Nonlinear Processes in Geophysics (1994) 1: 12 - 25

Nonlinear Processes in Geophysics

Searching for determinism in observed data: a review of the issues involved

A.A. Tsonis[1], G.N. Triantafyllou[1] and J.B. Elsner[2]

[1] Dept. of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA
[2] Dept. of Meteorology, Florida State University, Tallahassee, FL 32306, USA

Received 1 December 1993 - Accepted 16 February 1994 - Communicated by A.D. Kirwan

Abstract. In this paper we present a review of advances made and problems still existing in the application of the theory of chaos and dynamical systems to time series. In particular we discuss issues pertaining the estimation of dimensions, Lyapunov exponents and nonlinear prediction from an observable. We analyze the problems and discuss proper ways to deal with them.

1 Introduction

Lately, ideas from the theory of nonlinear dynamical systems and chaos have been applied to many problems from many different disciplines. The main goal is the search for low-dimensional chaos and the extraction of the properties of the underlying attractors, if any. The procedure often involves one observable (time series) and a reconstruction of the attractor. The reconstruction is achieved by taking a scalar time series $x(t_i)$ and its successive time shifts (delays) as coordinates of a vector time series given by:

$$\mathbf{X}(\mathbf{t_i}) = \{\mathbf{x}(\mathbf{t_i}), \mathbf{x}(\mathbf{t_i} + \tau), \ldots, \mathbf{x}(\mathbf{t_i} + (\mathbf{n} - \mathbf{1})\tau\} \quad (1)$$

where n is the dimension of the vector $\mathbf{X}(\mathbf{t_i})$ (often referred to as the embedding dimension) and τ is an appropriate delay (Packard et al. 1980; Ruelle 1981; Takens 1981). For proper reconstructions the embedding dimension n should be equal or greater to $2D + 1$, where D is the dimension of the manifold containing the attractor. Such an embedding preserves the topological properties of the attractor. More specifically the embedding will be a diffeomorphism, a differentiable mapping with differentiable inverse from the true phase space to the delay space. This is Whitney's theorem and, strictly speaking, is valid only when we have an infinite and dense set in our disposal. When we only have a limited dataset the theorem may not be valid. In fact, in those cases the word embedding is used loosely as any topologist will argue.

2 The first problem

Even before we begin to discuss methods to extract properties of the attractors we are faced with our first problem. What is a proper delay parameter τ in equation (1)?

When we reconstruct the attractor by producing a cloud of points at a given embedding dimension, those points should be independent of each other. If they are not the correlation dimension could be underestimated (Figure 1). Therefore, τ must be chosen so as to result in points that are not correlated to previously generated points. Thus, a first choice of τ should be in in terms of the decorrelation time of the time series under investigation. The question now arises: How do we define

Correspondence to: A.A. Tsonis

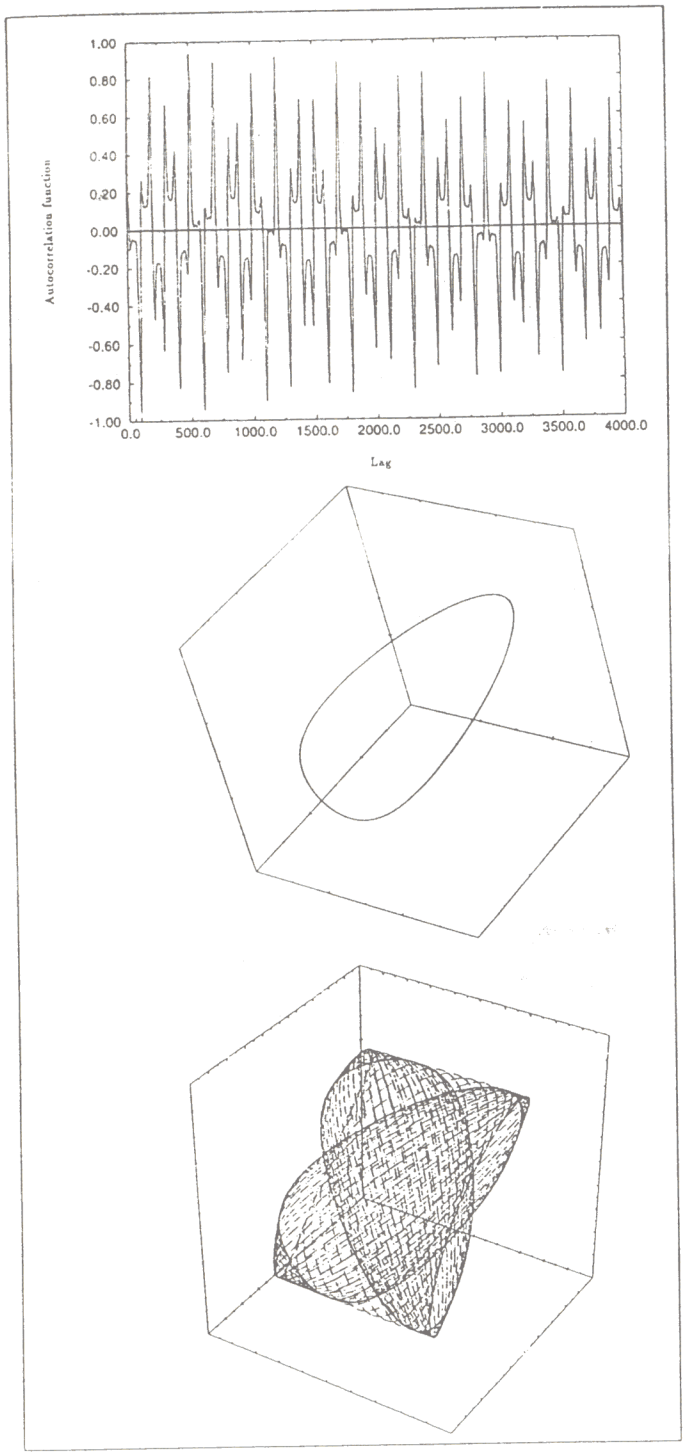

Fig. 1. a) Autocorrelation function of time series obtained from a dynamical system that has a 2-torus attractor.
b) Reconstruction in an embedding dimension three using $\tau = 100$. Note this τ corresponds to highly correlated values and as a consequence we obtain a limit cycle.
c) Reconstruction using $\tau = 50$ which corresponds to virtually uncorrelated values. Now we obtain a drastically different picture closer to the true torus.

the decorrelation time? A straightforward procedure is to consider the decorrelation time equal to the lag at which the autocorrelation function for the first time attains the value of zero. Other approaches consider the lag at which the autocorrelation function attains a certain value like $1/e$, 0.5, or 0.1 (Tsonis and Elsner 1988). Another suggestion for the choice is to take τ equal to T/n where T is the dominant periodicity (as revealed by Fourier analysis) and n is the embedding dimension. In this way τ gives some measure of statistical independence of the data average over an orbit and it is an appropriate approach if the autocorrelation function is periodic. As it was pointed out, however, by Frazer and Swinney (1986) the autocorrelation function measures the linear dependence among successive points and may not be appropriate when we are dealing with nonlinear dynamics. They argue that what should be used as τ is the local minimum of the mutual information that measures the general dependence among successive points. Evidently, no one of the aforementioned rules has emerged as the undisputed rule for choosing τ, but the mutual information approach appears to have the edge. Nevertheless, a very reassuring practice is experimenting with various τ's (while repeating the aforementioned constraints) in order to address possible effects of the choice of τ.

Once a delay τ has been chosen, the characterization of a dynamical system commonly includes estimation of the various dimensions, estimation of Lyapunov exponents and nonlinear prediction. As Figure 2 depicts there are several problems for each and every one of those procedures. Below we discuss and we propose ways to deal with those problems.

3 Estimating Dimensions

3.1 Existence of scaling

For an n-dimensional phase space, a "cloud" or a set of points will be observed. From this set the various dimensions and exponents that characterize the underlying attractor can be calculated. The most popular approach is to calculate the cor-

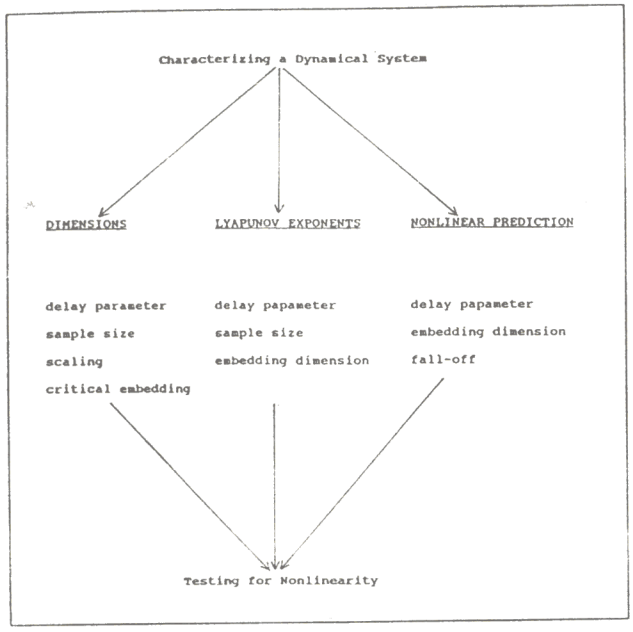

Fig. 2. Approaches used in search for determinism in observed data and the associated problems.

relation dimension. According to this approach (Grassberger and Procaccia 1983a,b), given the cloud of points one finds the number of pairs $N(r,n)$ with distances less than a distance r. In this case, if for significantly small r, we find that:

$$N(r,n) \propto r^{d_2} \qquad (2)$$

then the scaling exponent d_2 is the correlation dimension of the attractor for that n. Since the dimension of the underlying attractor is not known, we test the power law of equation (2) for increasing values of n and check for a saturation value D_2, which will be an estimation of the correlation dimension of the attractor.

The scaling exponent d_2 is often found via a least square fit over some range of scale where the $\log N$ vs $\log r$ plot appears linear. However, $\log-\log$ plots have a tendency to display "linear" regions even when they are completely nonlinear. As a result, estimating d_2 (and then D_2) as such, may not be very reliable. Figure 3 demonstrates the above point by showing the logarithm of the root mean square fluctuation $F(l)$ of some walk as a function of the logarithm of displacement l. According to the theory of random walks a power law $F(l) \propto l^a$, where a is the scaling exponent that characterizes

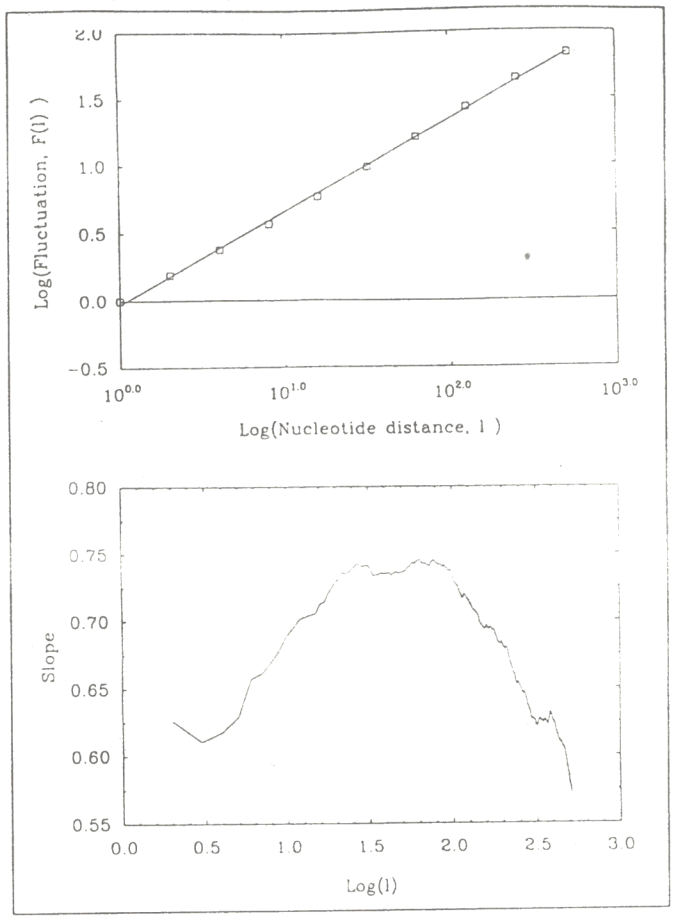

Fig. 3. a) $\log F(l)$ versus $\log l$ for some walk. The graph displays an apparent linear region over which a linear regression provides $F(l) \propto l^a$ with $a = 0.67$.
b) The first derivative (or local slope) of the data as a function of $\log l$ reveals that the graph in plot 3a is indeed nonlinear and therefore a scaling does not exist.

the random motion (see Tsonis et al. 1993a). In the $\log-\log$ plot an apparent (and tempting?) linear region with a slope of 0.67 exists over a wide range of scales which will indicate that the walk is indeed a *random* walk with $a \approx 0.67$. Figure 3b shows the local slope or more formally the first derivative of $\log F(l)$ with respect $\log l$. If scaling in figure such as 3a exists, then it follows that a plateau in figures such as 3b would be observed at $d\log F(l)/d\log l \approx a$. As can be seen from Fig. 3b no such a plateau can be justified, therefore the apparent scaling in Fig. 3a does not represent a real scaling, which means that in such cases a calculation of the exponent a is *meaningless!* There-

fore, the estimation of exponents that characterize power laws should not be found by "looking" at log – log plots, but their existence should be investigated via plots of slope versus log(scale). Unfortunately it is still the practice to estimate such exponents (and therefore dimensions) using the least squares approach and therefore many of the reported cases of scaling might not be accurate.

3.2 Critical embedding dimension

For a finite data set because of Equation (2) the population of pairs of points on smaller scales is smaller than the population of pairs on longer scales. Thus, if for a fixed n the number of points in the set becomes smaller, the population of pairs over the scales for which Equation (2) holds begins to be depleted. As we continue to decrease the number of points (for our fixed n always) we will observe:

1. more and more depletion at smaller scales (since less and less points will be found) and
2. large fluctuations of $N(r,n)$ due to small populations (and thus inadequate statistics) at slightly larger scales.

The net result is that the scaling region may be completely masked. Any straight line fitting at this point will result in a false correlation dimension for that n. Thus for an accurate estimation of the slope d_2 on a $\log N(r,n)$ versus $\log r$ plot requires a minimum number of points.

It should now be emphasized that by embedding the dataset into continually higher dimensions we effectively "distribute" the same number of points into continually higher dimensional space. In effect we go from a densely populated low-dimensional space to a sparsely occupied high-dimensional space. The result will be the same as before: for very small scales $N(r,n)$ goes to zero (depopulation) and for large scales (scales close to the radius of the set) $N(r,n) = N(N-1)/2$ independently of r and n (saturation). Thus, at some embedding dimension the scaling region will not be clearly defined as it will be "lost" between depopulation and saturation. The embedding dimension above which the scaling region cannot be accurately defined is called the critical embedding dimension n_c. The above points have been emphasized by Essex et al. (1987) and Tsonis and Elsner (1990).

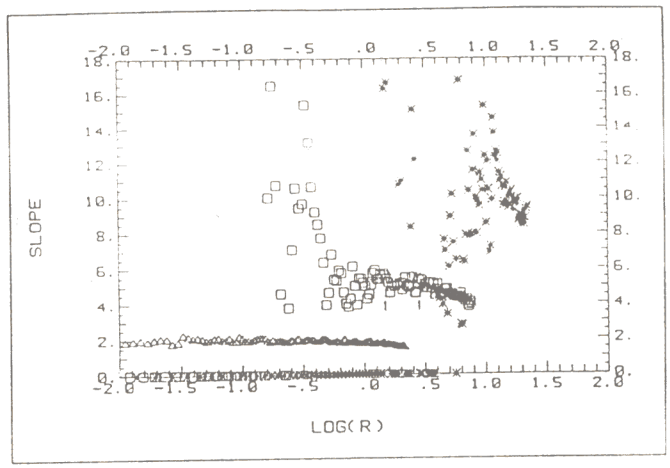

Fig. 4. $\Delta \log N(r,n)/\Delta \log r$ as a function of $\log r$ for embedding dimension 2 ($\triangle$), 6 ($\square$), and 15 ($*$). Based on a record of 500 white noise values. See text for details.

We now wish to demonstrate the above by starting with a time series of 500 white noise values. We know that in this case for any embedding dimension $d_2 = n$ (as long as we use the necessary number of points for each n). We started with embedding dimension $n = 2$ and found $\log N(r,n)$ as a function of $\log r$ and then we calculated

$$slope = \Delta \log N(r,n)/\Delta \log r$$

as a function of $\log r$. If there exists a clearly defined scaling region in the $\log N(r,n)$ versus $\log r$ plots then we should be able, on a slope versus $\log r$ plot, to observe a plateau. This plateau will provide an estimation for the exponent d_2 for a given n. Figure 4 shows slope versus $\log r$ for $n = 2(\triangle)$, $n = 6(\square)$, and $n = 15(*)$. For $n = 2$ we observe that slope is nearly constant at about a value of two for a wide range of scales. When the scales become too large, saturation indicated by a gradual decrease of slope is seen. Depopulation is not visible (at least within the scale range of the figure). Therefore we may conclude that 500 points are adequate in defining the scaling region when $n = 2$. For $n = 6$ we observe a very different picture. We see depopulation manifesting itself as many zero slope values over small scales, large fluctuations

over larger scales, and saturation over very large scales. A scaling region can be suggested (indicated by the arrows) but it is not as clearly defined or as wide as in the case of $n = 2$. Nevertheless this small plateau is found at about slope = 5.0, which is less than the true value of 6.0. Thus an attempt here to define a scaling region will at best result in an underestimation of the true value of the exponent d_2. Similar comments can be made for $n = 15$; here the difference shows that there is virtually no way one can define a scaling region.

Figure 5 is similar to Fig. 4 but for $n = 4$. Similar comments to those made in Fig. 4 for n greater or equal to six can be made for Fig. 5. A scaling region (indicated by the arrows) may be identified but it will produce a value of $d_2 = 3.5$ (which is less than the true value of 4.0). However, in Fig. 5b (with 5000 points), a well-defined scaling region exists with a value of $d_2 = 4.0$. It is important to know that the true scaling region here is *at smaller* scales compared to the scaling region that is identified when 500 points are used where the true scaling region is masked by large fluctuations.

3.3 Sample size requirements

The above clearly demonstrate what was discussed previously. We should be very careful not to exceed the critical embedding dimension that is a function of the data size. This unavoidably brings us to the following question: What is a sufficient number of points N for a given embedding dimension?

This problem can be approached by assuming that in all embedding dimensions n less than the dimension of the object in question, the object is space filling like uniformly distributed random numbers in the interval [0,1]. The painful exercise of determining the minimum number of points was first tackled by Smith (1988), who concluded that this number is equal to 42^m, where m is the smallest integer above the dimension of the object, which under the aforementioned assumption is the dimension in which the random numbers are embedded in. Thus, for $m = 4$, if N is not at least equal to 3 111 696 no accurate estimate of d_2 can be obtained. But as we saw in Fig. 5 ac-

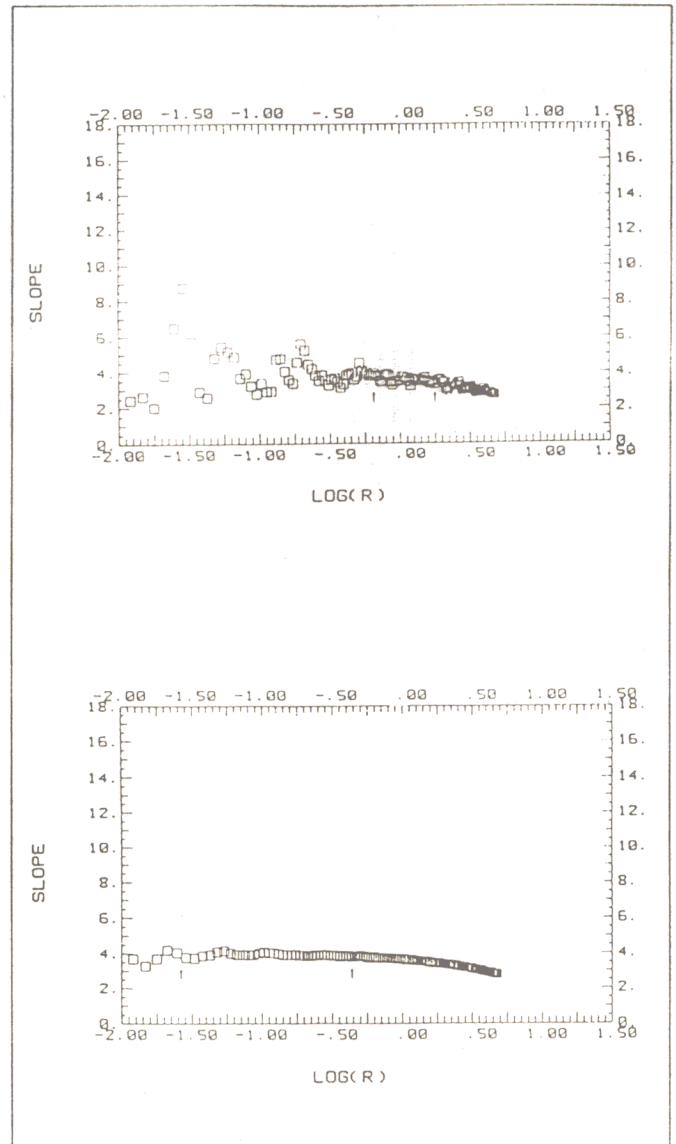

Fig. 5. a) $\Delta \log N(r,n)/\Delta \log r$ as a function of $\log r$ for embedding dimension 4.
b) As in Fig. 5a but for a record of 5000 white noise values.

curate estimates for $m = 4$ can be obtained with as little as 5,000 points. This number is significantly lower than 3 111 696. Why this great discrepancy? The only explanation is that the 42^m conclusion is in error. In fact, it has been recently demonstrated (Nerenberg and Essex 1990) that the data requirements are not as nearly as extreme. In fact, the minimum number of points N_{min} required to produce no more than an error A (typically $A = 0.05n$) is approximately (see Tsonis et al. 1993b)

$$N_{min} \propto 10^{2+0.4n}.$$

Thus, for $n = 4, N_{min} \sim 10^{3.6} \sim 4000$ points as mentioned above.

The aforementioned discussion unavoidably brings us to the next very important point. Because of the underlying assumptions all the theoretical calculations and derivations of the necessary number of points as a function of embedding dimension presented up to this point are valid only as long as the embedding dimension n is less than correlation dimension D_2. In addition, it is quite possible that estimates of the number of points would depend on the type of the attractor (nonuniform, fractal); as issue that has not yet been addressed in those calculations. Moreover, it is not certain whether or not (and especially for fractal sets) the need for data increases at the same rate with embedding dimension when $n > D_2$. Experimentation with known dynamical systems indicates that in the case even though the need for data may increase it may not be as severe as it is predicted by the aforementioned formulas! For example, Fig 6a shows slope versus $\log r$ plots of a sample size 2000 for an observable from the Hènon map for embedding dimensions 2, 4, 6, 8, 10, and 20. We observe that for $n = 2$, $d_2 = 1.25$ (which is the Hausdorff-Besicovitch dimension of the Hènon map). Subsequently, we observe that up to embedding dimension 10, fairly accurate estimates of the correlation dimension can be obtained. In fact, we do not observe too much fluctuation or an underestimation of d_2 with increasing embedding dimension. Of course, eventually (i.e., for $n \gg D_2$) fluctuations may mask any scaling region. The differences with Fig. 5b which show the same but for a sample of 2000 random numbers are clear and provide some

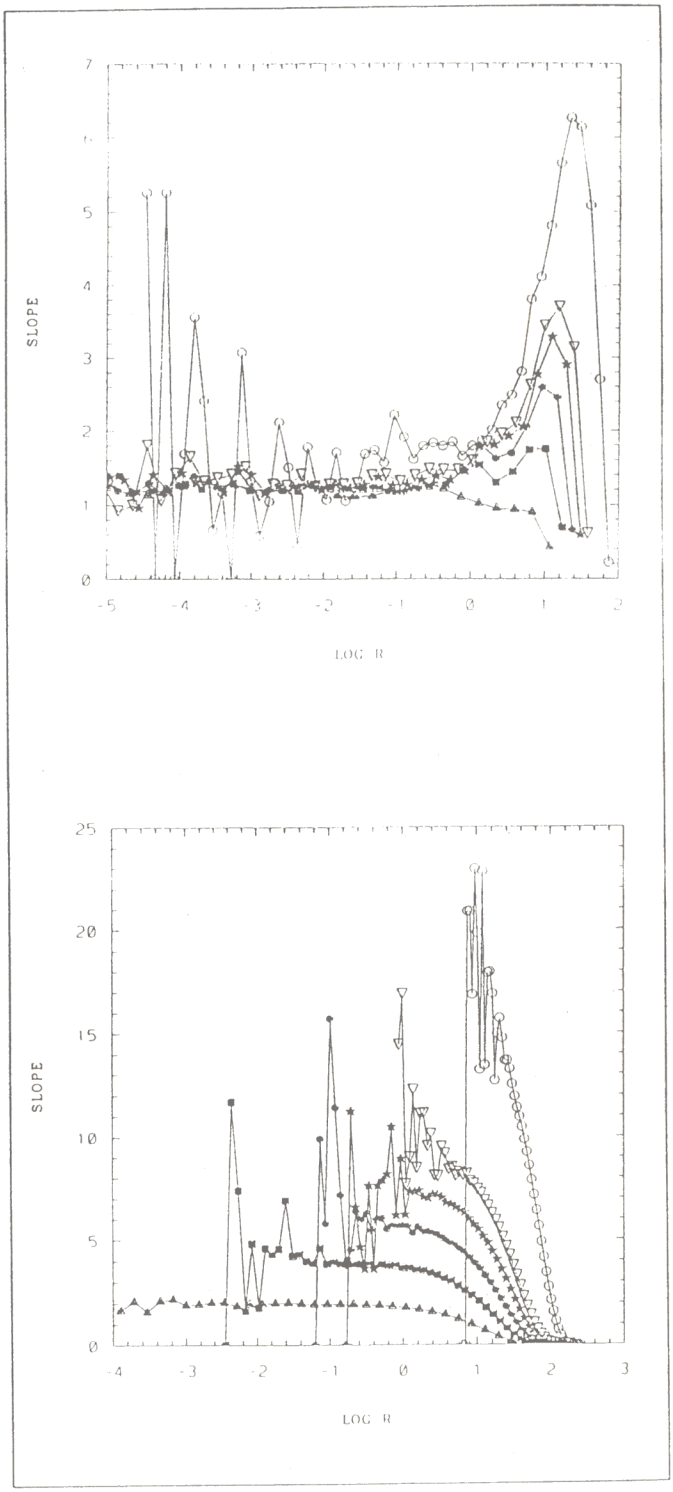

Fig. 6. a) Local slope as a function of $\log r$ for the Hènon map and for embedding dimensions 2 (full △), 4 (full □), 6 (•), 8 (⋆), 10 (△), 20 (○). Based on 2000 values.
b) As above but for a sample of uniformally distributed random numbers in $[0, 1]$.

evidence that for fractal sets the need for data may increase at a much slower rate for embedding dimensions higher than the correlation dimension of the attractor. We should point out, however, that such a conclusion may not be valid for every dynamical system or dataset. Do we, therefore, in cases where saturation is observed at an embedding dimension $n_* > D_2$, need $N \sim 10^{2+0.4n_*}$ points or just $N \sim 10^{2+0.4D_2}$ points (again provided that n_* is not much higher than D_2)? This is an open question and of course a conservative choice would be to consider the requirements for n_*.

3.4 Power-law spectra

In power law spectra the energy is high at low frequencies (large scales) and low at high frequencies (small scales). Since in nature energy is distributed from the large scales to the smaller scales it is expected that spectra of observables from nature would exhibit something like a power law spectra. Osborne et al. (1986) and Osborne and Provenzale (1989) have argued that a certain class of *random* sequences would exhibit a finite correlation dimension. This class of random sequences includes self-affine sequences that exhibit a power law spectra of the form $P(f_k) = Cf_k^{-a}$ for $1 < a < 3$ and are commonly referred to as fractional Brownian motions (fBms), a type of colored noise. In this case the *trail* of n-independent realizations (each one representing one phase space dimension) is self-similar with a theoretically predicted fractal dimension equal to $2/(a-1)$. When the Grassberger-Procaccia algorithm is applied to trails or to trajectories reconstructed from a single sequence via the method of delays, a finite correlation dimension (close to the fractal dimension of the trail) is obtained. Thus they suggested that a finite value for the correlation dimension cannot be indicative of a dynamical system with a finite number of degrees of freedom. It can only indicate a lower bound for the actual number of degrees of freedom, which might be infinite! Thus if the data in question exhibit power-law spectra of the form $P(f) \propto f^{-a}$ and the estimated correlation dimension is approximately equal to $2/(a-1)$ then certain tests must be contucted to make sure that we are looking at a deterministic process and not at some fBms.

One of the tests is to investigate the properties of the autocorrelation function of the data. For fBms for $N \to \infty$ the autocorrelation function scales as $C(\xi) \propto \xi^{1-a}$ where ξ is the lag (i.e., $C(\xi)$ never reaches the value of zero). In addition for finite N the decorrelation time (defined as the lag at which $C(\xi)$ attains some small value close or equal to zero) is a function of N. One, therefore, can check the decorrelation time of the original time series for gradually increasing N. If the decorrelation time does not keep on increasing then one can rule out fBms. Another test is a stationarity test. fBms are nonstationary processes while chaotic processes are stationary. If over the available data length the time series is stationary one again can rule out fBms. If over the available data the process is nonstationary both the above tests may fail. In this case an alternative test can be provided by nonlinear prediction (see the corresponding section below).

Before closing this section a point must be made clear. Because of their "long" memory it is very difficult to obtain points from an fBm that are truly independent. Therefore a correct application of the Grassberger-Procaccia algorithm requires extremely high N and in this case one finds that $D = n$ for any n as expected from stochastic processes (Theiler, 1991). For finite N's, however, the reported by Osborne and Provenzale "anomalous" scaling is observed. The importance of Osborne and Provenzale's observation lies in the fact that for a finite N (which is usually the case) a natural process might mimic an fBm, and thus a finite correlation dimension should be subjected to the tests outlined above.

4 Estimating Lyapunov exponents

The Lyapunov exponents measure the rate at which nearby trajectories in phase space diverge or converge. Positive Lyapunov exponents indicate divergence and therefore chaos. In theory the Lyapunov exponents λ_i are defined according to:

$$\lambda_i = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt \frac{d}{dt} \ln \left[\frac{p_i(t)}{p_i(0)} \right]$$

$$= \lim_{T\to\infty} \frac{1}{T} \ln \left[\frac{p_i(T)}{p_i(0)}\right] \quad (3)$$

Here $p_i(0)$ is the radius of the principal axis p_i at $t = 0$ of an initial hypersphere of dimension n and $p_i(T)$ is its radius after a long time T. The dimension n is the dimension of the Euclidean phase space in which the attractor is embedded. There are many Lyapunov exponents as the dimension of the phase space.

The estimation of the Lyapunov exponents from a system of ordinary differential equations is straight forward and it is based on the fact that the linearized equations which describe the local dynamics involve a Jacobian whose eigenvalues provide all the exponents (see Tsonis, 1992). In practice at first we embed the data in some space whose dimension is sufficient (normally $d_e > 2D$). Then the reconstructed phase space provides the information to estimate the Jacobian by monitoring the motion in space of selected points and their neighborhoods (Abarbanel and Kennel, 1991).

Because the linearized equation provide the local dynamics such an approach provides an estimation of the local Lyapunov exponents. Repeating the procedure for many points we can obtain an average picture which will be related to the Lyapunov exponents of the system. Note that, estimation of the Lyapunov exponents can provide an estimation of the dimension of the system. According to Kaplan-Yorke conjecture (Frederickson et al., 1983) the information dimension of a chaotic attractor, D_I is equal to:

$$D_I = j + \frac{\sum_{i=1}^{j} \lambda_i}{|\lambda_{j+1}|}$$

where j is defined by the condition that $\sum_{i=1}^{j} \lambda_i > 0$ and $\sum_{i=1}^{j+1} \lambda_i < 0$.

Apparently, unlike estimating dimensions here we need to know a proper embedding dimension n where the calculations have to be performed. A proper embedding requires an a priori knowledge of of the dimension and therefore the estimation of Lyapunov exponents suffers from the problems associated with dimension estimation. We still need a proper delay τ, we still have to worry about adequate sample sizes and correct procedure to extract the values of the exponents from experimental data.

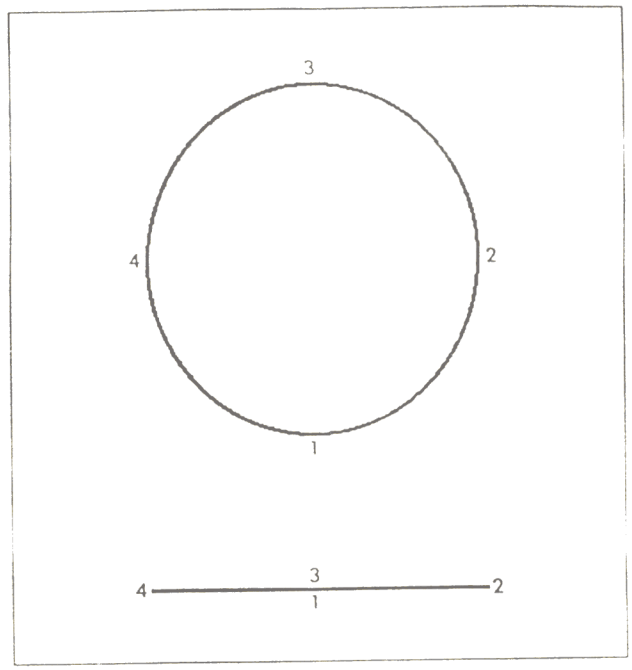

Fig. 7. An illustration of the principle behind estimating the embedding dimension using nearest neighbors (see text for details).

An alternative is provided by a new technique that defines a proper embedding dimension (Sugihara and May, 1990, Abarbanel et al. 1992). The basic philosophy is outlined in Figure 7. If the underlying attractor is assumed to be circle and we embed the motion in one dimension the points close to 1 and 3 will be close neighbors. We call these neighbors false neighbors because in the actual motion (i.e. in a two dimensional embedding space) are very far apart. Thus starting with $d = 1$ for each point in phase space we find the nearest neighbor. From all the available points we then find how many of those nearest neighbors remain nearest neighbors as we go to $d = 2, d = 3$ and so on. A proper embedding is defined as the dimension for which the percentage of false neighbors goes to zero. This is very promising technique and if the noise of the data is small good proper embedding dimensions can be obtained. Figure 8 shows the ratio of false neighbors to the total neighbors for a system that exhibits the two-torus attractor mentioned in Fig. 1. Note that after an em-

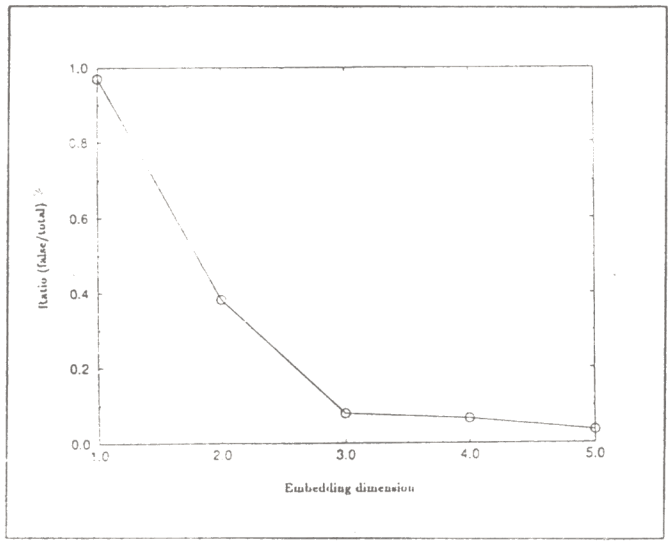

Fig. 8. Ratio of false neighbors to the total neighbors of an observable from a dynamical system with a 2-torus attractor. From this figure we can estimate (correctly) an embedding dimension of three.

bedding dimension three the ratio remains virtually unchanged indicating correctly that the proper embedding dimension is three. Another approach to define a proper embedding dimension is based on nonlinear prediction and it will discussed in the next section.

A final point which we cannot overstress. When we analyze some natural observed time series we are actually looking at a continuous process. Indeed, natural process are governed by some system of differential equations (flows). The dimension of any chaotic flow must be at least two. Otherwise the trajectory will have to be embedded in two dimensions and due to its continuous looping will eventually cross itself and afterwards will have to repeat exactly which is forbidden in chaotic systems. This is also evident from the Kaplan-Yorke conjecture. Since a chaotic system has at least one positive, one zero and the rest all negative Lyapunov exponents, $j \geq 2$ and $\frac{\sum_{i=1}^{j} \lambda_i}{|\lambda_{j+1}|} > 0$ which means then $D > 2$. Therefore a result $D < 2$ violates the above conjecture.

5 Nonlinear Prediction

While we look for additional tools in the search for determinism in measured time series, we should not forget that chaos is deterministic. Chaotic systems obey certain rules. Their limited predictive power is due to their sensitivity to initial conditions and to the fact that we cannot make perfect measurements. However, before their predictive power is lost (i.e. for short time scales) their predictability may be quite adequate and possibly better than the predictive power of linear statistical forecasting. The philosophy behind nonlinear forecasting is to explore the dynamics in order to improve predictions and to identify nonlinearities in the data.

Since we need to explain the dynamics we need to have the reconstructed attractor. Then we can begin to think how to improve short-term prediction. If an underlying deterministic mechanism exists, then the order with which the points appear in the attractor will also be deterministic. Thus, if we somehow are able to extract the rules that determine where the next point will be located in phase space we will obtain a very accurate prediction. In general, we can assume that the underlying dynamics can be written as a map f of the form:

$$\boldsymbol{x}(t+T) = f_T(\boldsymbol{x}(t))$$

where in phase space $\boldsymbol{x}(t)$ is the current state and $\boldsymbol{x}(t+T)$ is the state after some time interval T. For example, consider the sequence $\boldsymbol{x}(t)$: 0.12, 0.4224, 0.4759128, 0.094028, 0.3407468, 0.8985536, 0.36462, 0.9266888, ... If we plot $\boldsymbol{x}(t)$ vs $\boldsymbol{x}(t+1)$ we find that the points fall on a very well-defined parabola the expression of which we can easily find. In fact the sequence is the logistic map $x_{n+1} = 4x_n(1-x_n)$. In cases like that we can completely define the global dynamics. In reality, however, it is impossible to do that simply because we cannot always visualize the attractor or know a priori the form of the function f. In such cases a better approach is the local approximation (Farmer and Sidorowich, 1987). An illustration of how local approximation works is shown in Figure 9, where portions of a trajectory are shown in state space and a terminal point (present state) is denoted by

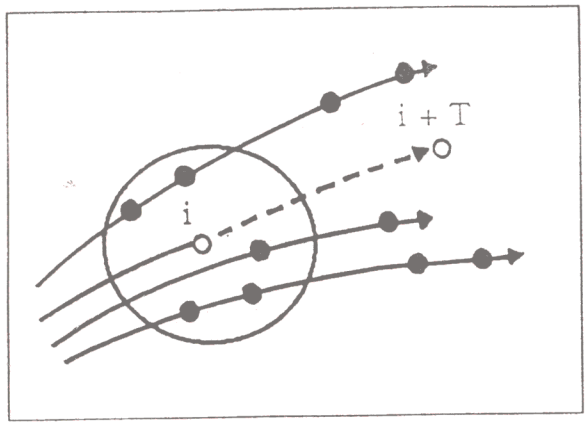

Fig. 9. An illustration of how local approximation works. The present state $x(i)$ and its unknown future value $x(i+T)$ are represented by open circles. The black dotes inside the circle define the neighborhood of $x(i)$ in this hypothetical state space. To make a prediction, we determine an appropriate mapping that takes the points in the neighborhood to some future time T, and then evaluate the mapping.

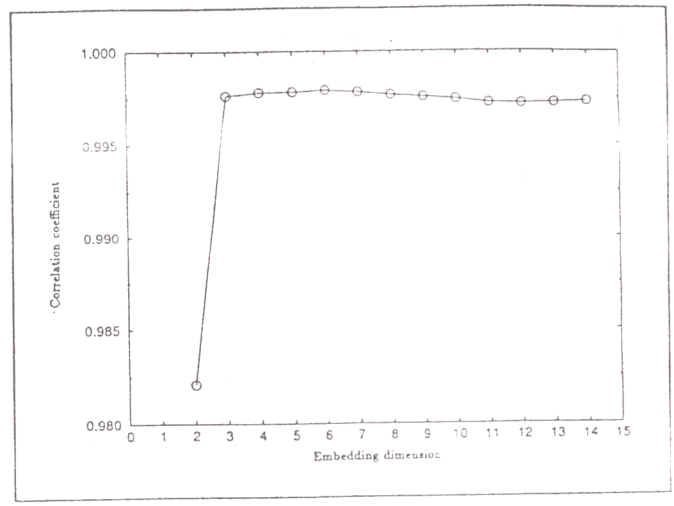

Fig. 10. Correlation coefficient between actual and predicted values for one step prediction ($r(1)$), as a function of the embedding dimension. The correlation increases and levels-off at an embedding dimension three which indicates that this is a proper embedding dimension.

an open circle. The solid circles indicate neighbors of the current state, and the arrowheads indicate movement of the neighbors through a local section of the embedding space. By finding a suitable function (linear or nonlinear) that describes how the neighbors advance, a prediction for the current state can be made. Since there are infinite nonlinear functions, the simplest way is to derive a linear mapping for each time step (i.e. construct f at each time step). Since the linear mapping may not be the same at each time step the overall procedure is **not** linear. Nonlinear prediction, unlike other methods for identifying chaos, maximizes the information in the available data and thus often works well with small data sets (Sugihara and May, 1990, Elsner and Tsonis, 1992). Calculation of the correlation dimension, for example, is based on the estimation of the scaling region which is typically small thus exploiting only a small subset of the available points in the phase space. Nonlinear prediction uses all the available data, thus requiring smaller samples. For that reason, nonlinear prediction has lately become a popular alternative to dimension estimates (for more information on nonlinear prediction see Tsonis, 1992).

Since we perform nonlinear prediction using the reconstructed attractor the previous problems of defining a suitable τ and an embedding dimension apply here as well. However according to Sugihara and May (1990) nonlinear prediction can by itself provide a may to determine a proper embedding for the data. The idea is that once a proper embedding has been found the correlation coefficient between actual and predicted values, $r(t)$ should not change. In other words predictability will increase as we approach the right embedding and after that the result will remain the same. Figure 10 shows the $r(1)$ (i.e. $r(t)$ for the first time step prediction) as a function of the embedding dimension for a time series that has as an attractor the same torus as in Figs. 1 and 8. Note that the correlation increases for $d_e < 3$ but remains unchanged for $d_e \geq 3$. Here again we find that the proper embedding is three. As a result the need for estimating the underlying dimension can be sidesteped.

The third problem, identified as the "fall-off" (of predictive power) in Fig. 2, has its source on the following. Since the predictive power of chaotic systems is limited one expect chaos to be characterized by a decrease of the correlation between predicted and actual values as prediction time increases. This property can be used to differentiate between chaos and additive uncorrelated noise

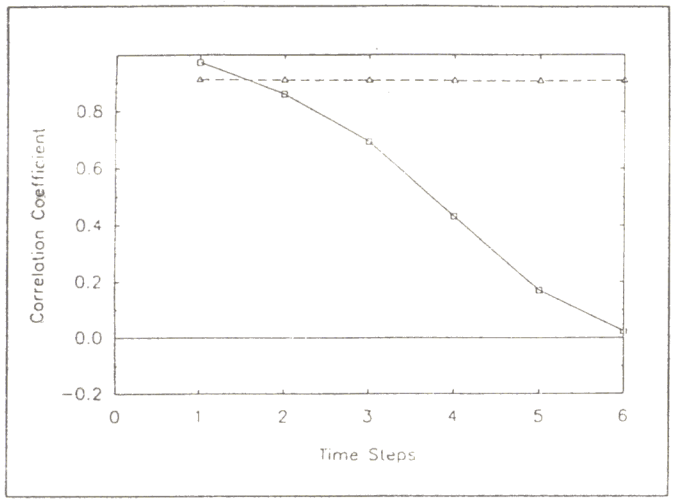

Fig. 11. Correlation coefficient $r(t)$ between actual and predicted values for the Lorenz system (solid line) and for a signal consisting of a sine wave plus noise (dashed line). The rapid drop of the correlation coefficient with prediction time is a characteristic of chaotic signals. In contrast, the independence of predictive skill with prediction time of the sine wave plus noise signal demonstrates that nonlinear prediction is capable of distinguishing between additive noise and chaos. Note that if there is no noise the predictive skill will level at $r(t) = 1.0$. As noise increase the $r(t) =$ constant level approaches zero.

(Sugihara and May, 1990). Additive noise produces a fixed amount of error regardless of the prediction time, as has been demonstrated for deterministic maps flows, and data from the natural world (see Figure 11). If a system is chaotic, then the decrease of predictive power with prediction time is equivalent to the presence of a positive Lyapunov exponent. The decrease or "fall-off" with time of $r(t)$ can indeed be used to infer the largest positive Lyapunov exponent (Wales, 1991).

The "fall-off", however, may not be a property of chaotic system only. Random fractal sequences (fBms) and other type of colored noise also exhibit a "fall-off". As it is demonstrated in Tsonis and Elsner (1992), however, $1 - r(t)$ for chaos scales with prediction time according to an exponential law (Figure 12) and for fBms it scales according to a power law (Figure 13). Therefore, nonlinear prediction can distinguish not only between chaos and periodic signals plus noise (or chaos and pure noise) but also between chaos and fBms, and possibly between chaos and other types of colored

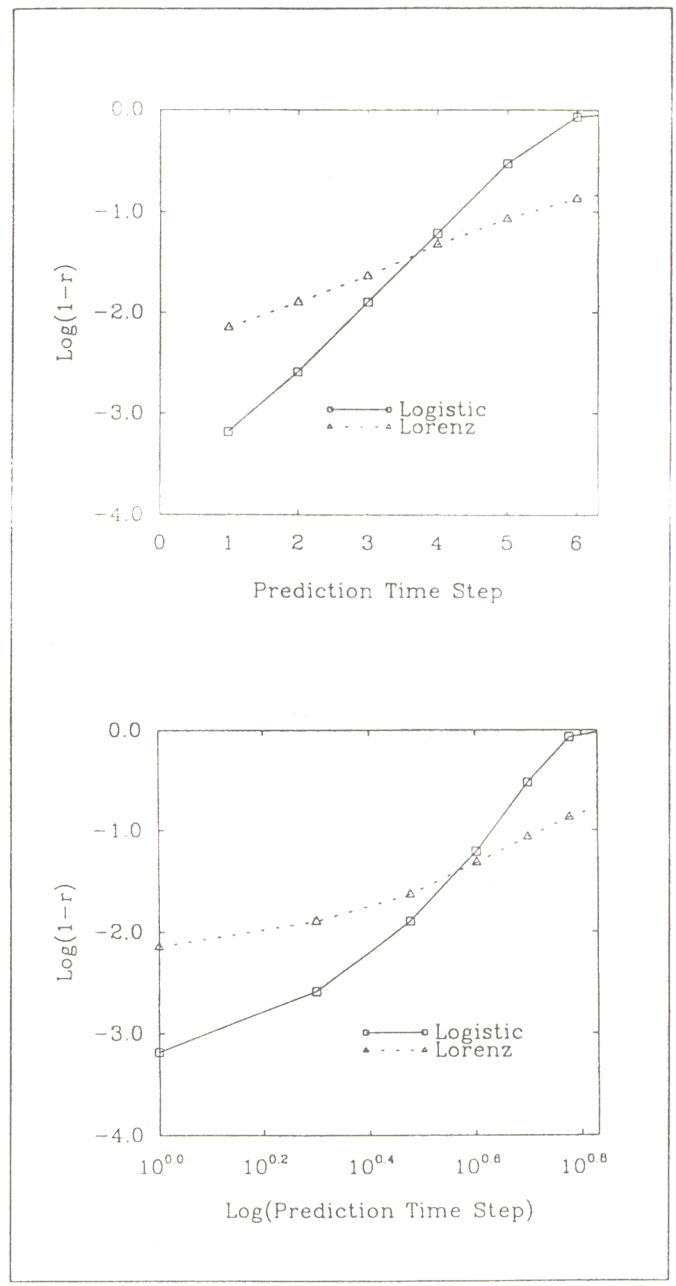

Fig. 12. Nonlinear prediction can be used to distinguish chaotic signals from random fractal sequences.
a) Logarithm of $1 - r(t)$ against prediction time step, t;
b) logarithm of $1 - r(t)$ against $\log t$ for single realizations of the x coordinate of the Lorenz system, and the logistic map.
The simulations show the expected scaling. For short prediction times, scaling is observed in the semi-log plot, indicating that the curve of $1 - r(t)$ against t curve for the chaotic signals is exponential.

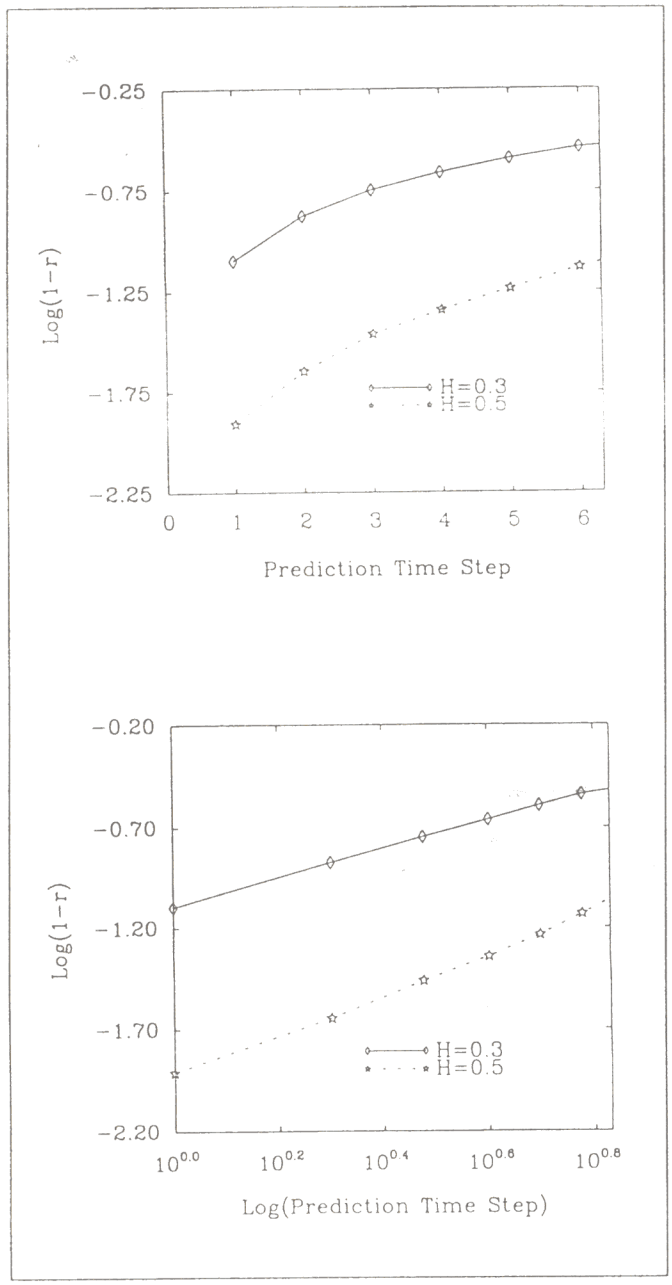

Fig. 13. As Fig. 12 but showing single realization results for two fBms with H=0.3 and 0.5. Here scaling is observed in the log-log plot indicating that for the fBms the curve $1 - r(t)$ against t is a power-low curve, as expected.

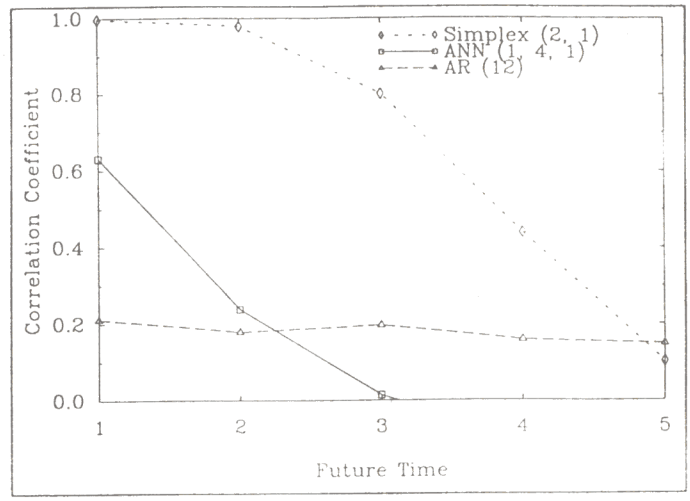

Fig. 14. Correlation coefficients between actual and predicted values as a function of future iterations of the logistic equation. The initial value used is $x_0 = 0.234889$. Comparisons are made between the two nonlinear models (ANN and SIMPLEX) and the linear autoregressive (AR) model. ANN(1,4,1) refers to an artificial neural network with one input node, four hidden nodes, and one output node, and Simplex(2,1) refers to a simplex model having an embedding dimension of 2 and a lag of 1. AR(12) refers to an autoregressive model of order 12. The example is chosen to emphasize that there is more to the evolution of dynamical systems than autocorrelation (autocorrelation is negligible in the logistic equation) and that certain nonlinear models are capable of short-term useful predictions by exploiting the nonlinear dynamics, even if the system in question is chaotic.

noise. As has been pointed out, however, by Tsonis and Elsner (1992) noise in the data may mask the actually scaling of $1 - r(t)$ with prediction time often making the distinction between chaos and colored noise difficult if not impossible. Nevertheless, nonlinear prediction has provided new alternatives to time series forecasting. When nonlinear structure exists in the data nonlinear prediction outperforms linear statistical approaches thus providing better forecasts even when the underlying dynamics are chaotic as the example in Figure 14 demonstrates.

6 Testing for nonlinearity

The bottom line from the above discussed issues is that many methods and not just one should be em-

ployed when we search for determinism in observed data. Dimension estimates should be accompanied by Lyapunov exponents estimates and by nonlinear prediction. Even then, however, the results may be inconclusive or challengable. Therefore, no matter what the approach and/or the results are, we should have a means to test our results. To this end over the past decade several approaches have been developed to test for nonlinearity. The latest and probably the most effective way is the method of surrogate data (Theiler et al. 1992). This procedure calls for the generation of a large number of random sequences of equal length as the time series to be tested. The idea is that the surrogate time series should be a non-deterministic record but similar in appearance to the original data. One method of surrogate generation is to preserve the amplitude spectrum of the raw data. First a Fourier transformation of the raw data is computed; then each complex amplitude is multiplied by $e^{i\phi}$ where ϕ is independently chosen from the interval $[0, 2\pi]$. As long as $\phi(f) = \phi(-f)$ it is guaranteed that the inverse transform is real. Finally, the inverse Fourier transform is the surrogate time series. The surrogate data have similar statistical properties such as mean variance etc to the original data. Their beauty, however, is that they preserve the autocorrelation structure which might be present in the original signal but not the dynamics (if any). An example of surrogate data is shown in Figure 15. In the past testing the significance of dimension estimates etc was based on the statistical comparison between the signal and a random sample obtained by "stirring" the original signal. Such a process results in surrogates that have the same mean and variance but not the autocorrelation structure that might exist even in otherwise random processes.

Once we have a way to produce such random processes we can define a null hypothesis against which the raw data can be tested using a discriminating statistic. According to the above algorithm for generating surrogate records, the null hypothesis is that the raw data come from a linearly autocorrelated Gaussian process. A variation of the above procedure where the null hypothesis is that the raw data come from a monotonic nonlinear transformation of a linear Gaussian process is also possible (Theiler et al. 1992). The discriminating statistic (e.g., Lyapunov spectrum, correlation dimension, correlation coefficient between predicted and actual values as a function of prediction time) is computed for each surrogate time series and its distribution approximated. If the discriminating statistic for the real data is significantly outside the mean of the distribution based on the surrogates, then the null hypothesis of linearly correlated noise is rejected and it can be concluded that significant nonlinear structure is present in the record. For a study case on surrogate data see Elsner and Tsonis (1993).

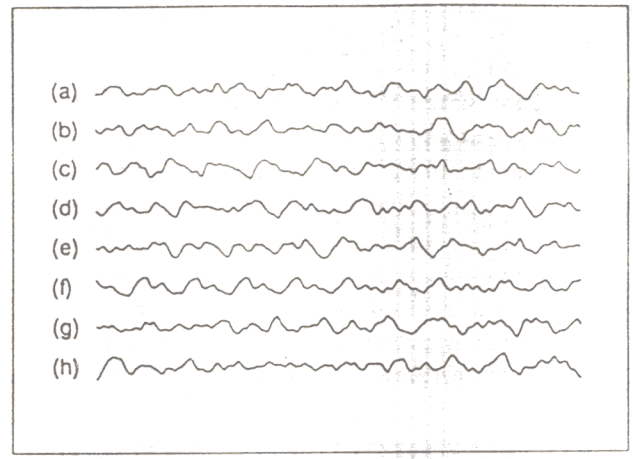

Fig. 15. Shown is eight time series. One of them is from the Mackey-Glass equation with $\tau = 30$ which is known to be low-dimensional and chaotic and the rest are surrogate obtained according to the Fourier method. Note the similarity between all time series. The surrogate here are random, but have preserved the autocorrelation structure but not the dynamics of the real data which by the way are time series (f) (figure of courtesy of Dr. James Theiler).

7 Conclusions

This work was designed to address some of the current issues involved in the search for determinism in observed time series. We have explained some ways according to which the algorithms might be working and at the same time we have demonstrated the existence of certain weaknesses. We like to conclude that special care should be considered when analyzing data. Preferably, not just one but all possible approaches should be considered,

and the results should be compared collectively. We suggest that the sensitivity of the conclusions be tested by repeating the analysis for various segments of the time series, for various τ's and various embedding dimensions. Finally a proper test for nonlinearity should be administrated. The theory of dynamical systems and chaos has provided us with new tools in analyzing observables. Whether or not low-dimensional attractors exist in nature is still debatable. Thus it is very critical that we apply these new theories with the utmost of care. Otherwise we may damage a theory that right now shows a great deal of promise.

References

Abarbanel, H. D. I., and M. B. Kennel, 1991: Lyapunov exponent in Chaotic systems: Their importance and their evaluation using observed data. *Mod. Phys. Lett. B.*, **5**, 1347-1375.

Abarbanel, H. D. I., and M. B. Kennel, 1992: Local false nearest neighbors and dynamical dimensions from observed chaotic data. Preprint, Institute for Nonlinear Science, University of California San Diego.

Elsner, J. B., and A. A. Tsonis, 1992: Nonlinear prediction, chaos and noise. *Bull. Amer. Meteor. Soc.*, **73**, 49-60.

—, and —, 1993: Nonlinear dynamics established in the ENSO. *Geoph. Res. Lett.*, **20**, 213-216.

Essex, C., T. Lookman, and M. A. H. Nerenberg, 1987: The climate attractor over sort time scales. *Nature*, **326**, 64-66.

Farmer, J. D., and J. J. Sidorowich, 1987: Predicting chaotic time series. *Phys. Rev. Lett.*, **59**, 845-848.

Frazer, A. M., and H. L. Swinney, 1986: Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, **33**, 1134-1140.

Fredericksom, P., J. Kaplan, E. Yorke, and J. Yorke, 1983: The Lyapunov dimension of strange attractors. *J. Diff. Eqs.*, **49**, 185-192.

Grassberger, P., and I. Procaccia, 1983a: Characterization of strange attractors. *Phys. Rev. Lett.*, **50**, 346-349.

—, and —, 1983b: Measuring the strangeness of strange attractors. *Physica*, **9D**, 189-208.

Nerenberg, M. A. H., and C. Essex, 1990: Correlation dimension and systematic geometric effects. *Phys. Rev. A*, **42**, 7065-7074.

Osborne, A. R., and A. Provenzale, 1989: Finite correlation dimension for stochastic systems with power-law spectra. *Physica D*, **35**, 357-381.

—, A. D. Kirwan, Jr., A. Provenzale, and L. Bergamasco, 1986: A search for chaotic behavior in large and mesoscale motions in the Pacific Ocean. *Physica*, **23D**, 75-83.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, 1980: Geometry from a time series. *Phys. Rev. Lett.*, **45**, 712-716.

Ruelle, D., 1981: Chemical kinetics and differentiable dynamical systems. *Nonlinear Phenomena in Chemical Dynamics*, A. Pacault and C. Vidal, Eds., Springer-Verlag.

Smith, L. A., 1988: Intrinsic limits on dimension calculations. *Phys. Lett. A.*, **133**, 283-288.

Sugihara, G., and R. M. May, 1990: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**, 734-741.

Takens, F., 1981: *Dynamical Systems and Turbulence. Lecture Notes in Mathematics 898.* Springer-Verlag.

Theiler, J., 1991: Some comments on the correlation dimension of $1/f^{\alpha}$ noise. *Phys. Lett. A*, **155**, 488-492.

—, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, 1992: Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, **58**, 77-94.

Tsonis, A. A., and J. B. Elsner, 1988: The weather attractor over very short time scales. *Nature*, **33**, 545-547.

—, and —, 1989: Chaos strange attractor and weather. *Bull. Amer. Meteor. Soc.*, **70**, 16-23.

—, and —, 1990: Comments on "Dimension analysis of climate data". *J. Climate*, **3**, 1502-1505.

—, and —, 1992: Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature*, **358**, 217-220.

Tsonis, A. A., 1992: Chaos: From theory to applications. Plenum, New York, 274 pp.

Tsonis, A. A., J. B. Elsner and P. A. Tsonis, 1993a: On the existence of scaling in DNA sequences. *Bioch. Biophys. Res. Comm.*, in press.

Tsonis, A. A., J. B. Elsner and K. P. Geprgakakos, 1993b: Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation. *J. Atmos. Sci.*, **50**, 2549-2555.

Wales, D. J., 1991: Calculating the rate of loss of information from chaotic time series by forecasting. *Nature*, **350**, 485-488.