant. In fact, even though according to the map (24) $x(n + 1)$ depends explicitly only on $x(n)$, δ_2 is significantly different from zero for some values of ε. This is due to the nonuniform curvature of the map (24), which induces changes in the nontrivial dependencies in higher condi-

tional probabilities as we shall discuss in more detail in section 5. This effect notwithstanding, it is nonetheless true that the major dependence indicated by the δ_j's is dependence on the immediately preceding variable, as we expect from eq. (24).

Finally, in table 3 we also show the values of δ_j calculated for sequences of 1000 points generated by a noisy logistic map. In particular, the form (23) was used with $T[x(n)]$ taken to be the right-hand side of (24). As with the tent map, δ_1 decreases with increasing noise, but even for fairly large values of σ, it is large enough to clearly signal the presence of an underlying deterministic process. The values of S_j are also included in this table. They follow the same qualitative pattern as those in table 2, and similar comments apply here also.

3.4. The Hénon map

The Hénon map has a chaotic regime which defines a strange attractor with correlation dimension $\nu \cong 1.22$. It is defined by the equation

$$x(n + 1) = 1 - ax^2(n) + bx(n - 1) \tag{25}$$

with $a = 1.4$ and $b = 0.3$. The δ_j for this map, listed in table 4, indicate two-dependence as expected from the form of (25). Although both δ_1 and δ_2 are large, δ_1 is significantly larger than δ_2 showing that the dependence of $x(n + 1)$ is

Table 3
δ_j and predictability index for sequences generated by the logistic map with noise of strength σ. See eq. (24).

j	$\sigma = 0.0$		$\sigma = 0.1$		$\sigma = 0.3$		$\sigma = 0.6$	
	δ_j	S_j	δ_j	S_j	δ_j	S_j	δ_j	S_j
0		0.296		0.290		0.283		0.284
1	0.452	0.540	0.390	0.475	0.364	0.445	0.193	0.352
2	0.0377	0.561	0.0726	0.513	0.0474	0.467	0.00841	0.355
3	−0.0166	0.552	−0.00315	0.511	0.00630	0.470	−0.00284	0.354
4	−0.0134	0.545	−0.00457	0.509	0.000818	0.470	0.0140	0.359
5	−0.0117	0.539	0.000965	0.509	−0.00296	0.469	0.0194	0.366
6	−0.000480	0.539	0.0180	0.519	0.00119	0.469	0.0143	0.371
7	−0.0165	0.530	−0.0337	0.502	−0.0142	0.463	0.0204	0.379

Table 4
δ_j and predictability index for a sequence generated by the Hénon map. See eq. (25).

j	δ_j	S_j
0		0.290
1	0.442	0.520
2	0.193	0.644
3	-0.0343	0.623
4	0.0397	0.648
5	0.0225	0.663
6	0.00574	0.667
7	0.0295	0.687
8	-0.00171	0.686
9	0.0435	0.717

Table 5
δ_j for interwoven sequences of logistic and Hénon maps. See eqs. (26) and (27).

	3 logistic maps	2 Hénon maps
δ_1	0.00163	0.0332
δ_2	-0.00215	0.428
δ_3	0.455	-0.00213
δ_4	0.00258	0.198
δ_5	-0.00800	0.00985
δ_6	0.0242	-0.0363
δ_7	-0.0196	-0.0165
δ_8	0.0242	0.0297
δ_9	-0.0280	-0.0241

stronger on $x(n)$ than is the conditional dependence on $x(n-1)$. One interesting practical implication of this result is the following: According to a theorem of Takens [7], if a sequence possesses an attractor that can be embedded in a manifold of dimension m, then an embedding dimension for the times series of dimension $2m + 1$ is sufficient to capture all the topological structure of the attractor. On the other hand, if from a practical point of view, one is only interested in the structure of the attractor to within ε, then a sufficient embedding dimension for the time series is just given by the largest value of j for which the δ_j's are significantly different from zero. Sometimes this will be smaller than the Takens limit.

We include in table 4 the values of S_j for the Hénon map. Note that $S_0 < S_1 < S_2$, and that the S_j for $j > 2$ are substantially unchanged, as we expect.

3.5. Interwoven sequences

To further test the efficacy of our procedure, we have studied some sequences composed of interwoven sets of numbers generated by simple maps. Table 5 shows the values of the δ_j generated from a sequence of three logistic maps inter-

woven according to the equations

$$x(3n + 3) = 4x(3n)[1 - x(3n)],$$

$$x(3n + 4) = 4x(3n + 1)[1 - x(3n + 1)],$$

$$x(3n + 5) = 4x(3n + 2)[1 - x(3n + 2)]. \quad (26)$$

These results are completely consistent with our expectations. The only sizable δ_j is δ_3. The small values of the rest of the δ_j indicate that there are no dependencies other than $j = 3$ in the data, as we expect from the map (26). (We are ignoring here the marginal induced dependence referred to in section 3.3.)

Also in table 5 are listed the results of applying our test to a sequence of two Hénon maps interwoven according to the equations

$$x(2n + 1) = 1 - ax^2(2n - 1) + bx(2n - 3),$$

$$x(2n + 2) = 1 - ax^2(2n) + bx(2n - 2). \quad (27)$$

Only δ_2 and δ_4 are significantly different from zero, consistent with eqs. (27).

4. Negative δ_j's

We have already noted that some of the δ_j's are negative. From (13) and (10) one sees that

$$\delta_j = 1 - \frac{P(t_{j+1}|t_j, \ldots, t_2)}{P(t_{j+1}|t_j, \ldots, t_1)}, \quad (28)$$

so

$$\delta_j < 0 \Leftrightarrow P\big(t_{j+1}|t_j,\ldots,t_1\big) < P\big(t_{j+1}|t_j,\ldots,t_2\big).$$

This statement, that we lose predictability (or, better, reproducibility) by considering more information, seems contradictory, but can be understood in light of the exact relation among conditional probabilities

$$P\big(t_{j+1}|t_j,\ldots,t_2\big) = P\big(t_{j+1}|t_j,\ldots,t_1\big)P(t_1)$$
$$+P\big(t_{j+1}|t_j,\ldots,\overline{t_1}\big)\,P(\overline{t_1}), \qquad (29)$$

where $\overline{t_1}$ stands for not t_1 and $P(\overline{t_1}) = 1 - P(t_1)$. From (29) it is easy to show that

$$\delta_j < 0 \Leftrightarrow P\big(t_{j+1}|t_j,\ldots,t_1\big) < P\big(t_{j+1}|t_j,\ldots,t_2\big)$$
$$< P\big(t_{j+1}|t_j,\ldots,\overline{t_1}\big) \qquad (30)$$

and

$$\delta_j = 0 \Leftrightarrow P\big(t_{j+1}|t_j,\ldots,t_1\big) = P\big(t_{j+1}|t_j,\ldots,t_2\big)$$
$$= P\big(t_{j+1}|t_{j+1}|t_j,\ldots,\overline{t_1}\big). \qquad (31)$$

If $\delta_j < 0$ we gain predictability for t_{j+1} by considering the information that t_1 is not true rather than the information that t_1 is true. In maps with nonuniform curvature, trajectories will diverge at different rates (or may even temporarily converge) depending on the local curvature. It is possible, therefore, that the probability that two trajectories will be close several time steps in the future will be larger if they are not close cur-

rently. Such an effect will result in negative values of δ_j which are statistically significant.

As a simple example of such a sequence, consider a sequence of the symbols 1, 2, and 3 formed according to the following rules:

1 is always followed by 2;

2 is always followed by 3;

3 is followed with 50% probability

by either 1 or 2. (32)

Such a sequence obviously does not have uniform divergence of trajectories since it is completely deterministic in some parts and uniformly random in others. The δ_j's for this sequence, presented in table 6, show a strong indication of one-dependence and a smaller indication of three-dependence as we might expect from the rules (32). In addition δ_2 is negative. In fact, one can calculate the probababilities exactly for this sequence. We find that $P(t_3|t_2) = \frac{7}{9}$, $P(t_3|t_2,t_1) = \frac{5}{7}$, and $P(t_3|t_2,\overline{t_1}) = 1$ so that (30) for $j = 2$ is indeed satisfied and $\delta_2 = -\frac{4}{45}$, consistent with table 6. A similar analysis shows that the negative value for δ_4 is also a significant indication and not just a statistical fluctuation. To test in general whether an observed negative value of δ_j is significant, one can study the sequence $\delta_j^{[k]}$ and see whether these numbers become increasingly negative as k increases. In the example (32), the $\delta_2^{[k]}$ and the $\delta_4^{[k]}$ remain negative and increase in magnitude with increasing k indicating further that the negative values provide significant information about the sequence as discussed at the

Table 6

The $\delta_j^{[k]}$, for $j = 1,\ldots,7$ and $k = 2,\ldots,8$ for a sequence generated by the rules (32).

k	$\delta_1^{[k]}$	$\delta_2^{[k]}$	$\delta_3^{[k]}$	$\delta_4^{[k]}$	$\delta_5^{[k]}$	$\delta_6^{[k]}$	$\delta_7^{[k]}$
2	0.538						
3	0.768	−0.0905					
4	0.896	−0.0610	0.108				
5	0.950	−0.101	0.151	−0.0667			
6	0.976	−0.144	0.191	−0.139	−0.000991		
7	0.989	−0.147	0.256	−0.174	0.033	0.0350	
8	0.995	−0.197	0.288	−0.259	0.0273	0.0302	−0.0415

end of section 2. Specifically in this case the negative values provide information about the presence of regions of alternating convergence and divergence of trajectories.

The logistic map, because of its nonuniform curvature, also shows regions of varying degrees of divergence as well as a region of convergence. (See the discussion in section 5.3.) The $\delta_j^{[k]}$ for this map are presented in table 7. Although the magnitudes of δ_3 through δ_7 are too small to indicate any dependence, the $\delta_j^{[k]}$ for these dependencies are negative and grow in magnitude with increasing k (particularly for $j = 3$ and 4). Hence, when ε equals one half the standard deviation of the sequence, there is a significant probability that two subsequences of the logistic map whose initial separation is greater than ε will converge to become temporarily separated by less than ε. At other values of ε, this effect may be less pronounced. The results for the logistic map may be contrasted with the results for the tent map presented in table 8. Note that while most of δ_3 through δ_7 are negative, there is no

trend of significant growth in magnitude of the $\delta_j^{[k]}$ with k as there is for the logistic map. But the tent map has a slope of constant magnitude and hence a uniform rate of divergence so these results are consistent with our expectations.

5. ε dependence

Thus far in this paper we have chosen the tolerance, ε, to be one half the standard deviation of the sequence being examined, and we have found that with this choice of ε our test gives a reasonable description of the dependencies involved. We now wish to examine the sensitivity of these results to the choice of ε. Specifically, we wish to examine the results in the limit that ε becomes small.

The ε dependence of the coefficients $C_d(\varepsilon)$ has already been studied by several investigators. Takens [7] has considered the case of a time sequence which has a chaotic attractor. He has shown that subject to some technical assump-

Table 7

The $\delta_j^{[k]}$, for $j = 1, \ldots, 7$ and $k = 2, \ldots, 8$ for a sequence generated by the logistic map.

k	$\delta_1^{[k]}$	$\delta_2^{[k]}$	$\delta_3^{[k]}$	$\delta_4^{[k]}$	$\delta_5^{[k]}$	$\delta_6^{[k]}$	$\delta_7^{[k]}$
2	0.452						
3	0.711	0.0377					
4	0.845	0.0586	−0.0166				
5	0.915	0.0667	−0.0473	−0.0134			
6	0.956	0.0639	−0.0916	−0.0391	−0.0117		
7	0.975	0.0606	−0.138	−0.0659	−0.0240	−0.000480	
8	0.986	0.0418	−0.207	−0.111	−0.0536	−0.0174	−0.0165

Table 8

The $\delta_j^{[k]}$, for $j = 1, \ldots, 7$ and $k = 2, \ldots, 8$ for a sequence generated by the tent map.

k	$\delta_1^{[k]}$	$\delta_2^{[k]}$	$\delta_3^{[k]}$	$\delta_4^{[k]}$	$\delta_5^{[k]}$	$\delta_6^{[k]}$	$\delta_7^{[k]}$
2	0.521						
3	0.774	0.0145					
4	0.892	0.0227	−0.0632				
5	0.949	0.0312	−0.0122	0.00047			
6	0.976	0.0356	−0.0225	−0.00331	−0.00424		
7	0.988	0.0440	−0.0284	−0.00281	−0.00421	−0.00426	
8	0.994	0.0442	−0.0434	−0.0110	−0.0129	−0.000124	−0.00870

tions[#3] the $C_d(\varepsilon)$ are related to the correlation dimension ν and K_2, the second-order Kolmogorov entropy, of the attractor by the following formulas:

$$\nu = \lim_{d \to \infty} \left(\lim_{\varepsilon \to 0} \frac{\ln C_d(\varepsilon)}{\ln \varepsilon} \right) \qquad (33)$$

and

$$K_2 = \lim_{\varepsilon \to 0} \left(\lim_{d \to \infty} \frac{\ln C_d(\varepsilon)}{-d} \right). \qquad (34)$$

Takens has further shown that if m is the dimension of a manifold containing the attractor, then the right-hand side of (33) reaches its $d \to \infty$ limit for every $d \geq 2m$. Accordingly, Grassberger and Procaccia [1] assume that

$$C_d(\varepsilon) \underset{\substack{d \to \infty \\ \varepsilon \to 0}}{\sim} \varepsilon^\nu e^{-dK_2}$$

and investigate corrections to this leading behavior. It is these corrections which determine the asymptotic behavior of the conditional probabilities and the $\delta_j(\varepsilon)$.

Without loss of generality we can define corrections to (33) and (34) by letting

$$C_d(\varepsilon) = \varepsilon^{\nu + \alpha_d} e^{-d[K_2 + \beta(\varepsilon)]} e^{\gamma_d(\varepsilon)}. \qquad (35)$$

As shown in the appendix, any $C_d(\varepsilon)$ satisfying (33) and (34) can be written in the form (35) with α, β and γ satisfying the limits

$$\lim_{d \to \infty} \alpha_d = \lim_{\varepsilon \to 0} \beta(\varepsilon) = \lim_{d \to \infty} \frac{\gamma_d(\varepsilon)}{-d}$$

$$= \lim_{\varepsilon \to 0} \frac{\gamma_d(\varepsilon)}{\ln \varepsilon} = 0. \qquad (36)$$

These corrections determine the ε dependence of the conditional probabilities since

$$P\left(t_{j+1} | t_j, t_{j-1}, \ldots, t_1\right)$$

$$= \frac{C_{j+1}(\varepsilon)}{C_j(\varepsilon)}$$

$$= e^{-[K_2 + \beta(\varepsilon)]} \varepsilon^{\alpha_{j+1} - \alpha_j} e^{\gamma_{j+1}(\varepsilon) - \gamma_j(\varepsilon)}$$

$$= e^{-[K_2 + \beta(\varepsilon)]} \varepsilon^{D_1(\alpha_{j+1})} e^{D_1(\gamma_{j+1}(\varepsilon))}, \qquad (37)$$

where D_n is the nth discrete difference operator. In addition from (13) we see that

$$\delta_j(\varepsilon) = 1 - e^{-\gamma_{j+1}(\varepsilon) + 2\gamma_j(\varepsilon) - \gamma_{j-1}(\varepsilon)} \varepsilon^{-\alpha_{j+1} + 2\alpha_j - \alpha_{j-1}}$$

$$= 1 - e^{-D_2(\gamma_j(\varepsilon))} \varepsilon^{-D_2(\alpha_j)}, \qquad (38)$$

so the curvature of α_j and γ_j as a function of j determines the ε-dependence of the δ_j[#4].

The discussion of section 4 notwithstanding, the most significant δ_j's are usually nonnegative so that the conditional probabilities in (37) do not generally decrease as the dependence increases, i.e.

$$P\left(t_{j+2} | t_{j+1}, t_j, \ldots, t_1\right) \geq P\left(t_{j+1} | t_j, t_{j-1}, \ldots, t_1\right).$$

If this inequality holds for all ε, then

$$\ln(\varepsilon) D_1(\alpha_{j+2}) + D_1(\gamma_{j+2}(\varepsilon))$$

$$\geq \ln(\varepsilon) D_1(\alpha_{j+1}) + D_1(\gamma_{j+1}(\varepsilon)), \qquad (39)$$

or

$$D_2(\alpha_j) + D_2\left(\frac{\gamma_j(\varepsilon)}{\ln \varepsilon}\right) \leq 0 \quad \text{for all } \varepsilon < 1. \qquad (40)$$

Then, taking the $\varepsilon \to 0$ limit and using the fourth limit from (36) we see that

$$D_2(\alpha_j) \leq 0. \qquad (41)$$

[#3]However, some unusual sequences do not obey these assumptions. See the discussion of the Zaslaskii map in ref. [1] for an example of a sequence whose $C_d(\varepsilon)$ do not exhibit power law behavior in ε and hence do not satisfy (33). The ε-dependence of the δ_j for such sequences has not yet been studied.

[#4]In order that (38) be correct for $j = 1$, and also that (8) fit the form of (13), we normalize by defining $C_0(\varepsilon) = 1$, $\beta(\varepsilon) = -K_2$ and $\gamma_0(\varepsilon) = 0$ for all ε, and $\alpha_0 = -\nu$.

This result, that in well-behaved cases the function α_j is not concave, restricts the limiting behavior of the $\delta_j(\varepsilon)$ considerably, as we shall see.

To study this limiting behavior of $\delta_j(\varepsilon)$ more closely, we consider separately the two cases $D_2(\alpha_j) < 0$ and $D_2(\alpha_j) = 0$. Whenever $D_2(\alpha_j) < 0$ the term $\varepsilon^{-D_2(\alpha_j)}$ in (38) approaches zero as ε approaches zero. Although the factor $e^{-D_2(\gamma_j(\varepsilon))}$ may diverge as ε approaches zero, it follows from the fourth limit in (36) that it will diverge slowly enough so that

$$\lim_{\varepsilon \to 0} e^{-D_2(\gamma_j(\varepsilon))} \varepsilon^{-D_2(\alpha_j)} = 0.$$

Hence from (38) whenever $D_2(\alpha_j) < 0$,

$$\lim_{\varepsilon \to 0} \delta_j(\varepsilon) = 1. \tag{42}$$

If $D_2(\alpha_j) = 0$ we see from (38) that

$$\delta_j(\varepsilon) \underset{\varepsilon \to 0}{\sim} 1 - e^{-D_2(\gamma_j(\varepsilon))}, \tag{43}$$

and from (40) that

$$D_2\left(\frac{\gamma_j(\varepsilon)}{\ln \varepsilon} \right) \leq 0 \quad \text{for all } \varepsilon < 1$$

or,

$$D_2(\gamma_j(\varepsilon)) \geq 0 \quad \text{for all } \varepsilon < 1. \tag{44}$$

Hence, whenever $D_2(\alpha_j) = 0$, $\delta_j(\varepsilon)$ is less than 1 for all ε less than 1. Note that this does not rule out the possibility that $\lim_{\varepsilon \to 0} \delta_j(\varepsilon) = 1$ when $D_2(\alpha_j) = 0$. But the fourth limit in (36) and eq. (43) imply that when $D_2(\alpha_j) = 0$, $\delta_j(\varepsilon)$ approaches its limiting value more slowly than ε^κ for any positive real exponent κ. Thus, since one is unable to observe the behavior of $\delta_j(\varepsilon)$ with infinite precision, it is difficult to discern the $\varepsilon \to 0$ limiting value of δ_j if $D_2(\alpha_j) = 0$. On the other hand, if $D_2(\alpha_j) < 0$, then the factor $\varepsilon^{-D_2(\alpha_j)}$ in (38) controls the approach of $\delta_j(\varepsilon)$ to one. This approach is obviously much easier to observe. In the examples below we study the behavior of

$\delta_j(\varepsilon)$ to a precision $\varepsilon \cong 10^{-4}$ and a clear approach to one is never observed unless $D_2(\alpha_j) < 0$.

The saturation of the $\lim_{d \to \infty}$ for $d \geq 2m$ (where m is the dimension of a manifold containing the chaotic attractor of the time sequence under investigation) in (33) further constrains the limiting behavior of the $\delta_j(\varepsilon)$ for $j \geq 2m + 1$. This saturation means that $\alpha_j = 0$ for $j \geq 2m$. The results we examine below indicate that in some cases the $d \to \infty$ limit in (33) becomes saturated at a value of d smaller than $2m$, so in these cases $\alpha_j = 0$ for some values of j less than $2m$. In general we can state that for every sequence for which the limits in (33) and (34) are well defined, there exists an integer $j_0 \leq 2m + 1$ such that

$$\alpha_j = 0 \quad \text{for } j \geq j_0 - 1,$$

$$\alpha_j \neq 0 \quad \text{for } j = j_0 - 2. \tag{45}$$

Then, using (41), we have $D_2(\alpha_j) = 0$ for $j \geq j_0$, and $D_2(\alpha_j) < 0$ for $j = j_0 - 1$. Therefore,

$$\lim_{\varepsilon \to 0} \delta_{j_0 - 1}(\varepsilon) = 1,$$

$$\delta_j(\varepsilon) < 1 \quad \text{when } \varepsilon < 1 \text{ for } j \geq j_0. \tag{46}$$

$\delta_j(\varepsilon)$ can exhibit either of these two behaviors when $j < j_0 - 1$.

Next we study some examples and examine the asymptotic behavior of the $\delta_j(\varepsilon)$ for the tent map, the logistic map and the Hénon map.

5.1. The tent map

Fig. 1a is a plot of $\delta_1(\varepsilon)$ for the tent map. Note that $\delta_1(\varepsilon)$ approaches 1 as ε approaches 0. By contrast in fig. 1b, we plot $\delta_2(\varepsilon)$, $\delta_3(\varepsilon)$ and $\delta_4(\varepsilon)$. All these $\delta_j(\varepsilon)$'s are consistent with zero for the range of ε shown. These results follow from certain general properties of the tent map. The tent map has correlation dimension $\nu = 1$ and so $\alpha_0 = -1$ (see footnote 4). Any time sequence defines a measure, called its natural measure, as follows: The measure $m(I)$ of an interval I is
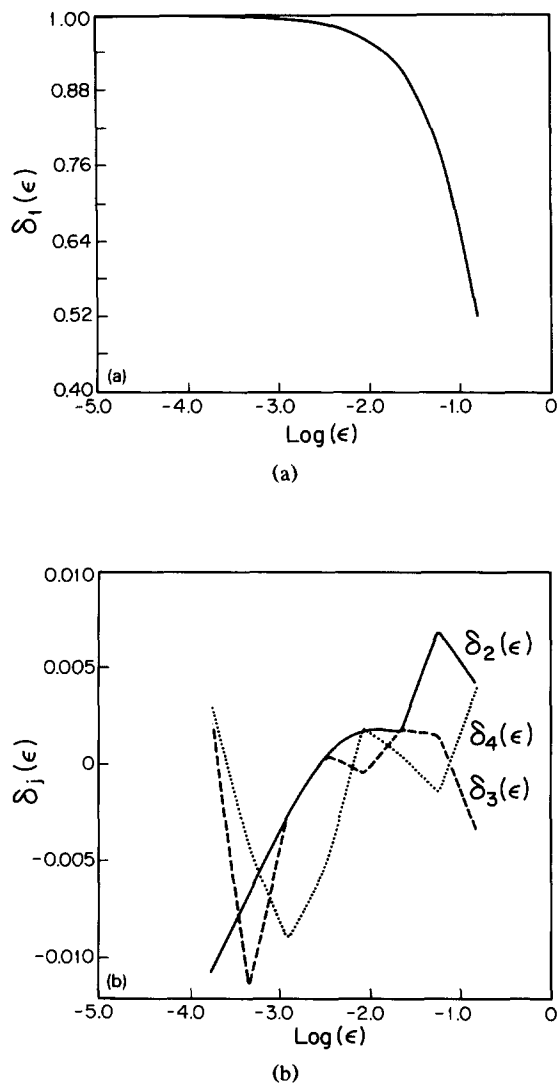
Fig. 1. (a) $\delta_1(\varepsilon)$ for the tent map. (b) $\delta_2(\varepsilon)$ through $\delta_4(\varepsilon)$ for the tent map.

defined to be the frequency with which the time sequence visits that interval. This allows us to define a density function $\rho(x)$ by requiring that for any interval I,

$$m(I) = \int_I \rho(x)\,\mathrm{d}x. \tag{47}$$

If for any x_0 and y, $g(y) \equiv \int_{x_0}^{y}\rho(x)\,\mathrm{d}x$ is continuous then we say that the sequence has a continu-

ous natural measure and we conclude that

$$C_1(\varepsilon) = \int \mathrm{d}y\,\rho(y)\int_{y-\varepsilon}^{y+\varepsilon}\rho(x)\,\mathrm{d}x$$

$$\underset{\varepsilon\to 0}{\sim}\, 2\left(\int\rho^2(y)\,\mathrm{d}y\right)\varepsilon \underset{\varepsilon\to 0}{\sim} \varepsilon^{\nu+\alpha_1}, \tag{48}$$

where the last equality follows from (35) and implies that $\alpha_1 + \nu = 1$. For the tent map $\rho(x) \equiv 1$ so that (48) applies, and in addition $\nu = 1$. Thus $\alpha_1 = 0$. Furthermore, it is clear from (22) that a one-dimensional manifold will contain the attractor of the tent map, and so (45) implies that for some $j_0 \leq 3$, α_j equals zero for $j \geq j_0 - 1$. In addition, from (41) we know that α_j is never concave. The only way all three of these conditions can be satisfied is if $\alpha_j = 0$ for $j \geq 1$ so that $j_0 = 2$. Hence $D_2(\alpha_1) = \alpha_0 = -1$ and from (43) we see that $\delta_1(\varepsilon)$ approaches 1 as ε approaches 0. Also from (43) we find that for $j \geq 2$, $\delta_j(\varepsilon)$ is less than one when ε is less than one. But the tent map is a simple one-dimensional deterministic map having a slope with constant magnitude, so for any value of ε we expect that we will attain maximum predictability with only a one-lag dependence. This is reflected in the observed small values for $\delta_2(\varepsilon)$ through $\delta_4(\varepsilon)$.

When $D_2(\alpha_j) = 0$ and $\delta_j(\varepsilon)$ is small (and conversely when $D_2(\gamma_j(\varepsilon))$ is small) it can be shown from (43) that $\delta_j(\varepsilon) \sim D_2(\gamma_j(\varepsilon))$, and hence that when $D_2(\alpha_j) = 0$

$$\delta_j(\varepsilon) \ll 1 \Leftrightarrow D_2\big(\gamma_j(\varepsilon)\big) \ll 1. \tag{49}$$

Thus the corrections $D_2(\gamma_j(\varepsilon))$ are also small for all $j \geq 2$. This result will be compared below to the quite different results from the logistic map where we will see that two simple heuristic arguments concerning the variation of slope of the two maps lead to a better understanding of the values for $\delta_2(\varepsilon)$ through $\delta_4(\varepsilon)$ in both maps.

5.2. The logistic map

The logistic map also has a correlation dimension $\nu = 1$ and a continuous natural measure so it

follows that $j_0 = 2$ exactly as for the tent map. Fig. 2a shows $\delta_1(\varepsilon)$ for the logistic map, and indeed it asymptotically approaches one as in the tent map. However, $\delta_2(\varepsilon)$ and $\delta_3(\varepsilon)$ for the logistic map (shown in figs. 2b and 2c) do not show the same small fluctuations about zero as they do in the tent map. Rather, as $\varepsilon \to 0$ they approach a non-zero limit less than one. For $j \geq 4$ the $\delta_j(\varepsilon)$ are statistically consistent with zero for small ε. Thus, for small enough ε the logistic map exhibits two- and three-dependence even though the map is explicitly only one-dependent. This extra dependence is induced by the non-uniform slope of the logistic map, which causes a nonuniform divergence of nearby trajectories and hence a slower approach to zero of the $\delta_j(\varepsilon)$ as a function of j [#5]. Since $\delta_2(\varepsilon)$ and $\delta_3(\varepsilon)$ are nonzero for the logistic map, $D_2(\gamma_j(\varepsilon))$ is also statistically nonzero for $j = 2$ and 3, according to eq. (49). This can be understood by noting that for the tent map, the parameters ν and K_2 are sufficient to describe the divergence of nearby trajectories (which after all is determined by only one parameter, the constant slope of the tent map). Hence, the corrections, α_d, $\beta(\varepsilon)$, and $\gamma_j(\varepsilon)$ are all small. By contrast, the nonuniform divergence of nearby trajectories determined by the continuous variation in slope of the logistic map requires the correction terms as well as ν and K_2 for a complete description. Thus, some of these corrections have significant values.

Fig. 3 illustrates more explicitly how the nonuniform slope changes the conditional probabilities and hence the $\delta_j(\varepsilon)$. For example, there is a large interval near the peak of the logistic map on which its slope is less than one, and so trajectories generated by the map which begin in this interval are initially converging. In this region $P(t_2|t_1) = 1$, which weights the conditional proba-

[#5]It is likely that strictly speaking, the $\delta_j(\varepsilon)$ for $j \geq 4$ are also nonzero in the limit $\varepsilon \to 0$. Such behavior could probably be seen with better statistics. In any case it is important to recognize that the strength of dependence of the time series on these variables is quite small and would only be relevant if great precision were required.
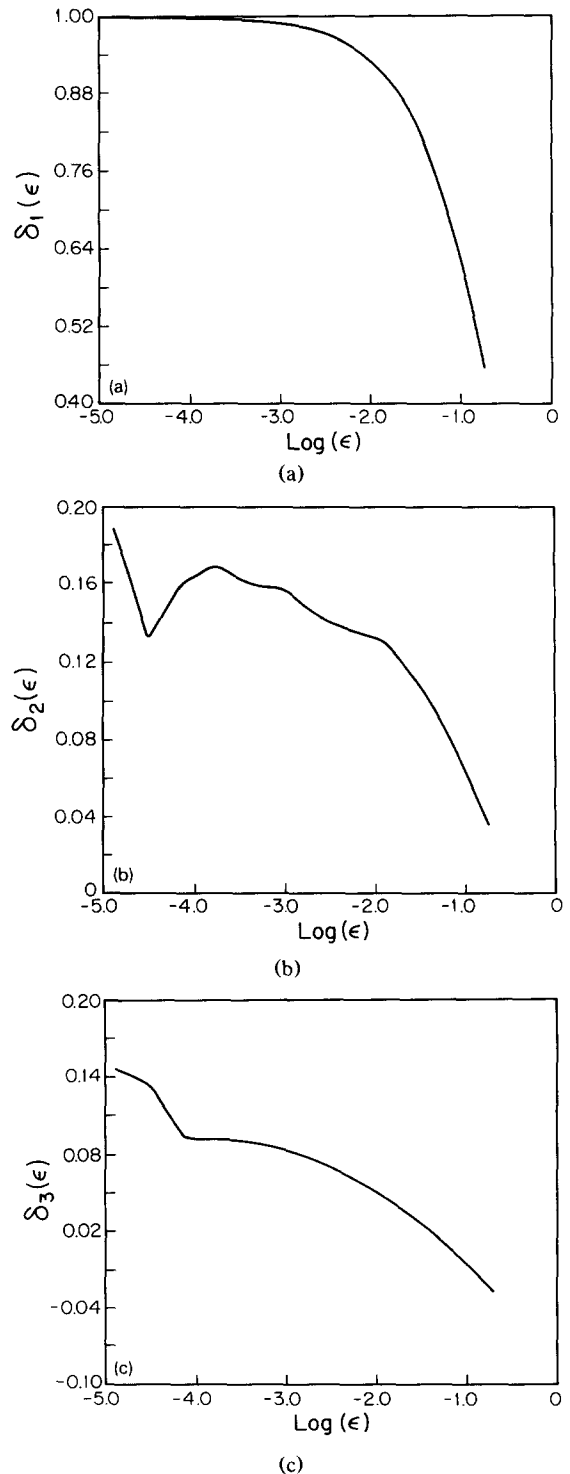

Fig. 2. (a) $\delta_1(\varepsilon)$ for the logistic map. (b) $\delta_2(\varepsilon)$ for the logistic map. (c) $\delta_3(\varepsilon)$ for the logistic map.
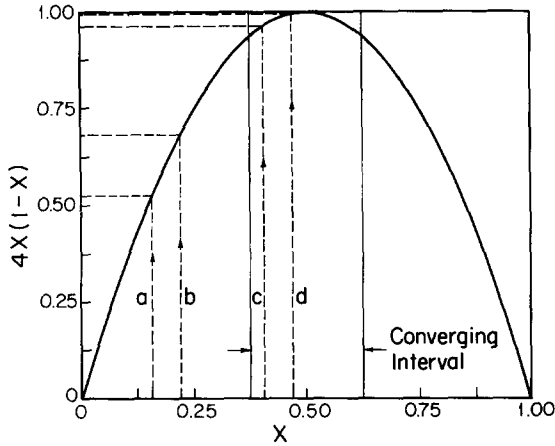
Fig. 3. Non-uniform divergence of the logistic map. Trajectories a and b have an initial separation of $1/16$ and are separated by $5/32$ after one iteration of the logistic map. Trajectories c and d lie inside the converging interval ($\frac{3}{8}, \frac{5}{8}$). They also have an initial separation of $1/16$ but they are separated by only $1/32$ after one iteration of the map. After the first iteration they diverge. For two trajectories such as c and d which lie initially inside ($\frac{3}{8}, \frac{5}{8}$), $P(t_2|t_1) = 1$.

bility averaged over the entire unit interval toward increased dependence. This effect becomes significant when ε becomes smaller than the length of the converging interval. While the presence of a large converging interval is the dominant cause of induced dependence in the logistic map, any variation in the slope of a map produces a similar effect. Thus, our test gives no indication of nontrivial induced dependence in the tent map, which is uniformly divergent. As discussed in section 1, any map exhibits certain kinds of induced dependencies, but only those induced dependencies which change the conditional probabilities are indicated by our test. It is these conditional probabilities that are the important variables in trying to reconstruct the underlying map.

It is also interesting to examine the effect of noise on the $\delta_j(\varepsilon)$. Figs. 4a–4c show the $\delta_j(\varepsilon)$, $j = 1, 2, 3$ for the logistic map plus uniform noise between -0.5 and 0.5 of strength $\sigma = 0.1$. Note that in all three cases the effect of the noise is to lower the limit approached as $\varepsilon \to 0$. (In fact, for $j = 2$ and 3 the value of $\delta_j(\varepsilon)$ for small ε is
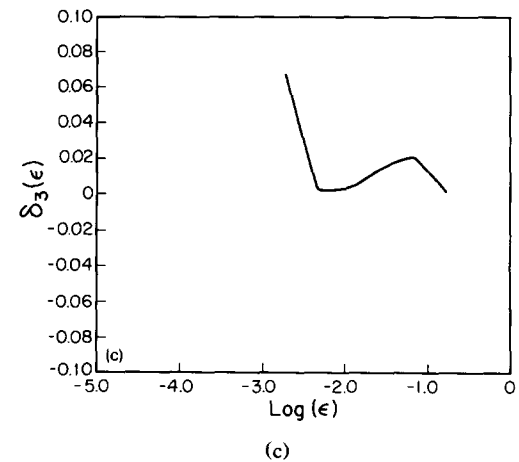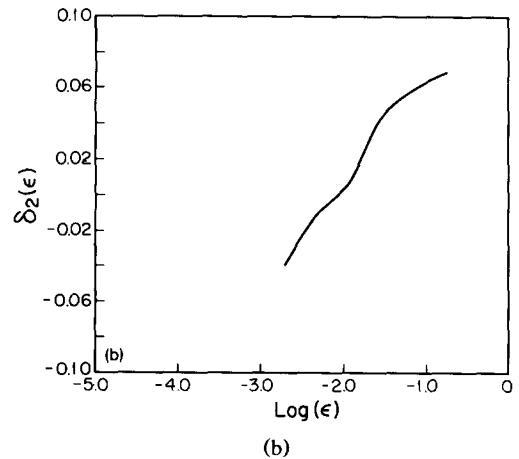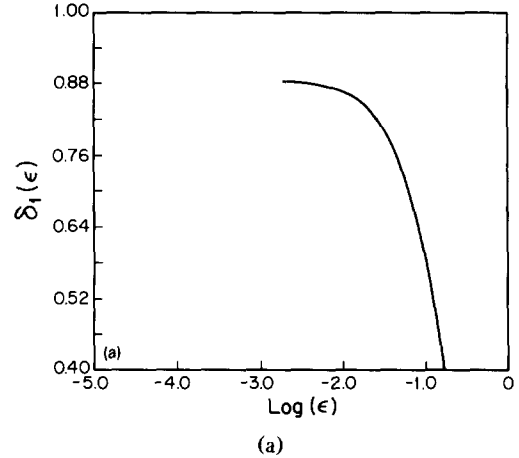

Fig. 4. (a) $\delta_1(\varepsilon)$ for the logistic map plus uniform noise between -0.5 and 0.5 of strength $\sigma = 0.1$, (b) $\delta_2(\varepsilon)$ for the logistic map plus uniform noise between -0.5 and 0.5 of strength $\sigma = 0.1$. (c) $\delta_3(\varepsilon)$ for the logistic map plus uniform noise between -0.5 and 0.5 of strength $\sigma = 0.1$.

consistent with zero.) This effect and the presence of dependencies induced by the nonuniform slope of the quadratic map have important consequences for the interpretation of our results. It is clear that the $\delta_j(\varepsilon)$ reveal not only information about the explicit dependencies in the underlying deterministic map but also information about those induced dependencies which lead to greater reproducibility of subsequences of the time series. In addition, different dependencies are indicated at different values of ε, so we conclude that the $\delta_j(\varepsilon)$ provide information about the reproducibility of a subsequence based on dependent variables *relative to the tolerance with which one examines the data*.

5.3. The Hénon map

Figs. 5a–5d show the $\delta_j(\varepsilon)$, $j = 1,\ldots,4$ for the Hénon map. Note that as $\varepsilon \to 0$, $\delta_1(\varepsilon)$ approaches one, $\delta_2(\varepsilon)$ appears to approach one or a value near one, and $\delta_3(\varepsilon)$ and $\delta_4(\varepsilon)$ are both sensibly zero. For the Hénon map $\nu = 1.22$ so $\alpha_0 = -1.22$ (see footnote 4). The Hénon map has a continuous natural measure so $\alpha_1 + \nu = 1$ as before, but now $\alpha_1 = -0.22$. At the level of our statistics α_d
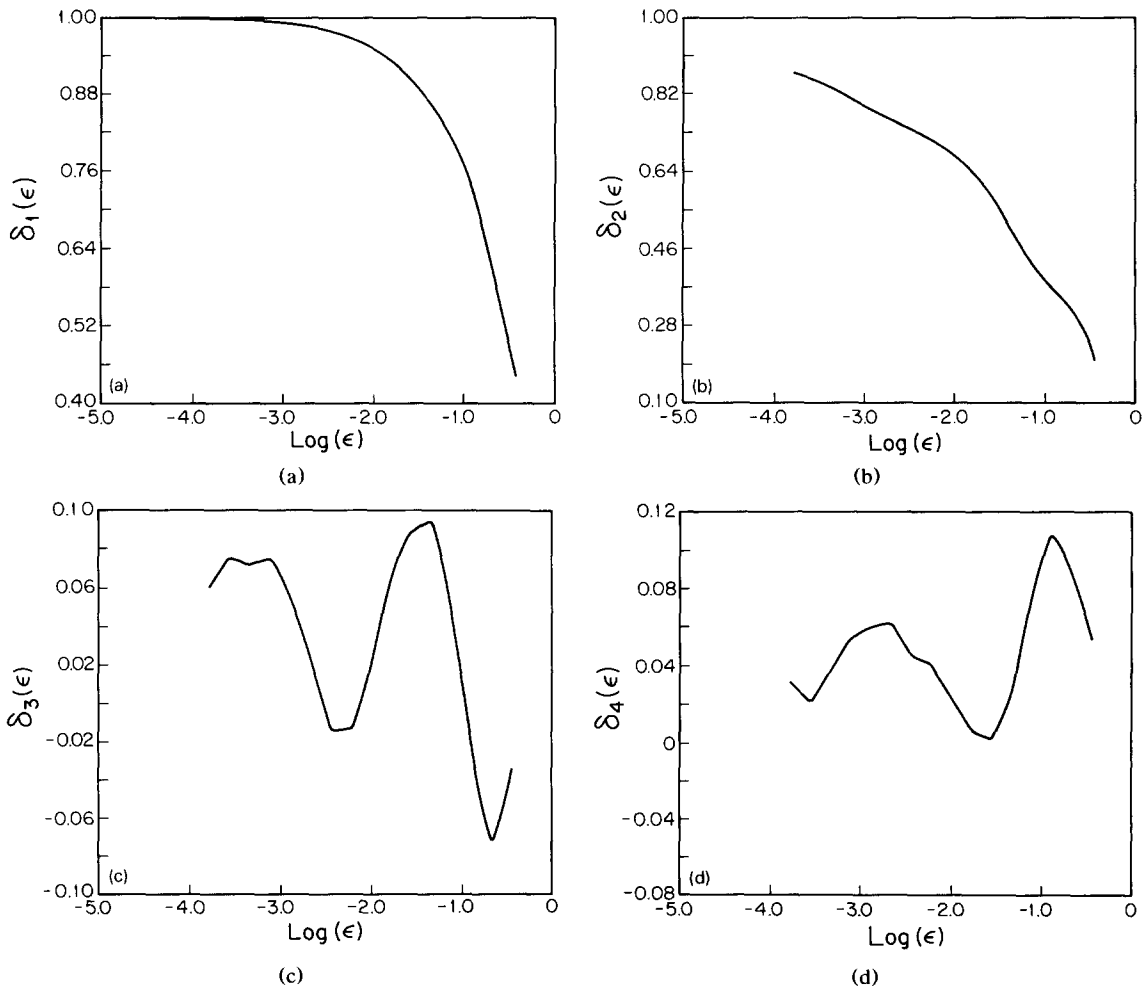

Fig. 5. (a) $\delta_1(\varepsilon)$ fo the Hénon map. (b) $\delta_2(\varepsilon)$ for the Hénon map. (c) $\delta_3(\varepsilon)$ for the Hénon map. (d) $\delta_4(\varepsilon)$ for the Hénon map.

is sensibly zero for all $d \geq 2$, and so from (45) $j_0 = 3$. One can therefore see that $D_2(\alpha_1) = -0.78 < 0$, so from (42) $\lim_{\varepsilon \to 0} \delta_1(\varepsilon) = 1$. In addition, (46) implies that $\lim_{\varepsilon \to 0} \delta_2(\varepsilon) = 1$, and $\delta_j(\varepsilon) < 1$ when $\varepsilon < 1$ for $j \geq 3$. The Hénon map does have nonuniform slope, but its effect is evidently not as pronounced as in the logistic map. Hence $D_2(\gamma_j(\varepsilon))$ rapidly approaches zero and $\delta_3(\varepsilon)$ and $\delta_4(\varepsilon)$ are both small.

The fact that $\delta_1(\varepsilon)$ approaches 1 for the tent and logistic maps and $\delta_1(\varepsilon)$ and $\delta_2(\varepsilon)$ both appear to approach 1 for the Hénon map as $\varepsilon \to 0$ raises the conjecture that this approach to 1 is an indicator of explicit dependence in noiseless deterministic maps. So far we have found no proof of this conjecture. In any case, it is primarily of only theoretical interest since noiseless deterministic maps are unlikely to be encountered in an experimental situation, and one rarely has the opportunity to observe phenomena with infinite precision.

6. Summary and discussion

In this paper we have presented a new approach to the study of time series, and in particular to the study of time series with broad-band spectra. In an effort to determine its efficacy, we have analyzed several one- and two-dimensional maps using our method. We have shown that the method is easily capable of identifying the non-trivial dependencies in chaotic time series generated by such maps, even in the presence of substantial IID noise. We also introduced a predictability index which measures the extent to which one can predict subsequent values of a time series given the history of the series. Finally, we studied the ε-dependence of our methods, especially the theoretically interesting case in which ε goes to zero. Based on these results, we believe that the approach described in this paper will prove to be very useful in analyzing series in which one expects to find some underlying intrinsic deterministic process.

We have used relatively simple low-dimensional systems to introduce our methods. One very useful aspect of these techniques is that they are also applicable to time series generated by higher dimensional systems. So long as new structure (in the sense of conditional probabilities) emerges in higher dimensions the deltas will continue to be statistically nonzero. As a result, these methods are more systematic, robust and transparently applicable than some other approaches such as the construction of phase portraits. This statement is also supported by our work on higher-dimensional continuous time series reported in ref. [8].

The techniques presented in this paper represent a significant departure from usual methods of time series analysis based on linear transforms. Such linear methods are next to useless for series with broad-band spectra. One way of viewing our techniques, is that the values of the deltas can be considered tests against a null-hypothesis of IID with high power for other alternative explanations based on the existence of conditional probabalistic dependence. Following this approach, one should, in practice, compare the values of the deltas obtained from an analysis of some time series with values obtained from a good quasi-random sequence of the same length and with a similar probability distribution. For asymptotically long time series, it turns out that such comparisons can be put on a statistically more precise mathematical footing. This work is in progress [9].

From another point of view, our work provides an alternate method for capturing the structure of an attractor given the tolerance, ε, with which one examines the data. Given an ε, one can determine which deltas are statistically nonzero, and use the number of such deltas as an estimate of an appropriate embedding dimension. However, this estimate may be larger or smaller than the true asymptotic embedding dimension. In the case of the logistic map, for example, δ_2 is nonzero when ε is one-half the standard deviation, even though it approaches zero in the small ε limit. But from an empirical point of view,

determining which deltas are statistically nonzero may be more useful than trying to determine the minimum embedding dimension of the attractor. It is also less computationally intensive: Using the methods of Grassberger and Procaccia [1], the calculation of the correlation dimension from time series data requires the calculation of the correlation integrals for a range of ε as ε approaches zero in order to elucidate a scaling region. The necessity of calculating the correlation integrals for small ε greatly increases the number of data points required to achieve accurate results and the calculation of correlation integrals for a range of ε in the scaling region is more time consuming than the calculation of results at a single value of ε. In addition, the dimension of the attractor is related to the underlying dynamics through the theorem of Takens, which states that a sufficient number of embedding dimensions to reconstruct a strange attractor is less than or equal to twice its Haussdorf dimension plus one. In many cases, this inequality is not saturated. One may well be interested only in an understanding (or empirical reconstruction) of the dynamics for a certain value of the tolerance. In such a case, knowing which deltas are nonzero for a specific value of ε, may provide more directly usable empirical information, and provide it more economically than attempting to determine the theoretical embedding dimension and reconstruct the attractor in detail. Of course, we do not mean to imply that the embedding dimension and detailed structure of an attractor is uninteresting, but our approach provides complementary, albeit related information.

Our work raises a number of other interesting questions. One obvious one concerns generalizations of this study to continuous time series. The methods presented here seem well suited to systems in which there may be an underlying discrete deterministic process, but their application to continuous time series requires significant generalization. In another paper [8], we report on the development of a generalized method applicable to continuous time series and illustrate its use in a wide variety of numerical examples, including the Lorentz model, the Hénon–Heiles equations, the forced Brussels oscillator, and the Mackey–Glass equation. Our generalized methods work well at picking out empirical dependencies in these more complicated and higher dimensional examples. The application of these statistics to the analysis of experimental data including data from chaotic chemical reactions, turbulent fluid flow, and radioactive decay is in progress.

Note added

After this work was completed, we became aware of the paper by Cenys and Pyragas [10], in which some of the ideas contained in the present work appear in a somewhat different form.

Acknowledgements

Appendix

We elaborate here our choice of the form (35) to express the corrections to the leading behavior of $C_d(\varepsilon)$ in the limits $\varepsilon \to 0$ and $d \to \infty$. First we define a general correction term $\xi_d(\varepsilon)$ such that

$$C_d(\varepsilon) \equiv \varepsilon^\nu \, e^{-dK_2} \, e^{\xi_d(\varepsilon)}. \tag{A.1}$$

We will apply, in proper order, the four limits in (33) and (34) to (A.1) and we will show that the existence of these limits[#6] restricts the asymptotic behavior of $\xi_d(\varepsilon)$. For future convenience, we

---

[#6] See ref. [9] for a more detailed discussion of the conditions required for the existence of these limits. Also see the discussion of the Zaslaskii map in ref. [1].

will also separate $\xi_d(\varepsilon)$ into two first-order correction terms and a third term containing all higher-order corrections.

Applying the first limit in (33) to (A.1) we see that

$$\lim_{\varepsilon \to 0} \frac{\ln C_d(\varepsilon)}{\ln \varepsilon} = \nu + \lim_{\varepsilon \to 0} \frac{\xi_d(\varepsilon)}{\ln \varepsilon}, \tag{A.2}$$

and so we identify the first-order correction

$$\alpha_d \equiv \lim_{\varepsilon \to 0} \frac{\xi_d(\varepsilon)}{\ln \varepsilon}. \tag{A.3}$$

Applying the second limit in (33) we derive a restriction to the asymptotic behavior of α_d because

$$\nu = \lim_{d \to \infty} \left(\lim_{\varepsilon \to 0} \frac{\ln C_d(\varepsilon)}{\ln \varepsilon} \right)$$
$$= \nu + \lim_{d \to \infty} \alpha_d \Rightarrow \lim_{d \to \infty} \alpha_d = 0. \tag{A.4}$$

Now we define a new correction term $\eta_d(\varepsilon)$ such that $\xi_d(\varepsilon) \equiv \alpha_d \ln \varepsilon + \eta_d(\varepsilon)$. Then (A.3) implies that

$$\lim_{\varepsilon \to 0} \frac{\eta_d(\varepsilon)}{\ln \varepsilon} = 0. \tag{A.5}$$

Consider now the limits in (34). From (A.1) we see that

$$\lim_{d \to \infty} \frac{\ln C_d(\varepsilon)}{-d} = K_2 + \lim_{d \to \infty} \frac{\xi_d(\varepsilon)}{-d}, \tag{A.6}$$

and we identify a new first-order correction

$$\beta(\varepsilon) \equiv \lim_{d \to \infty} \frac{\xi_d(\varepsilon)}{-d} = \lim_{d \to \infty} \frac{\alpha_d \ln \varepsilon + \eta_d(\varepsilon)}{-d}$$
$$= \lim_{d \to \infty} \frac{\eta_d(\varepsilon)}{-d}, \tag{A.7}$$

where the last equality follows from (A.4). Applying the second limit in (34) we derive a restriction

to the limiting behavior of $\beta(\varepsilon)$ because

$$K_2 = \lim_{\varepsilon \to 0} \left(\lim_{d \to \infty} \frac{\ln C_d(\varepsilon)}{-d} \right)$$
$$= K_2 + \lim_{\varepsilon \to 0} \beta(\varepsilon) \Rightarrow \lim_{\varepsilon \to 0} \beta(\varepsilon) = 0. \tag{A.8}$$

Finally, we define a correction $\gamma_d(\varepsilon)$ such that $\eta_d(\varepsilon) = -d\beta(\varepsilon) + \gamma_d(\varepsilon)$, which means that

$$\xi_d(\varepsilon) = \alpha_d \ln \varepsilon - d\beta(\varepsilon) + \gamma_d(\varepsilon). \tag{A.9}$$

Using (A.3), (A.5) and (A.8), we find

$$\lim_{\varepsilon \to 0} \frac{\gamma_d(\varepsilon)}{\ln \varepsilon} = 0, \tag{A.10}$$

and further (A.7) implies that

$$\lim_{d \to \infty} \frac{\gamma_d(\varepsilon)}{-d} = 0. \tag{A.11}$$

The results (A.8)–(A.11), and (A.4) establish that any $C_d(\varepsilon)$ satisfying (33) and (34) can be written as

$$C_d(\varepsilon) = \varepsilon^{\nu + \alpha_d} e^{-d(K_2 + \beta(\varepsilon))} e^{\gamma_d(\varepsilon)}, \tag{35}$$

with

$$\lim_{d \to \infty} \alpha_d = \lim_{\varepsilon \to 0} \beta(\varepsilon) = 0$$

and

$$\lim_{d \to \infty} \frac{\gamma_d(\varepsilon)}{-d} = \lim_{\varepsilon \to 0} \frac{\gamma_d(\varepsilon)}{\ln \varepsilon} = 0$$

and without loss of generality.

References

[1] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, Physica D 9 (1983) 189–208.
[2] W.A. Brock, W.D. Dechert, J.A. Scheinkman and B. LeBaron, A test for independence based on the correlation dimension, University of Wisconsin preprint (1988).
[3] J. Scheinkman and B. LeBaron, J. Business (Chicago) 62 (3) (1989) 311;
D. Hsieh, J. Business (Chicago) 62 (3) (1989) 339;
W. Brock, J. Economic Theory 40/168
[4] R. Savit, unpublished.

[5] A.N. Kolmogorov, Foundations of the Theory of Probability (translated by N. Morrison) (Chelsea Publ. Co., New York, 1956) pp. 12–13, 47–56.

[6] H. Sakai and H. Tokumaru, IEEE Trans. Acouts. Speech Signal Process, V.I. ASSP-28 (1980) 588.

[7] F. Takens, Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, Warwick 1980, eds.

D. Rand and B.S. Young, Lecture Notes in Mathematics 898 (Springer, Berlin, 1981)

[8] M. Green and R. Savit, Dependent variables in broad band continuous time series, University of Michigan preprint (1990).

[9] W. Brock, B. LeBaron and R. Savit, in preparation.

[10] A. Cenys and K. Pyragas, Phys. Lett. A 129 (1988) 227.