International Journal of Bifurcation and Chaos, Vol. 17, No. 12 (2007) 4229–4243

RECONSTRUCTING DYNAMICS FROM OBSERVABLES: THE ISSUE OF THE DELAY PARAMETER REVISITED

ANASTASIOS A. TSONIS
Department of Mathematical Sciences,
Atmospheric Sciences Group,
University of Wisconsin-Milwaukee,
Milwaukee, WI 53201-0413, USA

Received November 3, 2006

In this tutorial/review we revisit the problem of choosing a proper delay when we reconstruct attractors. The effect of this delay is investigated thoroughly and a definite answer to this problem as well as a robust procedure to reconstruct attractors is presented.

Keywords: Attractor reconstruction; nonlinear dynamics; chaos; dimension estimates.

1. Introduction

Dynamical systems possess attractors. Attractors are limit sets on which the evolution of the system from some initial condition is confined. For nonlinear chaotic systems, the underlying attractor is a fractal set whose Haussdorf–Besicovitch (or fractal) dimension is smaller than the Euclidean dimension of the system's state space. The state space is a Cartesian coordinate system where the coordinates are the variables involved in the formulation of the dynamical system. Fractal objects, unlike Euclidean objects, possess no characteristic sizes or length scales [Mandelbrot, 1983], which means that they display detailed structure on all length scales (a property that is described by power laws). As a result of this attribute, fractal sets exhibit strange properties such as infinite length boundaries enclosing finite areas. This allows motions in the system's state space that correspond to trajectories that have infinite length but are confined on finite areas. Due to that, the trajectories never repeat and never cross themselves. This makes many chaotic evolutions appear indistinguishable from pure random processes.

The fractal character of an attractor does not only imply nonperiodic orbits; it also causes nearby trajectories to diverge. As with all attractors, trajectories initiated from different initial conditions soon reach the attracting set, but in chaotic (or strange) attractors, unlike in limit cycles or other periodic attractors, two nearby trajectories do not stay close. They soon diverge and follow completely different paths on the attractor. The divergence of nearby trajectories is measured by the positive Lyapunov exponents of the system. This is known as sensitivity to the initial conditions and is a major property of chaotic systems, which imposes limits in predictability. As a consequence, even though the dynamical system is described by a set of equations and therefore is deterministic, small uncertainties in the initial condition are amplified thereby causing the true trajectory to follow a different path in the attractor than if there was no uncertainty. In natural systems (such as climate) uncertainties in the initial condition are caused by measurement error. In mathematical systems (such as the logistic equation $x_{n+1} = 4x_n(1 - x_n)$), which can be

iterated from an exactly defined initial value), small errors are introduced by the unavoidable truncation or round-off errors. For example, the five first values of the evolution of the logistic equation from the initial condition $x_0 = 0.4$ are:

$$x_0 = 0.4$$
$$x_1 = 0.96$$
$$x_2 = 0.1536$$
$$x_3 = 0.52002816$$
$$x_4 = 0.9983954912280576$$
$$x_5 = 0.0064077372941726395570612432896$$

The digits after the decimal point increase (in fact double) with every iteration. After seven iterations the result carries 128 digits. After twelve iterations, there are 2,048 digits. The number of digits is actually given by 2^k, where k is the number of iterations. Calculating exactly out to only 100 steps will require a computer that will carry calculations with 2^{100} decimal points. This number is approximately equal to 10^{30}, which is one trillion times greater than the age of the universe in seconds. No computer has been developed to handle even a thousand digits.

Despite these limitations, the discovery of chaos led to the realization that what appears random may not be random but the result of low-dimensional deterministic dynamics (i.e. described by only a few equations), and that chaos can provide the framework to explain and describe complex behavior and to define the limits of predictability of natural systems.

If the mathematical formulation of a dynamical system is known, then its state space is known and investigating its properties is straightforward. If, however, the mathematical formulation of a system is not known or it is incomplete, then the attractor has to somehow be reconstructed from one or more observables (time series) of the system. This reconstruction is achieved by considering a scalar time series $x(t)$ and its successive time shifts (delays) as coordinates of a vector time series given by

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

where n is the dimension of the vector $\mathbf{X}(t)$ (often referred to as the embedding dimension) and τ is an appropriate delay [Packard *et al.*, 1980; Ruelle, 1981; Takens, 1981]. The idea behind constructing a delay space is that the dynamics of the unknown system are "imprinted" in a measured outcome from that system and that delays simply approximate derivatives (in principle, any dynamical system of m nonlinear differential equations can be reduced to one nonlinear differential equation of mth order).

In practice, since the dimension of the attractor of the unknown system is not known *a priori*, the embedding dimension must vary until we "tune" to a structure (attractor) whose characteristics (fractal dimension, for example) become invariant after some embedding dimension (an indication that extra variables are not needed to explain the dynamics of the system under investigation).

- The most common approach to infer the dimension of an underlying attractor is by estimating the correlation dimension [Grassberger & Procaccia, 1983a, 1983b]. At some embedding dimension n in our delay space, we have a cloud of points. Within this cloud we may find the number of pairs of points, $N(r, n)$, that are separated by a distance less than, r. If we find that this number scales with r according to

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

  then the scaling exponent d_2 is the correlation dimension of the cloud of points for that n. Since the above equation is a power law, the value of d_2 is estimated by the slope of the plot $\log N(r, n)$ versus $\log r$. We then repeat this procedure and we check if d_2 reaches a saturation value D_2 as n increases. If this happens, it indicates that we have "locked" into the underlying attractor whose correlation dimension is D_2. Once we have the correct embedding dimension, we can proceed with estimating the Lyapunov exponents and perform nonlinear prediction.
- Since trajectories from different initial conditions are confined on an attractor with dimension less than that of the state space, these trajectories have to converge onto this attractor. This convergence is measured by the negative Lyapunov exponents. Once the trajectories have reached the attractor and if the attractor is chaotic then they diverge. This divergence is measured by the positive Lyapunov exponent(s). In addition to positive and negative exponent, any continuous dynamical system will have at least one zero exponent corresponding to the slowly changing magnitude of the principal axis tangent to

the flow. In theory the Lyapunov exponents 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 state space in which the attractor is embedded. There are as many Lyapunov exponents as coordinates in the state space. We can imagine this hypersphere being a set of initial conditions which under the operation of the dynamics are pulled onto the attractor and then are moved inside the attractor. This pull and subsequent motion distorts the shape of the sphere according to how fast and in which direction the sphere shrinks or expands. The estimation of the Lyapunov exponents from a system of ordinary differential equations is straightforward and is based on the fact that the linearized equations, describing the local dynamics, involve a Jacobian whose eigenvalues provide all the exponents (see [Tsonis, 1992]). In practice (i.e. from an observable), at first we embed the data in some appropriate space (a good recommendation is a space of dimension at least $2D_2$). Then we monitor the motion in that space of a point and also of points in its close neighborhood for some time called the decomposition length [Abarbanel *et al.*, 1991; Brown *et al.*, 1991]. From this monitoring one can estimate the Jacobian, which provides information about the Lyapunov exponents. Because the linearized equations provide the local dynamics such an approach provides an estimation of the local Lyapunov exponents. By repeating this procedure for many points we can obtain an average picture which will be related to the average Lyapunov exponents of the system.

- Chaotic systems obey certain rules. Their limited predictability is due to their sensitivity to initial conditions and to the fact that we cannot make perfect measurements nor have infinite computing power. 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 prediction is to explore the dynamics in order to improve predictions and 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 of 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 state space, then we will obtain very accurate prediction. In general, we can assume that the underlying dynamics can be written as a map f of the form

$$x(t+T) = f_T(x(t))$$

where in state space $x(t)$ is the current state and $x(t+T)$ is the state after some time interval T. In reality, however, we cannot easily find the actual form of the function f. A solution to this problem is an approach called the local linear approximation [Farmer & Sidorowich, 1987]. According to this approach, the future state is determined by a linear mapping that applies to the evolution of a small neighborhood around the current state. Since the linear mapping may not be the same at each time step the overall procedure is not linear. An improvement in forecasting against other linear or statistical approaches indicates nonlinearity and the presence of chaos. The rate at which predictability decreases should also reflect the loss of predictability due to the presence of positive Lyapunov exponents. In fact, an exponential decrease in predictability is a feature often used to identify chaos in data [Sugihara & May, 1990; Tsonis & Elsner, 1992; Wales, 1991].

In the mid and late 1980s when these methods were introduced several problems emerged. These problems had to do with sample size, defining the scaling region, the statistical significance of estimated dimension and other nonlinear measures, and the choice of the delay parameter τ.

The problem of the sample size had to do with what was the necessary number of points to accurately estimate a dimension D. Initially, theoretical results suggested that in order to estimate a dimension D one needed 42^D data points [Smith, 1988]. This is a large number even for low values of D. Given the small data sizes available in most disciplines, if this were true all attempts to estimate dimensions will be in vain. This issue was clarified

soon. It is now widely accepted that this number is closer to $10^{(2+0.4D)}$ rather than 42^D. This number which is known as the *Tsonis criterion* has been verified with systems of known dimensions [Nerenberg & Essex, 1990; Tsonis, 1992; Tsonis *et al.*, 1993; Tsonis *et al.*, 1994].

The issue of how to define the scaling region (the region where Eq. (2) holds) emerged from the fact that, due to the compressing properties of logarithms, in a log–log plot many nonpower-laws may appear linear. This issue was settled by adapting the process of identifying plateaus in the $\mathrm{dlog}\, N(r,n)/\mathrm{dlog}\, r$ versus $\mathrm{dlog}\, r$ plot [Tsonis, 1992; Tsonis *et al.*, 1994]. The idea is that if scaling exists, then $\mathrm{dlog}\, N(r,n)/\mathrm{dlog}\, r$ is constant and independent of r. Such a plot provides the *slope* (also referred to as the *local slope*) as a function of $\log r$. A second problem associated with the issue of scaling was exactly how wide the scaling region should be in order to qualify as true [Avnir *et al.*, 1998; Tsonis, 1998]. Yet another issue emerged when reports appeared in the literature that certain stochastic processes exhibiting power law spectra may fool the algorithms used to estimate dimensions and make nonlinear prediction [Osborne & Provenzale, 1989], thereby falsely indicating chaotic dynamics.

In view of the above issues, the issue of the statistical significance of a result linking a data set to low-dimensional chaos arose from the desire to provide more confidence about the results. In this effort the generation of surrogate data [Theiler *et al.*, 1992; Schreiber & Schmitz, 1996] emerged as a powerful tool. In general, surrogate data are random data, which however, preserve the autocorrelation function and power spectra of the original data set. They are usually generated by inverting the spectra of the original data and randomize the phases. This, while it destroys whatever dynamics are present, preserves the autocorrelation in the data. By generating a large number of such surrogate records one is then able to test the null hypothesis that the estimated dimension, or Lyapunov exponent, or scaling region, is statistically significant at a desired significance level. At the same time specific tests to address problems with limited scaling or questionable stochastic processes were proposed [Tsonis, 1992; Tsonis & Elsner, 1995].

The final issue of the proper choice of the delay parameter in Eq. (1) is the subject of this tutorial/review. Equation (1) provides no guidance about what τ should be. Initially, it was recommended that a delay parameter should be chosen so that it delineates nearly uncorrelated points in the reconstruction. Employing too small of a τ could result in highly correlated points, which tend to arrange themselves on narrow bands in phase space thus leading to an underestimation of the actual dimension. Possibilities ranged from the first zero of the autocorrelation function of the data (a linear measure, but easily estimated), to the time lag where the autocorrelation function attains the value of $1/e$, to the first minimum of the mutual information function (a nonlinear measure not easily estimated) [Tsonis *et al.*, 1994].

By the early 1990s all the above issues were thus addressed and a certain procedure was in place for proper nonlinear time series analysis. However, while the issues of sample size and scaling were addressed extensively and rigorously, the issue of τ was never really addressed rigorously either theoretically or experimentally. There was always some arbitrariness about the first zero of the autocorrelation function or the first minimum of the mutual information. The reason for this was that in the mid 1980s computer memory and speed was not adequate enough to research this issue in detail. Now that computer speed and memory is not a problem, we can go back and investigate this issue in detail. What we find is indeed very interesting. In what we present next it is assumed that a sufficiently large number of points (sample size) are available. Thus, issues of sample size requirement do not apply here.

2. The System Used

In order to demonstrate the effect of the delay parameter we will employ the Lorenz system [Lorenz, 1963]. Edward Lorenz, an atmospheric scientist at MIT, who was trying to explain why weather is unpredictable, reduced the complicated physics of the atmospheric circulation into three simple nonlinear ordinary differential equations, which modeled the behavior of a fluid layer heated from below. This is an approximation of what happens basically every day in the lower atmosphere. The sun rises; the surface absorbs solar radiation and gets warm. Subsequently, the air gets warm by contact with the warmer surface and rises. This rising motion leads to convective turbulent motion. The mathematical formulation of this system is the

following:

$$\frac{dx}{dt} = -ax + ay$$

$$\frac{dy}{dt} = bx - y - xz$$

$$\frac{dz}{dt} = xy - cz$$

where roughly speaking x is proportional to the intensity of convective motion, y is proportional to the horizontal temperature variation, z is proportional to the vertical temperature variation, and a, b, c are constants. Figure 1 shows details of this system's attractor for $a = 10$, $b = 28$ and $c = 8/3$ after integrating the above equations for 10,000 time steps, 5,000 of which are colored blue and 5,000 green. Panels (a)–(c) show the attractor from three different perspectives. Panel (d) is a close up revealing the fine nonrepeating fractal structure of the attractor. The apparent yellowish parts in the figure are very close blue and green lines, which due to

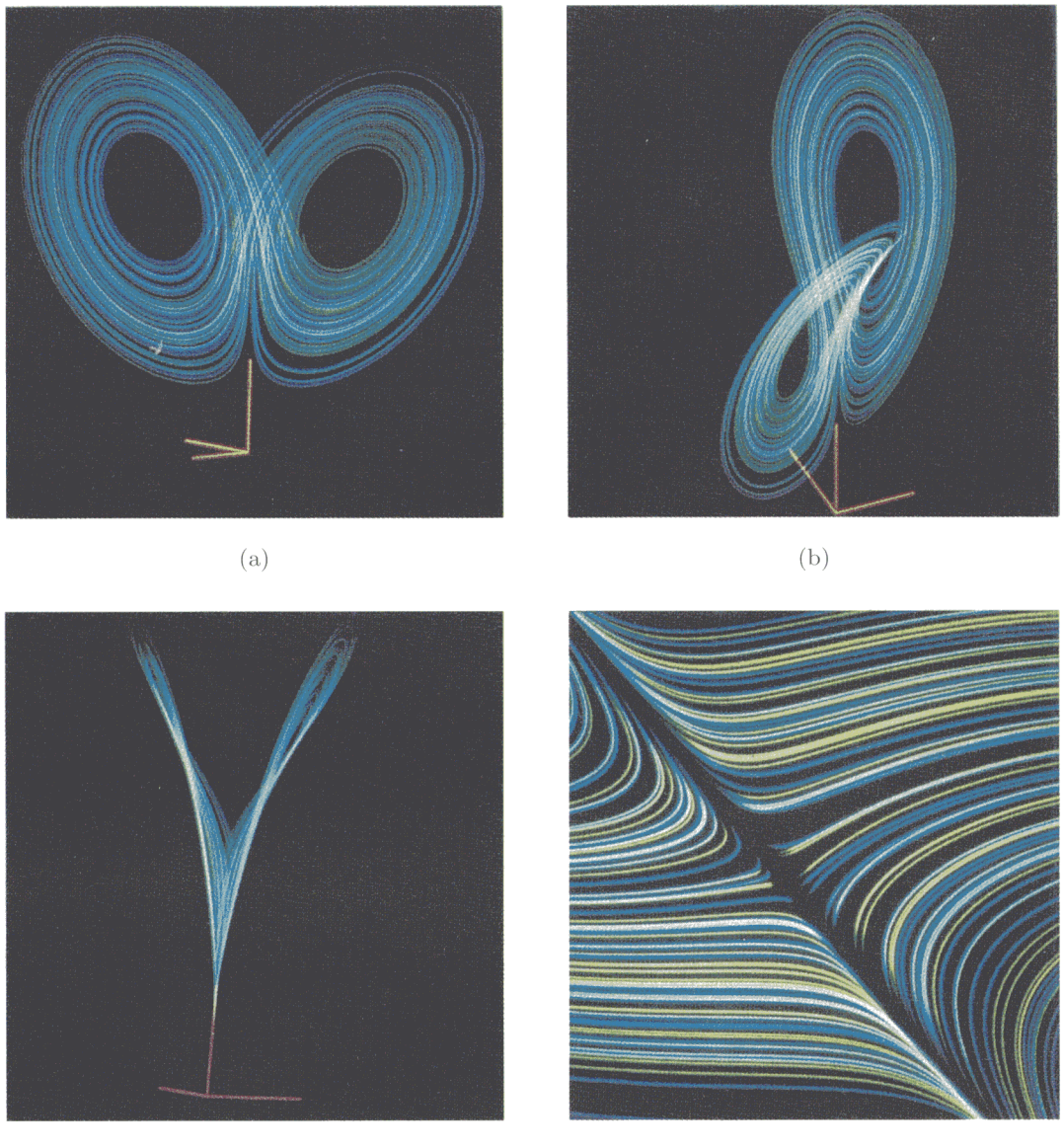


Fig. 1. The Lorenz [1963] attractor for $a = 10$, $b = 28$ and $c = 8/3$ after integrating the above equations for 10,000 time steps, 5,000 of which are colored blue and 5,000 green. Panels (a)–(c) show the attractor from three different perspectives. Panel (d) is a close up revealing the fine nonrepeating fractal structure of the attractor. The apparent yellowish parts in the figure are very close blue and green lines, which due to their closeness cannot be distinguished by the photographic device thereby appearing yellowish. [Figure is courtesy of Gottfried Mayers-Kress and has appeared in the book *From Cardinals to Chaos*, Cambridge University Press, 1989.]

their closeness cannot be distinguished by the photographic device thereby appearing yellowish. The attractor is definitely not planar but it consists of two rather thin lobes. Note that while here we will present results from the analysis of the Lorenz system, the general conclusions have been verified with other dynamical systems.

2.1. *Reconstructing the attractor*

Let us now assume that instead of the complete description of the Lorenz system we have a record of one of the variables, say $x(t)$ (Fig. 2). We obtain this "observable" by integrating the above system using the fourth order Runge–Kutta method and a time step of 0.01. Figures 3 and 4 show the autocorrelation function and mutual information (solid line) for this observable. The autocorrelation function attains its first zero at about a lag of 200 and

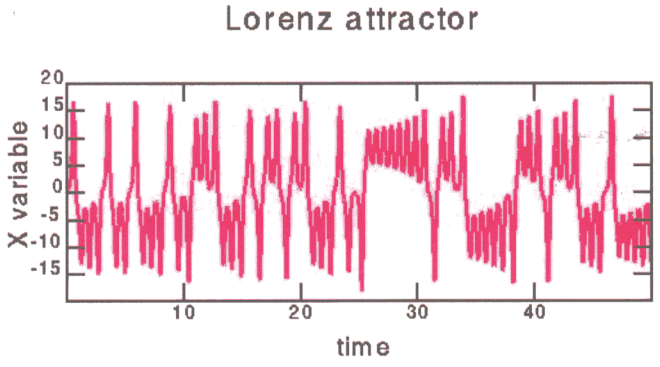


Fig. 2. A time series of the x variable of the Lorenz system (courtesy of Eric R. Weeks).

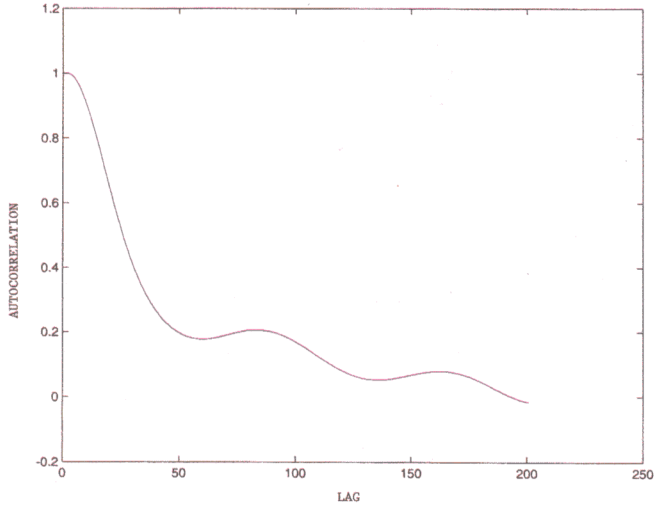


Fig. 3. The autocorrelation function of the x variable of the Lorenz system.

the mutual information attains its first minimum after about 12 time steps. We then proceed with reconstructing the attractor from this observable. Figures 5 and 6 show the reconstructed attractor in three dimensions using a time delay ranging from 2 to 200. In order to visualize what happens as τ increases, in Fig. 5, τ is varied very slowly. What we observe is that even at the very small $\tau = 2$ the two lobes of the attractor are visible. As τ increases the attractor widens (stretches) in one direction and folds on the top and bottom. This stretching and folding continues as τ keeps on increasing eventually resulting in a rather amorphous structure. A similar picture is observed when another variable $(z(t))$ is considered (see Fig. 7). The reason for this is that a larger τ teams values that are more separated from each other in the time series. When τ becomes too big there is simply no correlation between the points. These figures suggest that a τ corresponding to a zero in the autocorrelation function of $x(t)$ may not be appropriate. A delay close to the first minimum of the mutual information function may be more appropriate even though a smaller τ (4 or 6) gives a more realistic delineation of the attractor. So, what is really the proper τ when it comes to reconstructing the attractor?

The sequence of reconstructions in Figs. 5–7 and the stretching and folding observed as τ increases reminds us of how a strange attractor is born. We all know from theory that a volume element in state space under the action of a chaotic flow will at first be pulled along the direction of greatest instability. This stretching, however, cannot occupy more and more volume (because the attractor is confined in a finite area in state space). The mechanism that prevents this is folding. Thus, the attractor has to fold onto itself. Successive iterations of this process result in the asymptotic attractor with folds within folds *ad infinitum* (i.e. a fractal object). The fact that an increasing τ emulates this process provides a key observation, which points to a way of estimating the proper τ: the optimum τ should be the one that approximates more closely the actual stretching and folding of the attractor. Since the stretching and folding is a physical property, it may not relate well to statistical measures such as the autocorrelation function or mutual information. This will explain why neither choices of τ give consistent results [Martinerie *et al.*, 1992].

Now let us assume that our attractor is a set of points in a two-dimensional state space (Fig. 8, top)

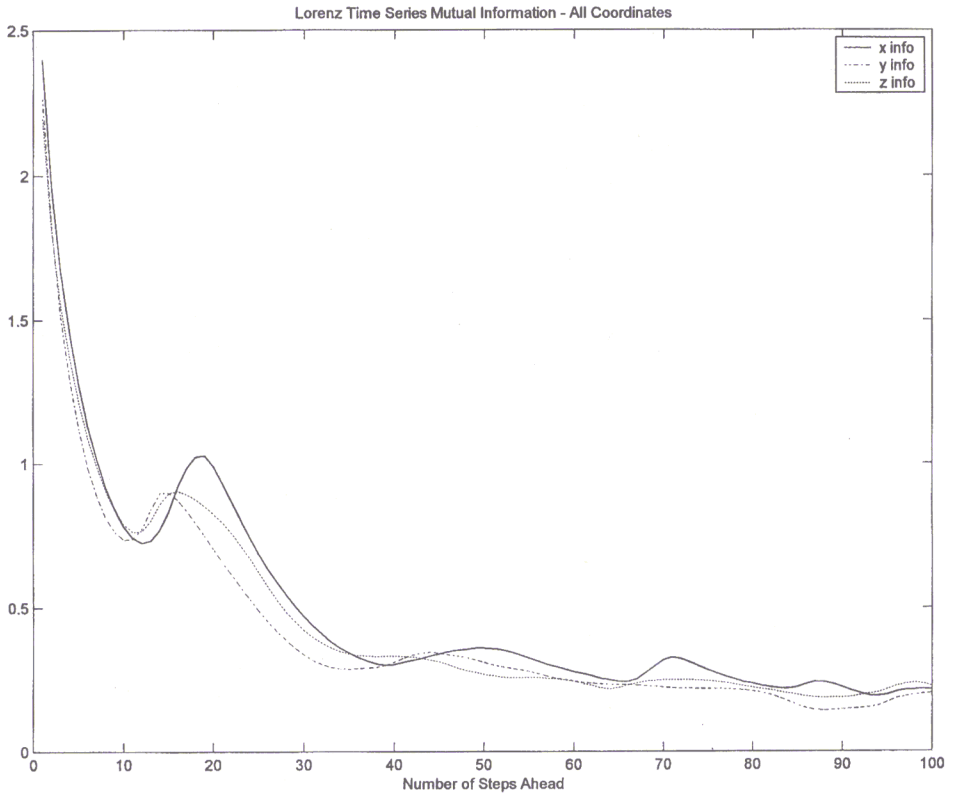


Fig. 4. The mutual information function of the x variable of the Lorenz system (solid line).

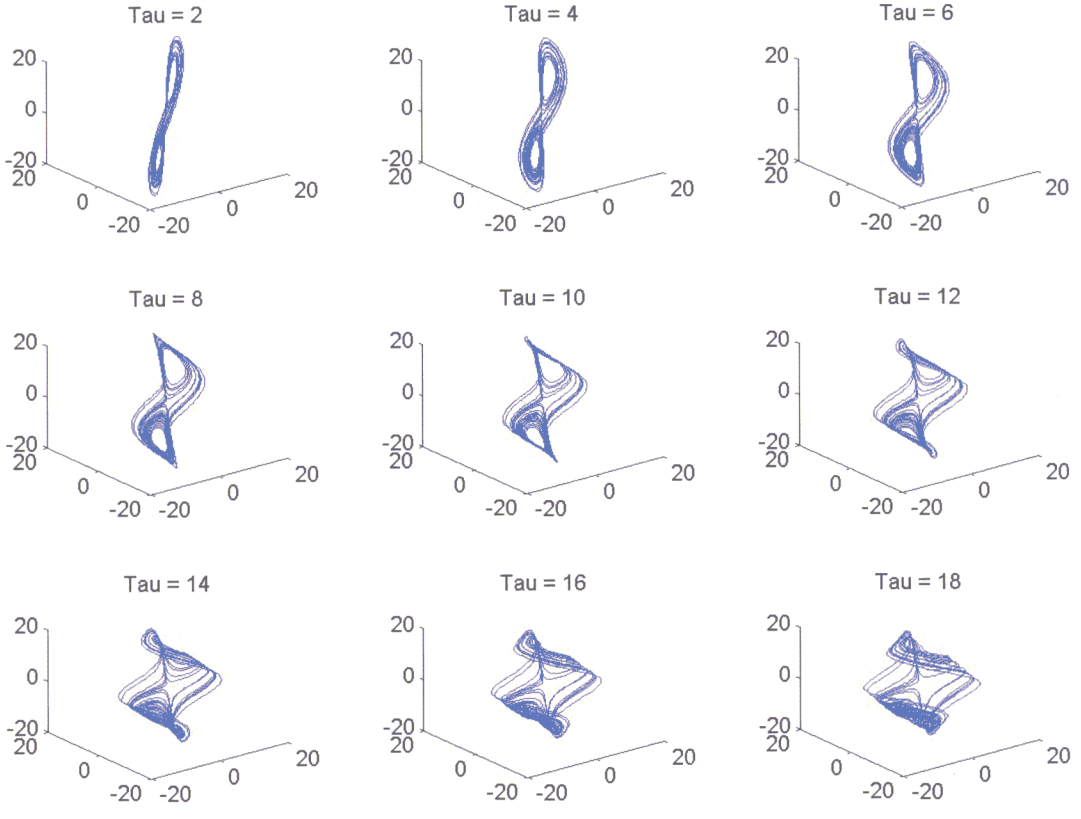


Fig. 5. Reconstruction of the Lorenz attractor from $x(t)$.

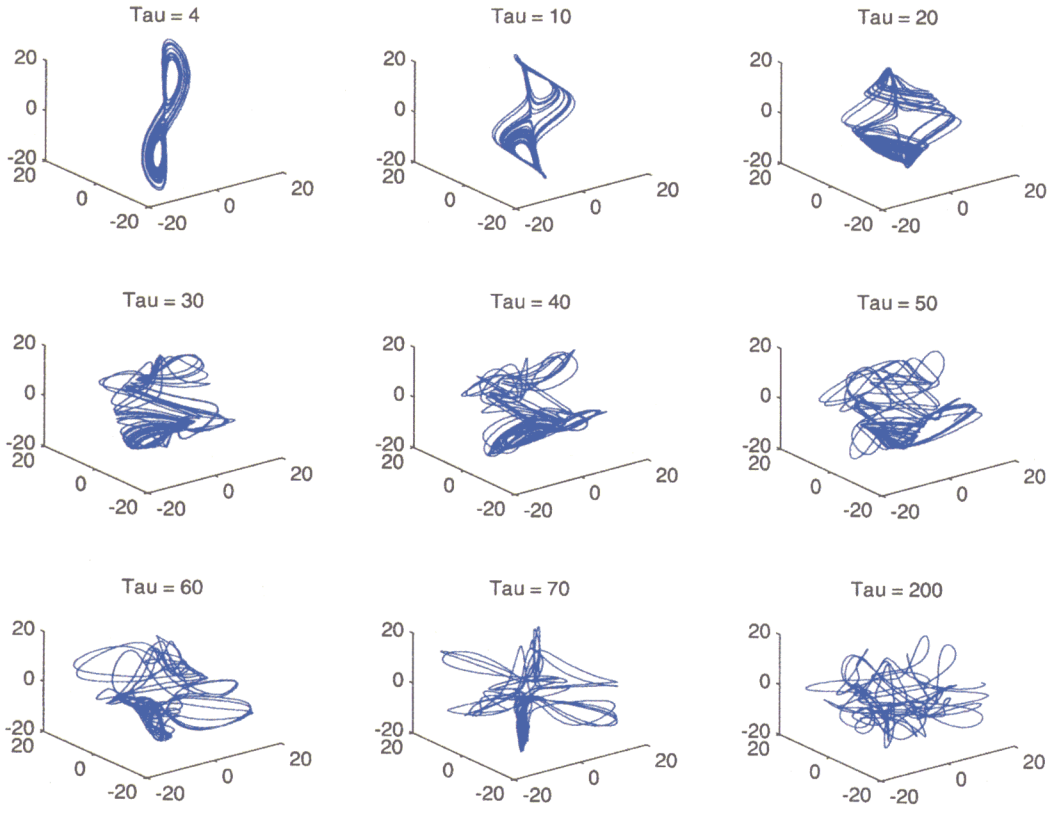


Fig. 6. Same as Fig. 5 but for additional delay parameters τ.

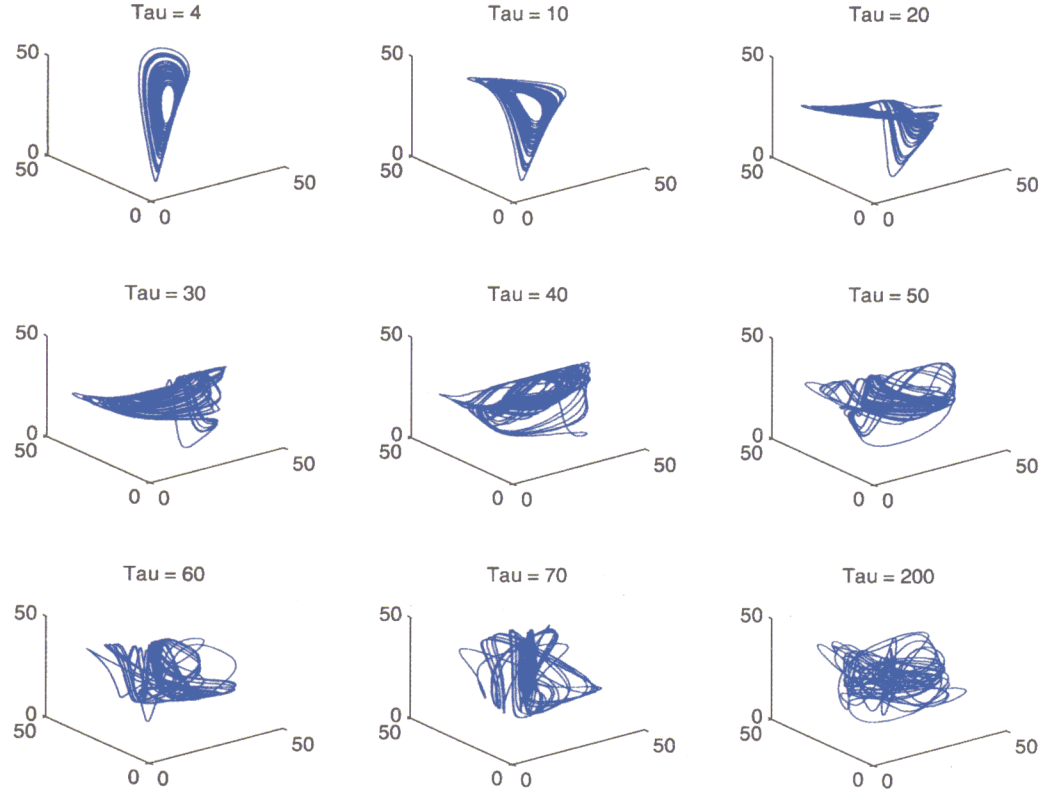
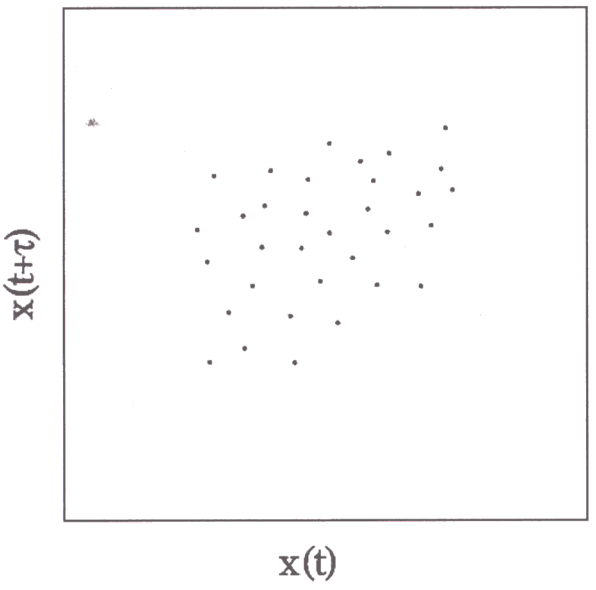


Fig. 7. Same as Fig. 6 but for the variable $z(t)$.



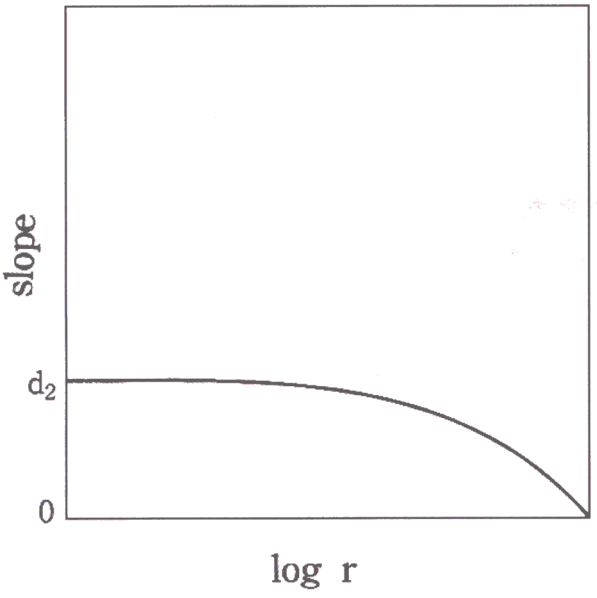


Fig. 8. A hypothetical cloud of points embedded in two dimensions and the expected *slope* versus $\log r$ function.

having a correlation dimension D_2. How does the *slope* versus $\log r$ plot would look like? For an embedding dimension of two, for scales smaller than the diameter of the attractor, the plot will show a plateau at slope $= d_2$. However, since the attractor has a finite diameter, for distances greater than this diameter $N(r, n)$ will reach a saturation value and will not change further with r. Because of that, the $\mathrm{d}\log N(r, n)/\mathrm{d}\log r$ function (i.e. the *slope*) will tend to zero for very large r's. These qualitative characteristics of the *slope* as a function of space scale are illustrated in Fig. 8 (bottom). Now what

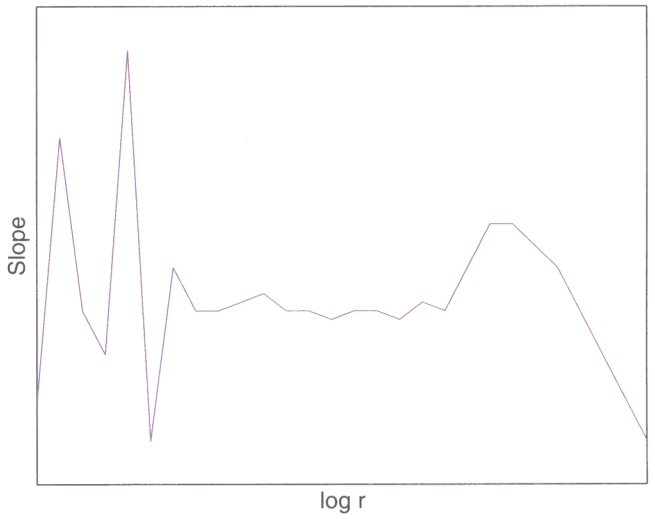


Fig. 9. Expected features of the *slope* versus $\log r$ function when data in Fig. 8 are embedded in a dimension higher than two.

do you think will happen if we embed the data in three dimensions? In this case we distribute the points in the attractor into a much larger volume. The effect of this is that the density of points per unit volume decreases. Now, a pair in two dimensions separated by a distance r is separated by a longer distance. This "depopulation" results in poor statistics over small scales and manifests itself as fluctuations about the plateau. On the other hand, because we have embedded the attractor in a higher dimension than needed, the larger scales will appear higher-dimensional than they actually are (much like when we curl a sheet of paper; the smaller scales remain planar but in large scales the sheet of paper appears three-dimensional). Thus when we embed the data in higher than needed dimension, the local slope plot will show large fluctuations at very small scales, a plateau at intermediate scales at the level slope $= d_2$, a tendency for higher slope values at large scales (manifesting itself as a "hump"), and finally an approach to zero values for very large scales (Fig. 9). It follows, that a proper embedding is one that minimizes the distortions at the two ends resulting in a well defined plateau over the widest range of scales.

Now recall from above that a proper τ is one that emulates the properties of the attractor closely and that the effect of an increasing τ is also to pull nearby points apart. One should then expect a similar behavior in the *slope* versus $\log r$ plot as τ varies. If this is true then the proper τ should be the one that also results in the widest and best

defined plateau. Thus, we should be able to define the proper τ as well as the dimension of the attractor in a common procedure where the *slope* versus $\log r$ plot is produced by simultaneously varying the embedding dimension and delay parameter.

This procedure is demonstrated in Fig. 10. Each panel corresponds to a specific τ and shows the first derivative of $\log N(r, n)$ versus $\log r$ function (or local slope) for embedding dimensions 2, 3 and 4. As we mentioned earlier a plateau in such a graph identifies the scaling region. If this plateau reaches a constant level as the embedding dimension increases, it indicates that we have tuned to the invariant attractor. This figure displays some

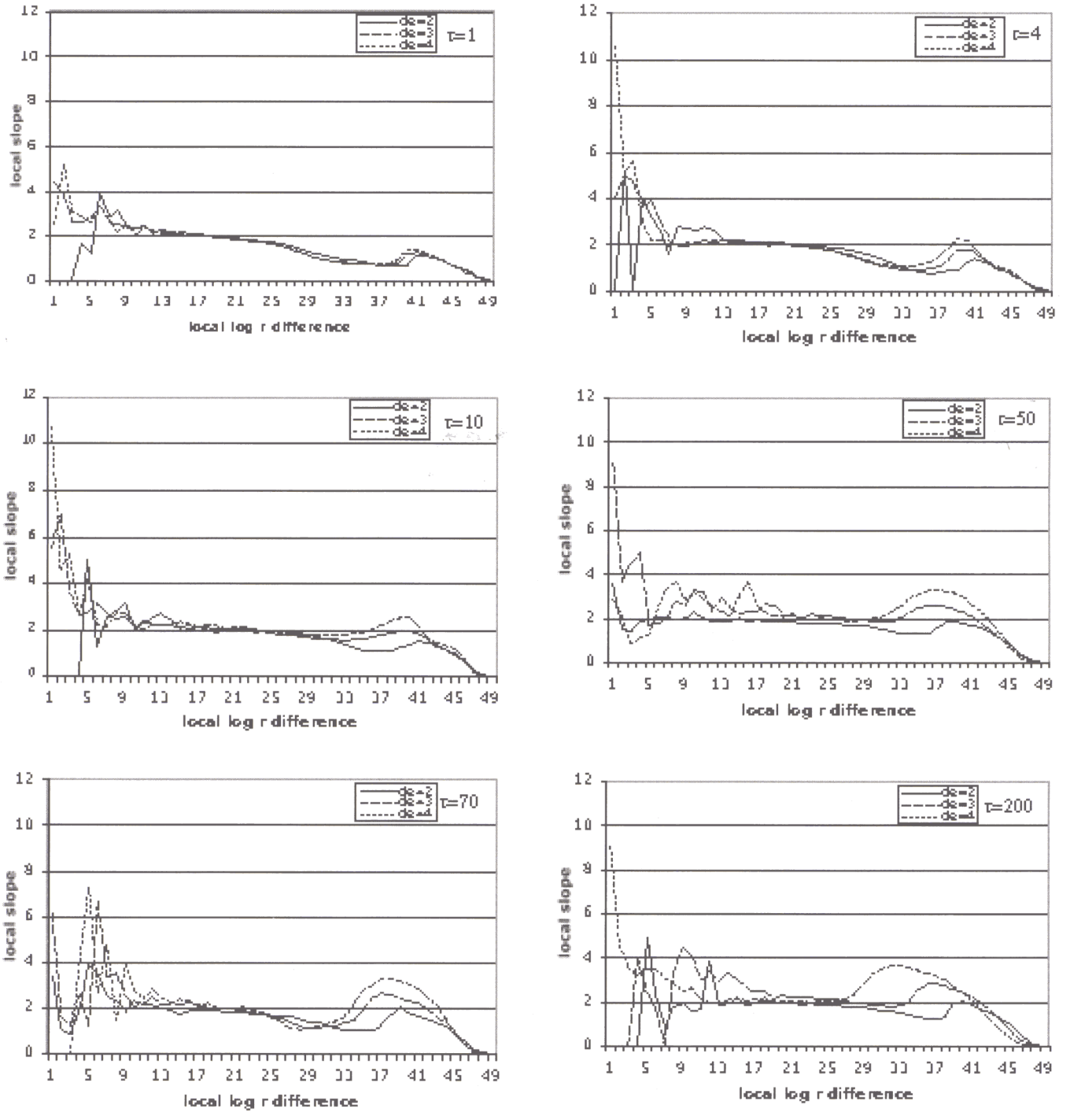


Fig. 10. *Slope* versus $\log r$ plots for $x(t)$ and for varying embedding dimension and delay parameter. The x-axis labeling from 1–49 reflects the fact that $N(r, n)$ was calculated for 50 r values.

important features. First, we observe the expected behavior from Fig. 9. For any fixed τ, as the embedding dimension increases, the fluctuations at small scales increase and the "hump" over the larger scales becomes more and more pronounced. The same is observed if we fix the embedding dimension and vary τ. This demonstrates that whether we fix τ and vary the embedding dimension or we fix the embedding dimension and vary τ, the two are equivalent. The proper (or optimum) τ should then be the one that, in graphs like the above, delineates the widest plateau. From our analysis it appears that this occurs around $\tau = 10$ where D_2 is slightly above 2.0 (this value of D_2 is in very good agreement with its estimates in the literature of 2.05). It is interesting to note that in Fig. 10 basically all τ's provide a very close estimation of the dimension. Now that we understand what an increasing τ does to the attractor (stretching and folding it) we can explain this result easily. Topologically, a stretching or folding does not affect the dimension as long as this action does not tear the structure or opens holes in it. Thus, we find that basically any τ will do when it comes to dimension estimates, even very small ones (this however, may not be true as we will see later for estimating other properties of attractors). Note that the proper τ in this example is consistent with the first minimum of the mutual information. This does not prove that the first minimum of the mutual information is always the proper choice of τ. It has been shown [Martinerie *et al.*, 1992] that this choice is not always consistent.

2.2. *Nonlinear prediction*

Let us now see what the effect of τ is on nonlinear prediction. As we mentioned in the introduction once we have an appropriate embedding we can consider a terminal point and its close neighborhood. Then we can find the linear mapping of the motion of its neighbors and extrapolate it to get a projection of the terminal point into the future. The performance of nonlinear prediction is evaluated by estimating the correlation between predicted and actual values as a function of the prediction time step [Farmer & Sidorowich, 1987; Wales, 1991]. Figure 11 shows this correlation for six different delays. We observe that prediction is very good

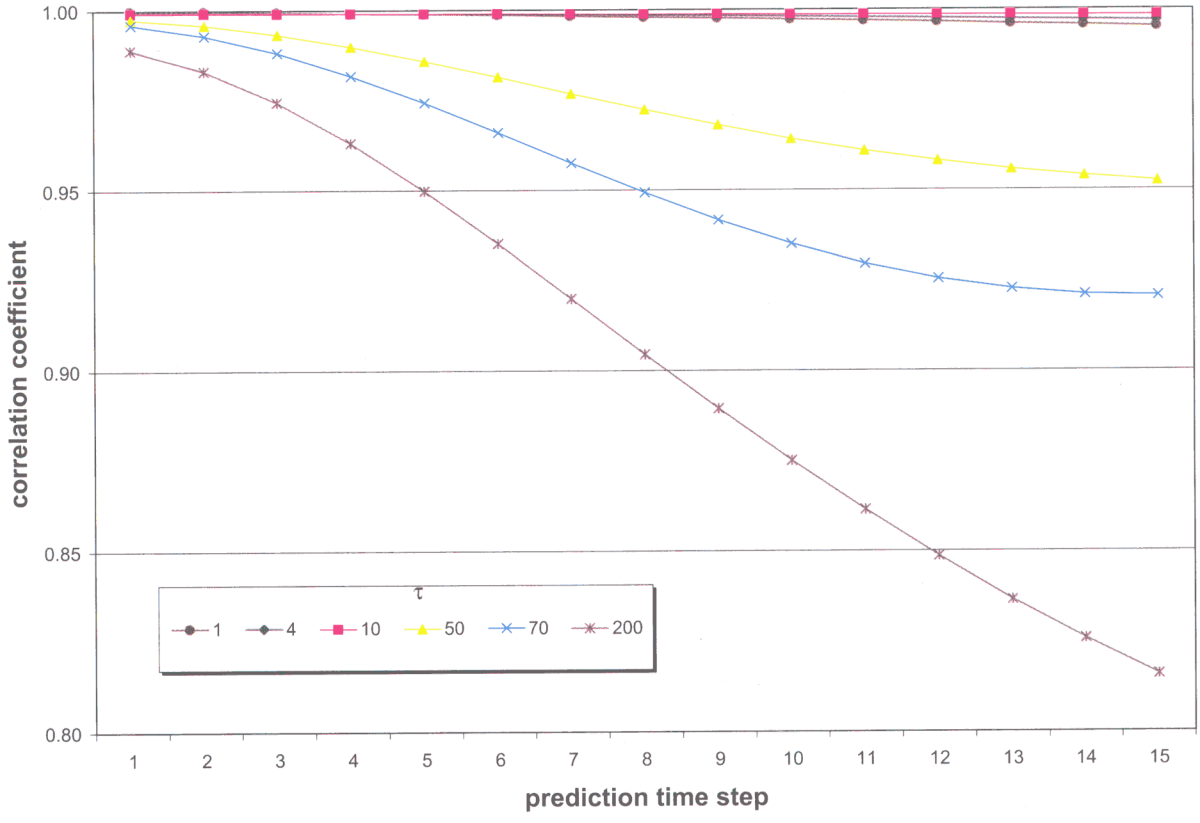


Fig. 11. Nonlinear prediction for different delays.

and more or less identical for $\tau = 1$, $\tau = 4$ and $\tau = 10$. For greater delays the accuracy of prediction decreases noticeably. Thus, unlike our conclusions with dimension estimates, here we cannot say that any delay will do a good job. The folding and stretching effect of increasing τ is in *addition* to that of the attractor. As such, when a greater than the optimum delay is used, this "extra" action expands the neighborhood and mixes points in the attractor more than necessary. This affects the local linear mapping thereby decreasing predictability. Interestingly, a delay parameter smaller than the optimum does not have this effect. We suspect that the squeezing of the attractor for very small delays simply makes the neighborhood denser (as if we had more points available). As a consequence the results are basically identical for smaller delays.

2.3. *Estimating the Lyapunov exponents*

Finally let us examine the effect of τ in the estimation of the Lyapunov exponent spectrum. The divergence or convergence rate of nearby trajectories in strange attractors is not uniform. Depending on the region in the attractor the divergence or convergence rate may be greater or smaller. As such the Lyapunov exponents are not constant but they vary as the trajectories move on the attractor. Figure 12 illustrates this by showing the positive Lyapunov exponent of the Lorenz system along a trajectory. The overall Lyapunov exponent is the average of all

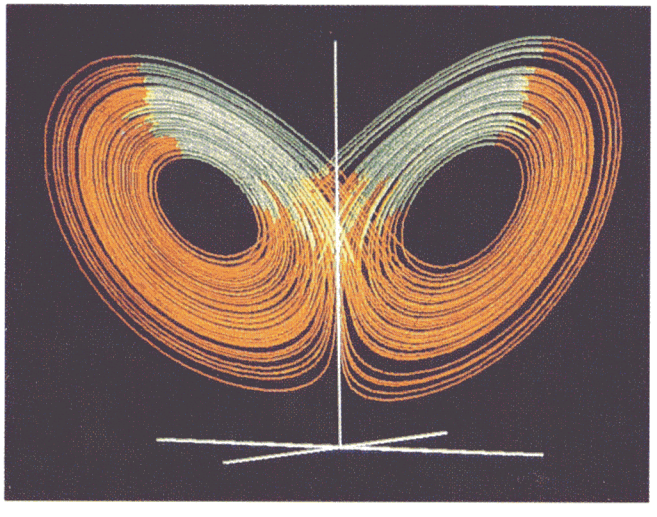
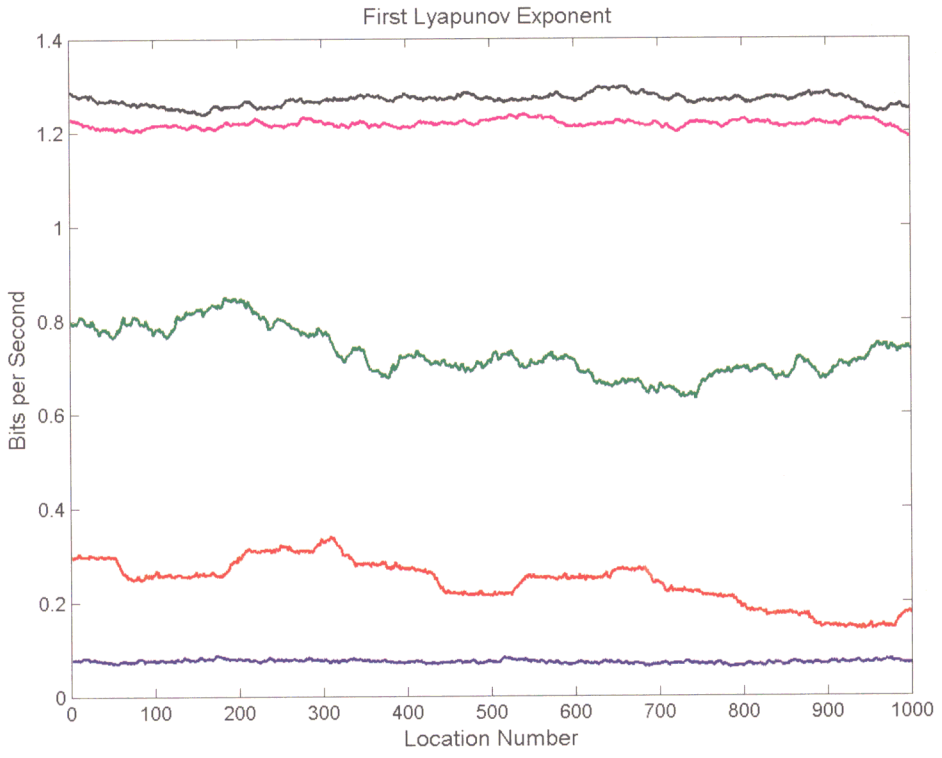
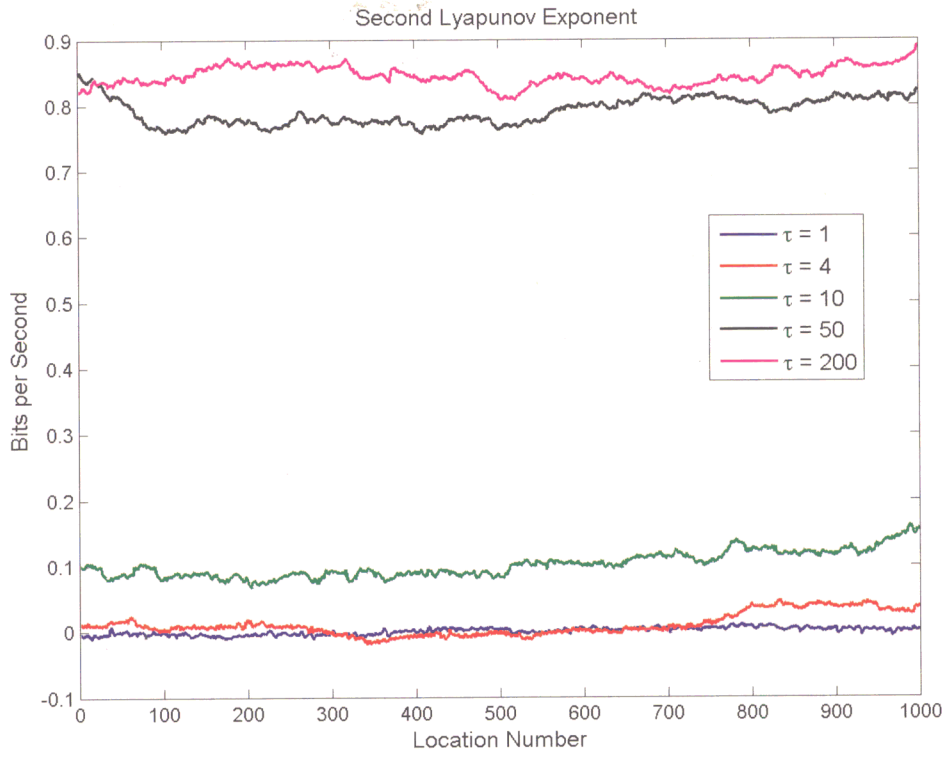


Fig. 12. The positive Lyapunov exponent of the Lorenz system along a trajectory. Green represents smaller local divergence rates and brown represents larger local rates. [Figure courtesy of Dr. J. Nese.]

local Lyapunov exponents along a long trajectory. In order to estimate the local Lyapunov exponents at a given location we consider our representative point in the reconstructed state space and the points in a small neighborhood around it. These neighbors have been introduced during the reconstruction at earlier or later times and play the role of small fluctuations about the reference point from which other "nearby" trajectories can be obtained. The motion of this neighborhood in a predefined time interval (called the decomposition length L) provides the information to estimate the Jacobian of these small deviations about the reference point, whose eigenvalues provide an estimate of the Lyapunov exponent spectrum at this location [Abarbanel *et al.*, 1991; Brown *et al.*, 1991]. By shifting the decomposition length one time step at a time we can obtain a record of any local Lyapunov exponent along the trajectory. The estimated exponents depend on the decomposition length and approach the true values of the system for a sufficiently long L. Figure 13 shows the first (top), the second (middle), and third (bottom) Lyapunov exponent along the trajectory (location number) for five different delays. In order to compare our results to those reported in the literature, the observable $x(t)$ used in this figure was obtained by integrating the Lorenz system with a time step of 0.05 and the decomposition length was assumed equal to 50 (in this case the first zero of the autocorrelation function is about 40 and the first minimum of the mutual information is 2). For this value of decomposition length the *expected* average Lyapunov exponents of the system are $\langle\lambda_1\rangle \approx 2$ bits/sec, $\langle\lambda_2\rangle \approx 0$ bits/sec, and $\langle\lambda_3\rangle \approx -18$ bits/sec [Abarbanel *et al.*, 1991; Brown *et al.*, 1991]. Note that, because of the integration time step used here, in order to derive the actual exponents from Fig. 13 we need to divide the values by 0.05. Accordingly, we find that from the above choices the one that is the closest to the expected values is $\tau = 1$. For this delay we estimate that $\langle\lambda_1\rangle \approx 0.085/0.05 = 1.7$ bits/sec, $\langle\lambda_2\rangle \approx 0.0/0.05 = 0$ bits/sec, and $\langle\lambda_3\rangle \approx -0.85/0.05 = -17$ bits/sec (note that this delay would correspond to $\tau = 5$ if the integration step were 0.01). From Fig. 13 we observe, as in the case with nonlinear prediction, that not any delay will provide useful results. A $\tau = 2$ (which corresponds to $\tau = 10$ when the integration step is 0.01) gives values slightly closer to the expected estimates (not shown in Fig. 13). However, for greater delays all Lyapunov exponents are overestimated significantly; a result that can lead

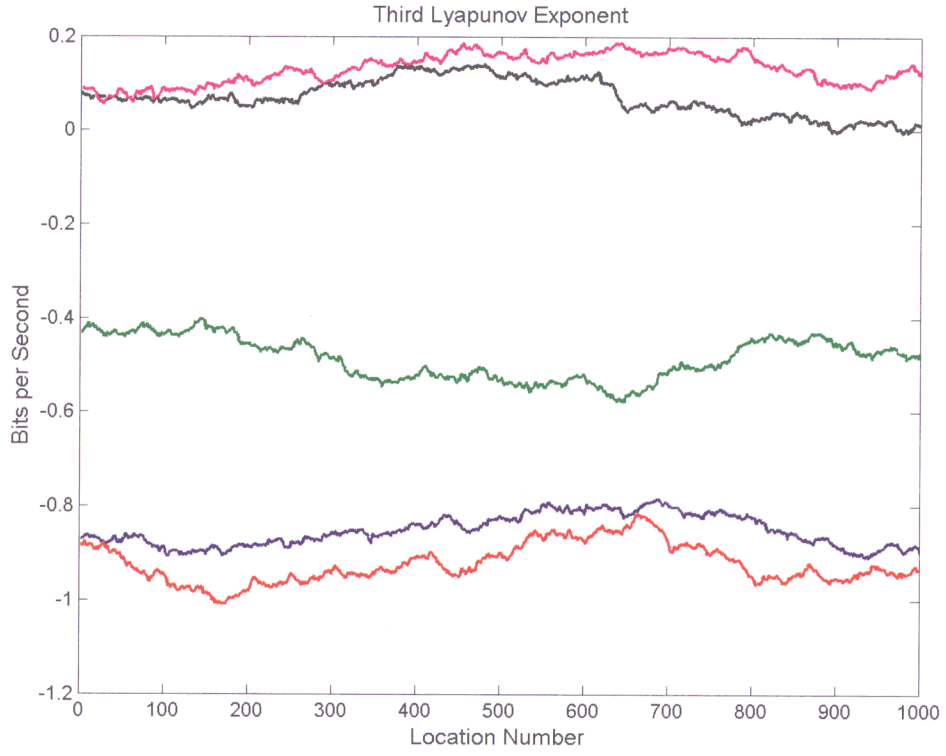


(a)



(b)

Fig. 13. The (a) positive, (b) zero, and (c) negative Lyapunov exponents of the Lorenz system as a function of the location number (i.e. 1,000 points along a trajectory).



(c)

Fig. 13. (*Continued*)

to false interpretation of the underlying physics. On the other hand, estimates based on smaller than the optimum delays are rather useful, probably for the same reason that very small delays result in very good prediction.

3. Concluding Remarks

When we reconstruct the dynamics from observables our first issue is to estimate the proper delay and the dimension of the attractor. Once this is done effectively, then we can embed the data in the appropriate space and perform nonlinear prediction and estimate the Lyapunov exponents. Here we show that an effective way to tackle the first problem is to take advantage of the fact that an increasing delay emulates the general properties of stretching and folding in strange attractors. In this case the proper delay is the one that approximates these properties best. We then show that the proper delay and embedding dimension can be simultaneously estimated by examining the *slope* versus $\log r$ plots and choose the optimum delay as the one resulting in the cleanest and widest plateau. There is no need to resort to measures such as the first minimum of the mutual information or the first zero of the autocorrelation function. Such measures are statistical and may not have a direct connection to the underlying physics and dynamics, which dictate the proper stretching and folding action of the attractor. Be this as it may, we also demonstrated that one can in effect get a very good idea of the proper embedding dimension even with very small delays, and that even with very small delays we can obtain very reliable dimension estimates, prediction and Lyapunov exponent estimates. Having said that, when we investigate nonlinearity in time series one has to apply all available approaches and research their collective behavior. With the computer power available today there is no reason not to thoroughly apply the methods described in this tutorial and arrive at a consistent and accurate result.

Acknowledgment

Part of this work was supported by NSF grant ATM-0552215.

References

Abarbanel, H. D. I., Brown, R. & Kennel, M. B. [1991] "Lyapunov exponents in chaotic systems: Their

importance and their evaluations using observed data," *Mod. Phys. Lett. B* **5**, 1347–1375.

Avnir, D., Biham, O., Lidar, D. & Malcai, O. [1998] "Is the geometry of nature fractal?" *Science* **279**, 39–40.

Brown, R., Bryant, P. & Abarbanel, H. D. I. [1991] "Computing the Lyapunov spectrum of a dynamical system from observed time series," *Phys. Rev. A* **43**, 2787–2806.

Farmer, J. D. & Sidorowich, J. J. [1987] "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845–848.

Grassberger, P. & Procaccia, I. [1983a] "Characterization of strange attractors," *Phys. Rev. Lett.* **50**, 346–349.

Grassberger, P. & Procaccia, I. [1983b] "Measuring the strangeness of strange attractors," *Physica D* **9**, 189–208.

Lorenz, E. N. [1963] "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130–141.

Mandelbrot, B. B. [1983] *The Fractal Geometry of Nature* (Freeman, NY).

Martinerie, J. M., Albano, A. M., Mees, A. I. & Rapp, P. E. [1992] "Mutual information, strange attractors, and the optimal estimation of dimension," *Phys. Rev. A* **45**, 7058–7064.

Nerenberg, M. A. H. & Essex, C. [1990] "Correlation dimension and systematic geometric effects," *Phys. Rev. A* **42**, 7065–7074.

Osborne, A. R. & Provenzale, A. [1989] "Finite correlation dimension for stochastic systems with power-law spectra," *Physica D* **35**, 357–381.

Packard, N. H., Crutchfield, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716.

Ruelle, D. [1981] "Chemical kinetics and differentiable dynamical systems," in *Nonlinear Phenomena in Chemical Dynamics*, eds. Pacault, A. & Vidal, C. (Springer-Verlag, Berlin), pp. 57–72.

Schreiber, T. & Schmitz, A. [1996] "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett.* **77**, 635–638.

Smith, L. A. [1988] "Intrinsic limits on dimension calculations," *Phys. Lett. A* **133**, 283–288.

Sugihara, G. & May, R. M. [1990] "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature* **344**, 734–741.

Takens, F. [1981] "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898, eds. Rand, D. & Young, L. S. (Springer-Verlag, Berlin), pp. 366–381.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. [1992] "Testing for nonlinearity in time series: The method of surrogate data," *Physica D* **58**, 77–94.

Tsonis, A. A. [1992] *Chaos: From Theory to Applications* (Plenum, NY).

Tsonis, A. A. & Elsner, J. B. [1992] "Nonlinear prediction as a way of distinguishing chaos from random fractal sequences," *Nature* **358**, 217–220.

Tsonis, A. A., Elsner, J. B. & Georgakakos, K. P. [1993] "Estimating the dimension of weather and climate attractors: Important issues about the procedure and interpretation," *J. Atmos. Sci.* **50**, 2549–2555.

Tsonis, A. A., Triantafyllou, G. N. & Elsner, J. B. [1994] "Searching for determinism in observed data: A review of the issues involved," *Nonlin. Process. Geophys.* **1**, 12–25.

Tsonis, A. A. & Elsner, J. B. [1995] "Testing for scaling in natural forms and observables," *J. Stat. Phys.* **81**, 869–880.

Tsonis, A. A. [1998] "Fractality in nature," *Science* **279**, 1614–1615.

Wales, D. J. [1991] "Calculating the rate of loss of information from chaotic time series by forecasting," *Nature* **350**, 485–488.