The analysis of observed chaotic data in physical systems

Henry D. I. Abarbanel

Department of Physics and Marine Physical Laboratory, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0402

Reggie Brown, John J. Sidorowich, and Lev Sh. Tsimring

Institute for Nonlinear Science, University of California at San Diego, Mail Code 0402, La Jolla, California 92093-0402

Chaotic time series data are observed routinely in experiments on physical systems and in observations in the field. The authors review developments in the extraction of information of physical importance from such measurements. They discuss methods for (1) separating the signal of physical interest from contamination ("noise reduction"), (2) constructing an appropriate state space or phase space for the data in which the full structure of the strange attractor associated with the chaotic observations is unfolded, (3) evaluating invariant properties of the dynamics such as dimensions, Lyapunov exponents, and topological characteristics, and (4) model making, local and global, for prediction and other goals. They briefly touch on the effects of linearly filtering data before analyzing it as a chaotic time series. Controlling chaotic physical systems and using them to synchronize and possibly communicate between source and receiver is considered. Finally, chaos in space-time systems, that is, the dynamics of fields, is briefly considered. While much is now known about the analysis of observed temporal chaos, spatio-temporal chaotic systems pose new challenges. The emphasis throughout the review is on the tools one now has for the realistic study of measured data in laboratory and field settings. It is the goal of this review to bring these tools into general use among physicists who study classical and semiclassical systems. Much of the progress in studying chaotic systems has rested on computational tools with some underlying rigorous mathematics. Heuristic and intuitive analysis tools guided by this mathematics and realizable on existing computers constitute the core of this review.

CONTENTS

I. INTRODUCTION

Interpretations of measurements on physical systems often turn on the tools one has to make those interpretations. In this review article we shall describe new tools for the interpretation of observations of physical systems where the time trace of the measured quantities is irregular or chaotic. These time series have been considered bothersome "noise" without the tools we shall describe

and, as such, have traditionally been taken as *contamination to be removed* from the interesting physical signals one sought to observe. The characterization of irregular, broadband signals, generic in nonlinear dynamical systems, and the extraction of physically interesting and useful information from such signals is the set of achievements we review here. The chaotic signals that we shall address have been discarded in the past as "noise," and among our goals here is to emphasize the valuable physics that lies in these signals. Further we want to indicate that one can go about this characterization by exploiting the structure in these chaotic time series in a systematic, nearly algorithmic, fashion. In a sense we shall describe new methods for the analysis of time series, but on another level we shall be providing handles for the investigation and exploitation of aspects of physical processes that could be simply dismissed as "stochastic" or random when seen with different tools. Indeed, the view we take in this review is that chaos is not an aspect of physical systems that is to be located and discarded, but is an attribute of physical behavior that is quite common and whose utilization for science and technology is just beginning. The tools we discuss here are likely also to be just the beginning of what we can hope to bring to bear in the understanding and use of this remarkable feature of physical dynamics.

The appearance of irregular behavior is restricted to nonlinear systems. In an autonomous linear system of f degrees of freedom, $\mathbf{u}(t) = [u_1(t), u_2(t), \ldots, u_f(t)]$, one has

$$\frac{d\mathbf{u}(t)}{dt} = \mathbf{A} \cdot \mathbf{u}(t) , \tag{1}$$

where $\mathbf{A}$ is a constant $f \times f$ matrix. The solution of this linear equation results in (1) directions in f-space along which the orbits $\mathbf{u}(t)$ shrink to zero—namely, directions along which the real part of the eigenvalues of $\mathbf{A}$ are negative, (2) directions along which the orbits unstably grow to infinity—namely, directions along which the real part of the eigenvalues of $\mathbf{A}$ are positive, and (3) directions where the eigenvalues occur in complex-conjugate pairs along with zero or negative real part. In any situation where the eigenvalue has a positive real part, this is a message that the linear dynamics is incorrect, and one must return to the nonlinear evolution equations that govern the process to make a better approximation to the dynamics.

Much of the work in nonlinear dynamics reported in the past has concentrated on learning how to *classify* the nonlinear systems by analyzing the output from known systems. These efforts have provided, and continue to provide, significant insights into the kinds of behavior one might expect from nonlinear dynamical systems as realized in practice and has led to an ability to evaluate now familiar quantities such as fractal dimensions, Lyapunov exponents, and other invariants of the nonlinear system. The capability for doing these calculations is strongest when we know the differential equations or

mappings of the dynamical system. When inferring such quantities from data alone, we find the issues more challenging, and this is the subject matter covered in this article.

At the outset it is appropriate to say that this is not, and actually may never be, an entirely settled issue. However, a certain insight and the establishment of certain general guidelines has been achieved, and that in itself is important. Only recently has the question of inferring from *measured chaotic time series* the properties of the system that produced those time series been addressed with any real vigor. The reason is partly that, until the classification tools were clearly established, it was probably too soon to tackle this harder problem. Moreover, by its nature, this challenge of "inverting" time series to deduce characteristics of the physical system—that is, model building for nonlinear dynamics—has required the development of tools beyond those of direct classification, and these too have taken time to evolve. While some general results are known and will be described here, much of the analysis of a physical system may be specific to that system. Thus the tools here should be regarded as a guide to the reader, not a strict set of rules to be followed.

This article is designed to bring to scientists and engineers a familiarity with developments in the area of model building based on signals from nonlinear systems. The key fact that makes this pursuit qualitatively different from conventional time series analysis is that, because of the nonlinearity of the systems involved, the familiar and critical tool of Fourier analysis does very little, if any, good in the subject. The Fourier transform of a linear system changes in what might be a tedious set of differential equations into an algebraic problem where decades of matrix analysis can be fruitfully brought to bear. Fourier analysis of nonlinear systems turns differential equations in time into integral equations in frequency space involving convolutions among the Fourier transforms of the dependent variables. This is rarely an improvement, so Fourier methods are to be discounted at the outset, though as an initial window through which to view the data, they may prove useful.

Having set aside at the outset the main analysis tool for linear signal processing, we must more or less reinvent signal processing for nonlinear systems. In rethinking the problem it is very useful to address the essential set of general issues faced by signal processors. We envision that the physicist will be presented with observations of one or perhaps, if fortunate, a few variables from the system and then be interested in deducing as much as possible about the characteristics of the system producing the time series—this is *system classification*. The next step would be creating a framework in which to predict, within limits determined from the data itself, the future behavior of the system or the evolution of new points in the system state or phase space—this is *model building and parameter estimation,* since the models that allow prediction are usually parametric.

Even before beginning the analysis, the practitioner has to address the question of how to deal with data contaminated by other signals or by the measurement process itself. This contamination can come from environmental variables unmeasured by the instruments or from properties of the observation instruments. In a general sense the contamination has broadband Fourier spectra with possible sharp lines embedded in the continuum. This general description will also be true of the Fourier spectrum of the chaotic signal, since it arises from nonperiodic motion and has a continuous broadband spectrum with possible sharp spectral lines in it. Separating these two spectrally similar phenomena is not a task for which standard time series analysis is well suited, but one which, using differing characteristics of the signal and contamination, we may address in the time domain. We shall indicate in sections of this article how one goes about this initial task in signal processing in chaos.

Broadly speaking, this article will follow the directions pointed out in Table I. This indicates the requirements on the physicist for creating some order from the time series presented, and it is clear that the same problems, with different solutions, are faced whether the system is linear or nonlinear.

In an ambitious fashion, once one can clean up dirty signals, classify the systems that give rise to those signals, and reliably predict the behavior of the system in the future, within limits set by the dynamics themselves, then consideration of control strategies for nonlinear systems is in order. This topic will be touched on in later sections of this review.

One important insight into dynamical systems is the role played by *information theory*. There is an intuitive notion that a dynamical system that has chaotic behavior is precisely a realization of Shannon's concept of an ergodic information source (Gallager, 1968; Shaw, 1981, 1984). This source is assumed to produce a sequence of symbols X_t: ..., X_{-1}, X_0, X_1, ... drawn from an alphabet $x = 0, 1, \ldots, J - 1$. This is for a discrete source with a finite number of states available to it, namely, any source we would realize on a finite-state digital computer or in an experiment. The distribution of orbit points in the alphabet $P(x)$ gives the probability for realizing the symbol x by measurements of the X_t. The entropy (in bits) of a joint set of measurements is defined using the joint probability $P(x_1, x_2, \ldots, x_N)$ via

$$H_N(X) = - \sum_{x_j} P(x_1, x_2, \ldots, x_N) \log_2[P(x_1, x_2, \ldots, x_N)] .$$

(2)

As N becomes large this entropy, divided by N, has a finite limit,

$$\lim_{N \to \infty} \frac{H_N(X)}{N} = h(X) ,$$

(3)

and this is the amount one learns about the source of the symbols by the sequence of measurements on the set X.

Clearly the source is no more than an abstract statement of the system producing our measurements, and the concept of entropy of a deterministic dynamical system, where no notion of probability is immediately evident, seems much the same as in the information-theoretic setting. This idea was realized over thirty years ago by Kolmogorov (1958), who gave a definition of this entropy which, in practice, is precisely the same as $h(X)$ above. Further, this entropy is an invariant of the orbits of the system, that is, independent of the initial conditions or the specific orbits observed. Additional work on this quantity was done by Sinai (1959), and the quantity $h(X)$ has become known as the Kolmogorov-Sinai (or KS) entropy of a dynamical system. Strengthening this connection is a theorem of Pesin (1977) equating the KS entropy to the sum of the positive Lyapunov exponents. If a dynamical system has no positive Lyapunov exponents, it does not possess chaos in the usual sense.

These comments indicate the connection between dynamical systems and information theory, but do not explain the interest for someone concerned with the analysis and modeling of chaotic data. The KS entropy gives a precise statement of the *limits to predictability* for a nonlinear system. Think of a phase space for the system which is reconstructed or, if one actually knows the differential equations or mapping, precise. Any realizable statement about the system is known only to some precision in this state space. Within that precision or resolution we cannot distinguish at a given time $t = 0$ between two state-space points lying in the resolution cell. A nonlinear system with positive $h(X)$ is intrinsically unstable, however, and the two points unresolvable at $t = 0$ will move after some time T to differing parts of the phase space. $2^{[h(X)T]}$ is a measure of the number of states occupied by the system after the passage of a time T (Gallager, 1968; Rabinovich, 1978). When this number of states is approximately the total number of states available for the orbits of the system, then all prediction of the future of an orbit becomes untenable. This occurs in approximately $T \approx 1/h(X)$. One still has statistical information about the system that is quite useful and interesting, but knowledge of the evolution of a specific orbit is lost. Thus, $h(X)$ becomes an object of considerable interest for the physicist. Since it is the sum of the positive Lyapunov exponents, one can use the time series itself to determine how predictable it is.

A. Observed chaos

Actual observations of irregular evolution that we consider are typically *measurements of a single scalar observable at a fixed spatial point*: call the measured quantity $s(t_0 + n\tau_s) = s(n)$, where t_0 is some initial time and τ_s is the sampling time of the instrument used in the experiment. $s(n)$ could be a voltage or current in a nonlinear circuit or a density or velocity in a fluid dynamic or plasma dynamic experiment or could be the temperature at some location in a laboratory container or an atmospher-

ic or oceanic field experiment. The full structure of some of these measurements lies in the observation of a field of quantities, namely, $s(\mathbf{x}, t_0 + n\tau_s)$ with $\mathbf{x} = (x, y, z)$ the spatial coordinates. Most of this review will focus on methods for the analysis of scalar measurements at a fixed spatial point. This is done for the simple reason that the subject has matured to a point where one can productively review what is known with the hope that the methods can then become part of every physicist's experimental analysis toolkit in much the same way that Fourier analysis now is part of that toolbox. At the end of our discussion we shall have some things to say about spatio-temporal systems, but the developments there which parallel what we know about purely temporal measurements remain to be completed.

The first task will be to explain how to begin with this set of scalar measurements and to *reconstruct the system phase space* from it. This requires some explanation. In fluid dynamics, for example, the system phase space is infinite dimensional when the description of the fluid by the Navier-Stokes partial differential equations is correct. We do not reestablish from the $s(n)$ an infinite number of degrees of freedom, but we take the point of view that, since the observations of a dissipative system such as a fluid lie on a geometric object of much smaller dimension than that of the original state space, we must seek methods for modeling the evolution on the attractor itself and not evolution in the full, infinite-dimensional, original phase space. In some circumstances this dimension may be quite small, as low as five or so. The techniques we shall discuss are then directed toward the description of the *effective* time-asymptotic behavior of the observed dynamics. In rare circumstances this might be expected to yield the original differential equations.

The measurements $s(n)$ and the finite number of quantities formed from them in the course of the analysis provide coordinates for the effective finite-dimensional space in which the system is acting after transients have died out. This gives rise to two approaches to modeling the dynamics from the point of view of the physicist interested in understanding the system producing the measured signals:

● The first approach is that of making physical models in the coordinates $s(n)$ and quantities made from the $s(n)$—and doing all this in the finite-dimensional space where the observed orbits evolve. This description of the system differs markedly from the original differential equations, partial or ordinary, which one might have written down from first principles or from arguments about the physics of the experiment. Verification of the model by future experiment or different features of the experiment proceeds as usual, but the framework of the models is quite different from the usual variations of Newton's laws, which constitute most physics model making. The net result of this will be a nonlinear relationship or map from values of the observed variables $s(n)$ to their values $s(n + T)$ some time T later. If one of

the Lyapunov exponents is zero, as determined by the data using methods that we shall describe later, then one could construct differential equations to capture the dynamics. This suggests we are confident that extrapolation to $\tau_s \rightarrow 0$, implicit in the idea of a differential equation rather than a finite time map, captures all the relevant frequencies or time scales or degrees of freedom of the source. In either case—map or differential equation—no rigid rules appear to be available to guide the modeler.

● The second approach is more traditional and consists of making models of the experiment in the usual variables of velocity, pressure, temperature, voltage, etc. The job of the physicist is then to explore the implications of the model for various *invariant* quantities under the dynamics, such as the dimensions of the attractor and the Lyapunov exponents of the system. These invariants, which we shall discuss below, are then the testing grounds of the models. The critical feature of the invariants is that they are not sensitive to initial conditions or small perturbations of an orbit, while individual orbits of the system are exponentially sensitive to such perturbations. The exponential sensitivity is the manifestation of the instability of all orbits in the phase space of the dynamics. Because of this intrinsic instability no individual orbit can be compared with experiment or with a computed orbit, since any orbit is effectively uncorrelated with any other orbit, and numerical roundoff or experimental precision will make every orbit distinct. What remains, and what we shall emphasize in this review, is a *statistics of the nonlinear deterministic system* which produces the orbits. An additional set of invariants, *topological invariants,* are also quite useful with regard to this task of characterizing physical systems and identifying which models of the system are appropriate (Mindlin *et al.*, 1991; Papoff *et al.*, 1992), and we shall discuss them as well. Topological invariants are distinct from the better known *metric* invariants mentioned above in that they do not rest on the determination of distance or metric properties of orbit points on an attractor. They appear to be quite robust in the presence of "noise" and under changes in system parameters.

The statistical aspect of chaotic dynamics was discussed and its importance stressed in an earlier article in this journal by Eckmann and Ruelle (1985). Our review now is in many ways an update of features of that earlier article with an emphasis on practical aspects of the analysis of experimental data and model building for prediction and control. Of course, we also include many developments unanticipated by the earlier review, but we certainly recommend it as an introduction to many of the topics here, often from a rather more mathematical viewpoint than we project.

Many items we shall mention here, especially in the sections on signal separation, control of chaotic systems, and synchronization of chaotic systems, have direct practical implications which move beyond the usual preoccu-

pation of physicists with model building, prediction, and model verification. The authors expect that many of these topics will find further development in the engineering literature, and we cannot anticipate the full spectrum of those developments. We mention some throughout the review, since we think it quite important to underline the utility of chaotic states of a physical system in the hopes that physicists will seek new phenomena suggested by chaos and uncover further interesting physical behavior connected with this disordered, but not random, feature of deterministic systems.

Just a note before we move ahead: we interchangeably denote the space in which orbits of dynamical systems evolve as *state space* or *phase space*. State space is more commonly used in the engineering literature, while phase space is more common in the physics literature. In physics, phase space is also used to denote the canonical set of coordinates of Lagrange or Hamilton for conservative systems. We do not use phase space in this restricted sense, nor does our use of "phase" refer to the argument of a complex variable.

B. Outline of the review

The main body of the review begins in the next section with some details of approaches to the analysis of time series, first from the traditional or linear point of view. Then we move on to the idea of reconstructing the phase space of the system by the use of time delays of observed data, $s(n)$. In the subsequent section we discuss methods of determining the time delay to use in the practical reconstruction of phase space, and in doing so introduce the ideas of information theory in greater detail. Having established the time delay to use, we move on to a discussion of the choice, from analysis of the observed data, of the dimension of the phase space in which we must work to unfold the attractor on which the dynamics evolves.

With that we end our discussion of the state space in which we are to view the dynamics and we move on to the issue of classifying the dynamics that underlies the observations. The classifiers in a linear system are the resonant Fourier frequencies of the system. In a nonlinear system we have statistical aspects of the deterministic chaotic evolution which we can use for this same purpose. The quantities invariant under the dynamics are used as classifiers, and we shall discuss the physical and mathematical meaning of some of the most widely utilized of these invariants.

After we have established ways to classify the physical system leading to the observations $s(n)$, we move on to a discussion of model building to allow the codification of the measurements into a simple evolution rule. This is an easy task, as we shall see, if all we desire is numerically accurate predictors. If we wish insight into the physical processes needed to produce, reproduce, and predict the signals we observe, then the task is harder, and no clear-cut method has emerged, or for that matter may ever emerge. Our discussion will cover various ideas on this problem, all of which have limitations or unattractive features, however successful they may be in a numerical sense.

The next topic we address is that of separating the desired "clean" chaotic signal from the effects of contamination by the environment through which it passes on its way from the source to the measurement device or contamination by properties of the measurement process itself. This contamination is often called "noise," and the task of signal separation goes under the label of "noise reduction" in common parlance. The ideas involved in signal separation carry one far beyond traditional views of noise reduction, and we shall explore some of these as well. In any case, the methods one must employ are to be used in the time domain of reconstructed phase space, and they differ in detail and conception from Fourier-based methods, which are much more familiar. The penultimate topic we discuss is that of controlling chaotic systems by utilizing properties of the structure of the attractor to move the system from its "natural" autonomous chaotic state to motion about an unstable periodic motion—unstable from the point of view of the autonomous system, but stable from the point of view of the original system plus a control force.

Finally we turn to aspects of the experimental analysis of spatio-temporal systems with disorder or chaotic behavior. In this area, as noted earlier, we know much less, so much of what we present is speculative. Clearly a review of this subject five or so years from now will, without question, supplant this aspect of the review. Nonetheless, we felt it important to end the review with a glimpse, however imperfect, of the problems that will dominate the future of this kind of work. In any case, the analysis of chaos of fields is essential for doing real physics in chaotic systems, so the problems we outline and the potential solutions we discuss certainly will affect our ability to work with chaotic fields.

II. SIGNALS, DYNAMICAL SYSTEMS, AND CHAOS

In the analysis of signals from physical systems we assume from the outset that a dynamical system in the form of a differential equation or a discrete-time evolution rule is responsible for the observations. For continuous-time dynamics we imagine that there is a set of f ordinary differential equations for variables $\mathbf{u}(t) = [u_1(t), u_2(t), \ldots, u_f(t)]$,

$$\frac{d\mathbf{u}(t)}{dt} = \mathbf{G}(\mathbf{u}(t)) , \qquad (4)$$

where the *vector field* $\mathbf{G}(\mathbf{u})$ is always taken to be continuous in its variables and also taken to be differentiable as often as needed. When time is discrete, which is the realistic situation when observations are only sampled every τ_s, the evolution is given by a map from vectors in R^f to other vectors in R^f, each labeled by discrete time: $\mathbf{u}(n) = \mathbf{u}(t_0 + n\tau_s)$,

$$\mathbf{u}(n+1)=\mathbf{F}(\mathbf{u}(n)) \ . \tag{5}$$

The continuous-time and discrete-time views of the dynamics can be connected by thinking of the time derivative as approximated by

$$\frac{d\mathbf{u}(t)}{dt} \approx \frac{\mathbf{u}(t_0+(n+1)\tau_s)-\mathbf{u}(t_0+n\tau_s)}{\tau_s} \ , \tag{6}$$

which would lead to

$$\mathbf{F}(\mathbf{u}(n))\approx\mathbf{u}(n)+\tau_s\mathbf{G}(\mathbf{u}(n)) \ , \tag{7}$$

as the relation between the continuous and discrete dynamics.

Discrete dynamics can also arise by sampling the continuous dynamics as it passes through a hyperplane in the f-dimensional space. This method of Poincaré sections leads to an $(f-1)$-dimensional discrete system that is not easily related in an analytic fashion to the continuous system behind it. Often qualitative properties of the continuous system are manifest in the Poincaré section: an orbit that is a recurrent point (fixed point) in the section is a periodic orbit in the continuous flow.

The number of degrees of freedom or, equivalently, the dimension of the state space of a dynamical system is the same as the number of first-order differential equations required to describe the evolution. Partial differential equations generalize the number of degrees of freedom from a finite number, f, to a continuum labeled by a vector $\mathbf{x}=(x_1,x_2,\ldots,x_D)$. For a field with K components $\mathbf{u}(\mathbf{x},t)=[u_1(\mathbf{x},t),u_2(\mathbf{x},t),\ldots,u_K(\mathbf{x},t)]$, we write the differential equation

$$\frac{\partial\mathbf{u}(\mathbf{x},t)}{\partial t}=\mathbf{G}(\mathbf{u}(\mathbf{x},t)) \ , \tag{8}$$

and the labels of the dynamical variable now are both continuous (namely, $\mathbf{x}$ and t) and discrete. Broadly speaking the dynamics of fields differs from that of low-dimensional dynamical systems only because the number of degrees of freedom has become continuous. Many new phenomena can appear because of this, of course, but when we think of the analysis of real data it will certainly be finitely sampled in space $\mathbf{x}$ as well as in time, and we may consider the partial differential equation to have been reduced to a very large number of ordinary differential equations for purposes of discussion.

There is one important distinction between discrete dynamics that comes from a surface of a section of a flow (continuous-time dynamics) and that from finite time sampling of the flow. In the latter case we can associate the direction of the vector field with evolution which can be explored by the system. In the former case evolution along the direction of the vector field is off the surface of the section and is not available for study. This difference will be manifest when we discuss Lyapunov exponents below.

The vector field $\mathbf{F}(\mathbf{u})$ has parameters that reflect the external settings of forces, frequencies, boundary conditions, and physical properties of the system. In the case of a nonlinear circuit, for example, these parameters would include the amplitude and frequency of any driving voltage or current, the resistance or capacitance or inductance of any lumped linear element, and the IV or other characteristics of any nonlinear elements.

The dynamical system

$$\mathbf{u}(n+1)=\mathbf{F}(\mathbf{u}(n)) \tag{9}$$

is generally not volume preserving in f-dimensional space. In evolving from a volume of points $d^f u(n)$ to a volume of points $d^f u(n+1)$ the volume changes by the determinant of the Jacobian

$$DF(\mathbf{u})=\det\left[\frac{\partial\mathbf{F}(\mathbf{u})}{\partial\mathbf{u}}\right] \ . \tag{10}$$

When this Jacobian is unity, which is a very special occurrence and is usually associated with Hamiltonian evolution in a physical setting, phase-space volume remains constant under the evolution of the system. Generally the Jacobian has a determinant less than unity, which means volumes in phase space shrink as the system evolves. The physical origin of this phase-space volume contraction is dissipation in the dynamics.

As the physical parameters in the vector field are varied, the behavior of $\mathbf{u}(t)$ or $\mathbf{u}(n)$, considered as an initial-value problem starting from $\mathbf{u}(t_0)$ or $\mathbf{u}(0)$, will change in detail as t or n becomes large. Some of these changes are smooth and preserve the geometry of the set of points in R^f visited by the dynamics. Some changes reflect a sudden alteration in the qualitative behavior of the system. To have an explicit example to discuss, we consider the three differential equations of Lorenz (1963)[1]

$$\frac{dx(t)}{dt}=\sigma(y(t)-x(t)) \ ,$$

$$\frac{dy(t)}{dt}=-x(t)z(t)+rx(t)-y(t) \ , \tag{11}$$

$$\frac{dz(t)}{dt}=x(t)y(t)-bz(t) \ .$$

These equations are derived from a finite mode truncation of the partial differential equations describing the thermal driving of convection in the lower atmosphere. The ratio of thermal to viscous dissipation is called σ, the ratio of the Rayleigh number to that critical Rayleigh number where convection begins is called r, and the dimensionless scale of a convective cell is called b. The parameters σ and b are properties of the system and the geometry—planar here—while the parameter r is deter-

---

[1]In a later paper (Lorenz, 1984) there is a delightful description of how he came to consider this model. He was concerned with linear prediction of nonlinear models and used this three-degree-of-freedom model as a testing ground for this study. He concluded that the idea of linear prediction for this kind of system would not do.

mined by geometry, material properties, and the strength of the driving forces. The physically interesting behavior of the Lorenz system comes as r is varied, since that represents variations in the external forces on the "atmosphere." The derivation of the three ordinary differential equations shows that they provide a representation of thermal convection only near $r \approx 1$ (Salzmann, 1962),[2] but we, as have many others, have adopted them as a model of low-dimensional chaotic behavior, and we shall push the parameters completely out of the appropriate physical regime.

When r is less than unity, all orbits tend to the fixed point of the vector field,

$$G_1(x,y,z) = -\sigma x + \sigma y ,$$

$$G_2(x,y,z) = -y + rx - xz , \qquad (12)$$

$$G_3(x,y,z) = -bz + xy ,$$

at $(x,y,z) = (0,0,0)$. This represents steady thermal conduction. As r increases beyond unity, the state $(0,0,0)$ is linearly unstable, and the vector field has two symmetric linearly stable fixed points at $(x_+,y_+,z) = (\pm\sqrt{b(r-1)}, \pm\sqrt{b(r-1)}, r-1)$. For all $r < 1$, the geometry of the time-asymptotic state of the system is the same, namely, all initial conditions tend to $(0,0,0)$. For $r > 1$ and until r reaches $r_c = \sigma(\sigma+b+3)/(\sigma-b-1)$, orbits end up at (x_+,y_+,z) depending on the value of $(x(0),y(0),z(0))$. The geometry of the final state of the orbits is the same, namely, all volumes of initial conditions shrink to zero-dimensional objects. These zero-dimensional limit sets change with parameters, but retain their geometrical nature until $r = r_c$. When $r > r_c$, the situation changes dramatically. The two fixed points of the vector field at (x_+,y_+,z) become linearly unstable, and no stable fixed points remain. As r is increased from r_c, the time-asymptotic state emerging from an initial condition will either perform irregular motion characterized by a broad Fourier spectrum or undergo periodic motion.

The analysis of the sequence of allowed states in a differential equation or a map as parameters of the system are varied is called bifurcation theory. We shall not dwell on this topic as it is covered in numerous monographs and literature articles (Guckenheimer and Holmes, 1983; McKay, 1992; Crawford, 1991; Crawford and Knobloch, 1991). What is important for us here is that as we change the ratio of forcing to damping, represented by the parameter r in the Lorenz equations, the physical system undergoes a sequence of transitions from regular to more complex temporal evolution. The physical picture underlying these transitions is the requirement of passing energy through the system from whatever forcing is present to the dissipation mechanism for the particular system. As the amount of energy flux

increases, the system chooses different geometrical configurations in phase space to improve its ability to transport this energy. These configurations in the early stages frequently involve the appearance of higher Fourier modes than would be dictated by the symmetry of the driving forces and the geometry of the container. In a fluid heated from below the transition from conduction of heat to convection is an example of this (Chandrasekhar, 1961). The sequence of bifurcations that would be suggested by these first instabilities and changes in geometry does not persist. After a few new frequencies have been induced in the system by increasing the forcing to dissipation ratio, the system generically undergoes a structural transition to "strange" behavior in which the entire phase space is filled with unstable orbits. These unstable orbits do not go off to infinity in the phase space, as a rule, since in addition to the instabilities one has dissipation which limits the excursions of physical variables. The resulting stretching by instabilities and folding by dissipative forces is at the heart of the "strange" behavior we call chaos. When the forcing-to-dissipation ratio is small and regular behavior is observed, the time-asymptotic orbits evolve from whatever is the initial condition to motion on an attracting set of points that is regular and is composed of simple geometric figures such as points or circles or tori. The dimension of these geometric figures is integer and, because of the dissipation, smaller than the dimension f of the underlying dynamics. When the system becomes unstable, the orbits are attracted from a wide basin of initial conditions to a limiting set, which is fractional dimensional and has an infinite number of points. The complex appearance of the time series in such a situation is due to the *nonperiodic* way in which the orbit visits regions of phase space while moving along the geometric object we now call a strange attractor.

In Fig. 1 we display the time series for $x(t)$ which results from solving the Lorenz equations using a fourth-order Runge-Kutta integrator for time step $dT = 0.01$.
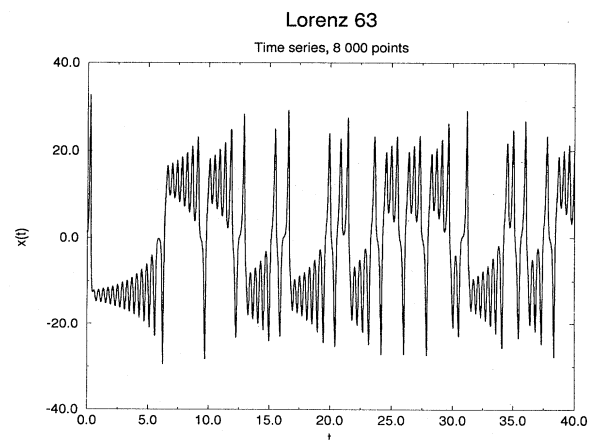

FIG. 1. Chaotic time series $x(t)$ produced by Lorenz (1963) equations (11) with parameter values $r = 45.92$, $b = 4.0$, $\sigma = 16.0$.

---

[2] The same equations were derived in the context of laser physics by Orayevsky *et al.* (1965).

The parameter values were $r = 45.92$, $b = 4.0$, and $\sigma = 16.0$. In Fig. 2 we plot the orbit coming from the three degrees of freedom $[x(t), y(t), z(t)]$ in a three-dimensional space. The familiar and characteristic structure of the long-term orbit of the Lorenz system—the Lorenz attractor—is apparent.

Much of the work on dynamical systems over the past fifteen years has focused on the analysis of specific physically, biologically, or otherwise interesting differential equations or maps for purposes of illuminating the various kinds of behavior allowed by nonlinear evolution equations. The sets of points in R^f visited by the orbits of such systems is probably not yet completely classified from a mathematical point of view, but the broad and useful description suggested above, as the ratio of forcing to dissipation is varied, has emerged. The details of the bifurcations involved in the transition from regular behavior, where the limit sets are points or periodic orbits, to irregular behavior such as that exhibited by the Lorenz model are often complicated. They are of interest in any specific physical setting, of course, but we shall deal with situations in which the system has arrived in a state where the parameter settings lead to irregular behavior characterized by a broad Fourier power spectrum, perhaps with some sharp lines in it as well.

This broad power spectrum is a first indication of chaotic behavior, though it alone does not characterize chaos. For the Lorenz model in the parameter regime indicated, the Fourier power spectrum is shown in Fig. 3. It is broad and continuous and falls as a power of frequency. Throughout this review we shall also work with data from a nonlinear circuit built by Carroll and Pecora at the U.S. Naval Research Laboratory (Carroll and Pecora, 1991) as an example for many of our algorithms and arguments. The time series for this circuit is taken
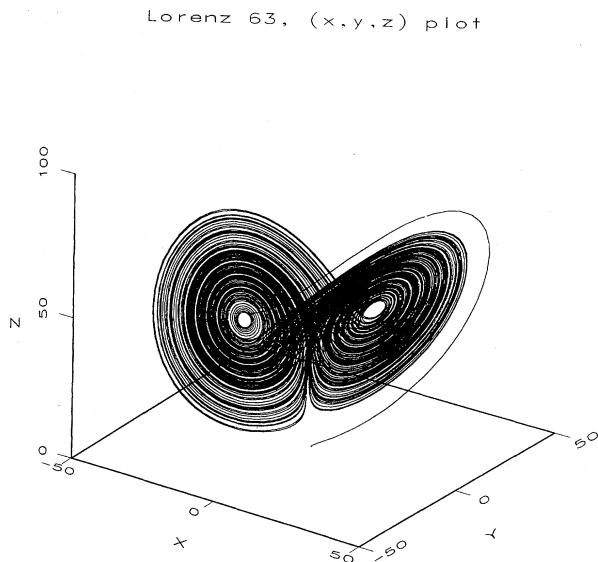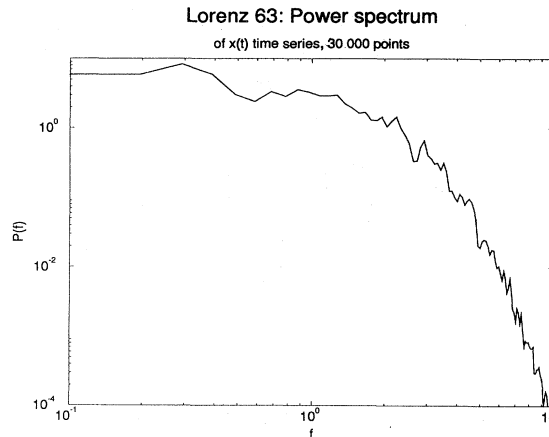

FIG. 3. Power spectrum of the Lorenz data $x(t)$.

as a tapped voltage from a location indicated in their original paper with $\tau_s = 0.1$ ms. The voltage as a function of time is shown in Fig. 4 and the corresponding Fourier spectrum in Fig. 5. Anticipating things to come, we plot in Fig. 6 a reconstruction of the phase space or state space of the hysteretic circuit orbits. The coordinates are the measured voltage, the same voltage 0.6 ms later and the same voltage 1.2 ms later. In contrast to the irregular voltage time trace in Fig. 4 and the Fourier spectrum of this voltage in Fig. 5, the phase-space plot demonstrates simplicity and regularity. These data indicate that the circuit is a candidate for a chaotic system, and we shall systematically investigate other features of the system from this voltage measurement. It is important to note that the time series alone does not determine whether the signal is chaotic. A superposition of several incommensurate frequencies will result in a complicated time trace but a simple line spectrum in Fourier space.

Linear analysis is not well tuned to working with signals with continuous, broad power spectra. Since stable linear dynamics produces either zero frequency, that is,
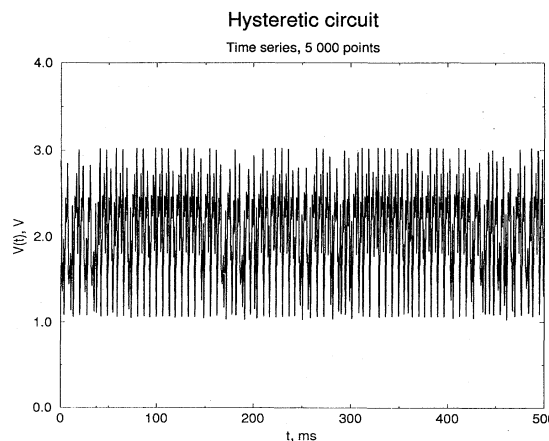

FIG. 2. Lorenz attractor in three-dimensional phase space $(x(t), y(t), z(t))$.

FIG. 4. Time series of voltage $V(t)$ produced by a hysteretic circuit (Carroll and Pecora, 1991).

Hysteretic circuit: Power spectrum
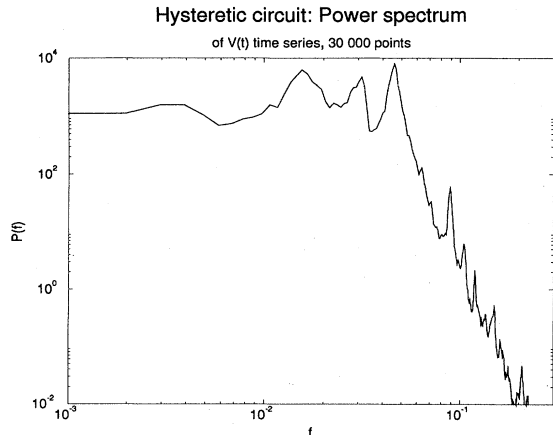
of V(t) time series, 30 000 points

FIG. 5. Power spectrum of the time series $V(t)$ from the hysteretic circuit.

fixed points, or oscillations at a few frequencies, it is not surprising that the use of linear models is not likely to give much insight into the behavior of signals arising from nonlinear systems. Traditional autoregressive/moving-average models (Oppenheim and Schafer, 1989), then, are not going to do very well in the description of chaotic signals. Indeed, if one tries to use linear models to fit data from chaotic sources and chooses to fit the unknown parameters in the model by a least-squares fit of the model to the observations, the rms error in the fit is about the same size as the attractor; that is, nothing has been accomplished. If one tries a *local linear fit* connecting neighborhoods in the state space of the system (this is certainly one of the simplest nonlinear models), the rms error can essentially be made as small as one likes depending on the quantity of data and size of the neighborhoods. This is not to suggest that local linear models answer all questions—indeed, we shall see below this is not the case—but that even small excur-

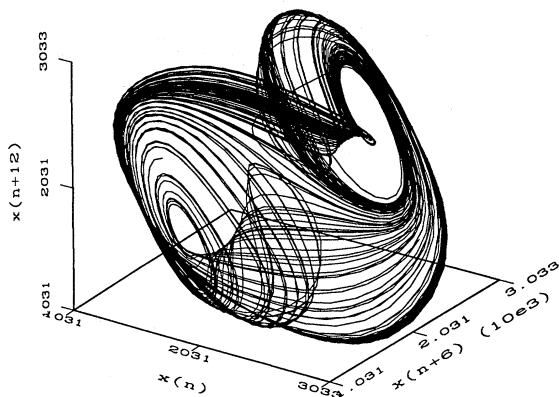Hysteretic Circuit; 64 000 Points

FIG. 6. Hysteretic time series embedded in a three-dimensional phase space: $[V(t), V(t+6), V(t+12)]$. See Fig. 4.

sions from global linear modeling can significantly improve performance in modeling nonlinear signals.

III. ANALYZING MEASURED SIGNALS—LINEAR AND NONLINEAR

The tasks faced by the analyst of signals observed from linear or nonlinear systems are really much the same. The methods for the analysis are substantially different. This section is devoted to a discussion of the similarities and differences in these requirements. The discussion is centered around Table I.

A. Signal separation

The very first job of the physicist is to *find the signal*. This means identifying the signal of interest in an observation that is possibly contaminated by the environment, by perturbations on the source of the signal, or by properties of the measuring instrument. In a graphic way we can describe the process of observing as (1) production of the signal by the source, potentially disturbed during the transmission, (2) propagation of the signal through some communications channel, potentially contaminated by the environment, and then (3) measurement of the signal at the receiver, potentially disturbed during the reception and potentially acting as a filter on the properties of the transmitted signal of interest.

The problems in finding the signal are common to the linear and the nonlinear observer. The linear observer has the advantage of being able to assume that the source emits spectrally sharp monochromatic signals which might be contaminated by broadband interference. The separation of the signal of interest from the unwanted background becomes an exercise in distinguishing narrowband signals in a broadband environment. Methods for this (Van Trees, 1968) are easily fifty years old and quite well developed.

Indeed, the main issue is really *signal separation*, and in order to carry out the processing required we must establish some difference between the information-bearing signal and the interference. In the case of a narrowband signal in a broadband environment, the difference is clear and Fourier space is the right domain in which to perform the separation. Similarly if the signal and the contamination are located in quite different bands of the spectrum, the Fourier techniques are certainly indicated. In the case of signals from chaotic sources, both the signal of interest and the background are typically broadband, and Fourier analysis will not be of much assistance in making the separation. In the first entry of the nonlinear side of Table I we identify several methods for separating the signal from its contamination. All of them rest on the idea that the signal of interest is low- or few-dimensional chaos with specific geometric structure in its state space. The geometric structure in the phase space is characteristic of each chaotic source, so we can use

this to effect the separation of that signal from others. When the contamination is "noise," which we shall see as an example of high-dimensional chaos, the signal separation problem is often called "noise reduction," but it remains at heart a separation problem. In nonlinear signal separation, three cases seem easy to distinguish. Each of these will be discussed in some detail in Sec. VII:

● *We know the dynamics.* We know the evolution equations $u(n+1)=F(u(n))$ and can use this knowledge to extract the signal satisfying the dynamics from the contaminating signals.

● *We know a signal from the system.* We have measured some signal from the system of interest at some earlier time. We can use this earlier measurement as a reference signal to establish the statistics of the evolution on the attractor and use these statistics to separate a subsequent transmission of the chaotic signal from contamination of the kind indicated just above.

● *We know nothing.* We know only a single instance of measuring a signal from the physical source. In this case

we are "flying blind" and must make models of the signal which allow us to establish what part of our observations are deterministic in low-dimensional phase space and what part a result of processes in higher-dimensional state space and thus to be discarded. It should be clear that this is much more problematic than the two other scenarios, but it may often be the realistic situation.

B. Finding the space

If we now suppose we have found the signal of interest, we may move on to establishing the correct space in which to view the signal. If the source of our signal is a linear physical process, perhaps a drum being beaten in a linear range or a seismic measurement far from the epicenter, so the displacements are very small, then the natural space in which to view the signal is Fourier space. If the source is invariant under translations in time, then sines and cosines are the natural basis functions in which to expand the problem, and the display of the information in the measurement will be most effective in Fourier

TABLE I. Connection between the requirements in linear and nonlinear signal processing. The goals are much the same in each case; the techniques are quite different.

Linear signal processing		Nonlinear signal processing
	Finding the signal	
Noise reduction; detection Separate broadband noise from narrowband signal using different spectral characteristics. If system is known, make matched filter in frequency domain.		Noise reduction; detection Separate broadband signal from broadband noise using deterministic nature of signal. If system is known or observed, make "matched filter" in time domain. Use dynamics or invariant distribution and Markov transition probabilities.
	Finding the space	
Fourier transforms Use Fourier space methods to turn differential equations or recursion relations into algebraic forms. $x(n)$ is observed; $x(f)=\sum x(n)\exp[i2\pi nf]$ is used.		Phase-space reconstruction Using timelagged variables, form coordinates for the phase space in d dimensions: $y(n)=[x(n),x(n+T), \ldots ,x(n+(d-1)T)]$ How are we to determine d and T? Mutual information; False neighbors.
	Classify the signal	
Sharp spectral peaks		Invariants of orbits. Lyapunov exponents; Various fractal dimensions; Topological invariants; Linking numbers of unstable periodic orbits
Resonant frequencies of the system Quantities independent of initial conditions		Quantities independent of initial conditions
	Make models, predict	
$x(n+1)=\sum c_j x(n-j)$ Find parameters c_j consistent with invariant classifiers (spectral peaks)		$y(n)\rightarrow y(n+1)$ as time evolution $y(n+1)=F[y(n),a_1,a_2, \ldots ,a_p]$ Find parameters a_j consistent with invariant classifiers (Lyapunov exponents, dimensions). Find dynamical dimensions d_L from data.

space. If the source has transients or significant high-frequency content with many localized events (in the time domain), then a linear transform, such as wavelets, that is adapted to such localized events will be more useful. Under our assumption that the underlying physical processes are governed by differential equations, we would expect them to have time-independent coefficients when the process is stationary, so Fourier space is again suggested, as it will reduce the differential equations to algebraic equations, which are much more tractable. Much of the contemporary toolkit of the signal processor rests on this ability to move to frequency domain and perform matrix manipulations on that representation of the measurements.

In the case of a nonlinear source, we are not likely to find any simplification from Fourier-transforming the scalar measurements $s(n)$, since the processes that give rise to chaotic behavior are fundamentally multivariate. So we need to reconstruct the phase space of the system as well as possible using the information in $s(n)$. This reconstruction process will result in vectors in a d-dimensional space. How we choose the components of the d-dimensional vectors and how we determine the value of d itself now become central topics. The answer will lie in a combination of dynamical ideas about nonlinear systems as generators of information and geometrical ideas about how one unfolds an attractor using coordinates established on the basis of their information-theoretic content. The result of the operations is a set of d-dimensional vectors $\mathbf{y}(n)$ which replace the scalar data we have observed.

C. Classification and identification

With a clean signal and the proper phase space in which to work, we can proceed to ask and answer interesting questions about the physical source of the observations. One critical question we need to address is that of classifying or identifying the source. In the case of linear systems we have Fourier analysis to fall back on again. The locations of the sharp lines in the Fourier spectrum are characteristic of the physical system being measured. If we drive the linear system harder, this will result in more power under each of these lines, and if we start the system at a different time, the phase of the signal will be altered, but the location of the Fourier peaks will remain unchanged. Thus the characteristic frequencies, usually called resonant frequencies for the linear system, are *invariants* of the dynamics and can be used to classify the physics. If we see a particular set of frequencies emerging from a driven linear system, we can recognize that set of frequencies at a later date and distinguish the source with confidence. Recognizing the hydrogen spectrum and codifying it with the Rydberg law is a classic (no pun intended) example of this. Similarly classifying acoustic sources of their spectral content is a familiar operation.

Nonlinear systems do not have useful spectral content when they are operating in a chaotic regime, but there are other invariants that are specific in classifying and identifying these sources. These invariants are quantities that are unchanged under various operations on the dynamics or the orbit. The unifying ingredient is that they are all unchanged under small variations in initial conditions. Some of the invariants we discuss have the property of being unchanged under the operation of the dynamics $\mathbf{x} \rightarrow \mathbf{F}(\mathbf{x})$. This guarantees that they are insensitive to initial conditions, which is definitely not true of individual orbits. Some of the invariants are further unchanged under any smooth change of coordinates used to describe the evolution, and some, the topological invariants, are purely geometric properties of the vector field describing the dynamics. Among these invariants are the various fractal dimensions and—even more useful for physics— the local and global Lyapunov exponents. These latter quantities also establish the predictability of the nonlinear source. Chaotic systems are notorious for the unpredictability of their orbits. This, of course, is directly connected with the instabilities in phase space we discussed earlier. The chaotic motion of a physical system is not unpredictable even with all the instabilities one encounters in traversing state space. The limited predictability of the chaos is quantified by the local and global Lyapunov exponents, which can be determined from the measurements themselves. So in one sense, chaotic systems allow themselves to be classified and to have their intrinsic predictability established at the same time.

One of the hallmarks of chaotic behavior is the sensitivity of any orbit to small changes in initial condition or small perturbations anywhere along the orbit. Because of this sensitivity, which is quantified in the presence of positive Lyapunov exponents, it is inappropriate to compare two orbits of a nonlinear system directly with each other; generically, they will be totally uncorrelated. The invariants of the geometric figure called the attractor, which each orbit visits during its evolution, will be the same. These are independent of initial conditions and provide a direct analogy to the Fourier frequencies of a linear dynamical system. So these invariants can be compared if we wish to establish whether the observations of the two orbits mean they arise from the same physical system. System identification in nonlinear chaotic systems means establishing a set of invariants for each system of interest and then comparing observations to that library of invariants.

As noted above there are also topological invariants of any dynamical system which can be used to the same effect. These are not dynamical in the same fashion as the Lyapunov exponents, but can serve much the same function as regards classification. Metric invariants such as Lyapunov exponents or the various fractal dimensions are unlikely to provide a "complete" set of invariants, so using them and the topological invariants together may provide sufficient discrimination power to allow for the practical separation of observed physical systems.

D. Modeling—linear and nonlinear

With all this analysis of the source of the measurements now done, we may address the problem of building models of the physical processes acting at the source. The general plan for this was outlined in the Introduction, and here we focus only on the first kind of model making: working within the coordinate system established by the analysis of the signal to build up local or global models of how the system evolves in the reconstructed phase space. In linear systems the task is relatively easy. One has observations $s(n)$ and they must somehow be linearly related to observations at earlier times and to the driving forces at earlier times. This leads to models of the form

$$s(n) = \sum_{k=1}^{N} a_k s(n-k) + \sum_{l=1}^{L} b_l g(n-l) , \qquad (13)$$

where the coefficients $\{a_k\}$ and $\{b_l\}$ are to be determined by fits to the data, typically using a least-squares or an information-theoretic criterion, and the $g(n)$ are some deterministic or stochastic forcing terms, which are specified by the modeler. This linear relationship becomes simpler and algebraic when we take the "z transform" by defining

$$S(z) = \sum_{n=-\infty}^{+\infty} z^n s(n) , \qquad (14)$$

which leads to

$$S(z) = \frac{\sum_{l=1}^{L} b_l z^l}{1 - \sum_{k=1}^{K} a_k z^k} G(z) , \qquad (15)$$

with $G(z)$ the z transform of the forcing. This is just a discrete form of Fourier transform, of course.

Once the coefficients are established by one criterion or another the model equation (13) is then used for prediction or as part of a control scheme. The choice of coefficients should be consistent with any knowledge one has of spectral peaks in the system, and the denominator in the z transform holds all that information in terms of poles in the z plane. Much of the literature on linear signal processing can be regarded as efficient and thoughtful ways of choosing the coefficients in this kind of linear model in situations of varying complexity. From the point of view of dynamical systems this kind of model consists of simple linear dynamics—the terms involving the $\{a_k\}$—and simple linear averaging over the forcing. The first part of the modeling, which involves dynamics, is often called autoregressive (AR) and the second moving average (MA), and the combination labeled ARMA modeling.

This kind of model will always have zero or negative Lyapunov exponents, zero KS entropy, and will never be chaotic. It is, as a global model of interesting nonlinear physics, not going to do the job.

Nonlinear modeling of chaotic processes revolves around the idea of a compact geometric attractor on which our observations evolve. Since the orbit is continually folded back on itself by the dissipative forces and the nonlinear part of the dynamics, we expect to find in the neighborhood of any orbit point $\mathbf{y}(n)$ other orbit points $\mathbf{y}^{(r)}(n); r = 1, 2, \ldots, N_B$, which arrive in the neighborhood of $\mathbf{y}(n)$ at quite different times than n. One can then build various forms of interpolation function, which account for whole neighborhoods of phase space and how they evolve from near $\mathbf{y}(n)$ to the whole set of points near $\mathbf{y}(n+1)$. The use of phase-space information in the modeling of the temporal evolution of the physical process is the key innovation in modeling chaotic processes. The general method would work for regular processes, but is likely to be less successful because the neighborhoods are less populated.

The implementation of this idea is to build parametrized nonlinear functions $\mathbf{F}(\mathbf{x}, \mathbf{a})$ which take $\mathbf{y}(n)$ into $\mathbf{y}(n+1) = \mathbf{F}(\mathbf{y}(n), \mathbf{a})$ and then use various criteria to determine the parameters $\mathbf{a}$. Further, since one has the notion of local neighborhoods, one can build up one's model of the process neighborhood by neighborhood and, by piecing together these local models, produce a global nonlinear model that captures much of the structure in the attractor itself.

Indeed, in some sense the main departure from linear modeling is to realize that the dynamics evolves in a multivariate space whose size and structure is dictated by the data itself. No clue is given by the data as to the kind of model that would be appropriate for the source of chaotic data. Indeed, since quite simple polynomial models and quite complex models can both give rise to time-asymptotic orbits with strange attractors, the critical part of the modeling—the connection to the physics—remains in the hands of the physicist. It is likely there is no algorithmic solution (Rissanen, 1989) to how to choose models from data alone, and that, of course, is both good news and bad news.

E. Signal synthesis

The main emphasis of this review and actually of most work on nonlinear signals has been on their analysis; that is, given an observed signal, can we tell if it came from a low-dimensional chaotic attractor, can we characterize that attractor, can we predict the evolution of state-space points near the attractor or into the future, etc.

Another direction now being taken concerns the *synthesis* of new signals using chaotic sources. Can one utilize these signals for communication between distant locations or synchronization of activities at these locations, etc. This, in our opinion, is likely to be the main direction of engineering applications of the knowledge we have garnered about nonlinear systems operating in a chaotic mode. We shall touch upon these ideas as we proceed here, but their review will have to await a more

substantial set of developments, and this we are confident will be available over the next few years.

IV. RECONSTRUCTING PHASE SPACE OR STATE SPACE

We now turn to methods for reconstructing the state space of a system from information on scalar measurements $s(n) = s(t_0 + n\tau_s)$. If more than just a single scalar were measured, that would put us in an advantageous position, as will become clear, but for the moment we proceed with one type of measurement.

From the discussion of dynamical systems we see that if we were able to establish values for the time derivatives of the measured variable, we could imagine finding the connection among the various derivatives and the state variables, namely, the differential equation, which produced the observations. To create a value for

$$\dot{s}(n) = \frac{ds(t)}{dt}\bigg|_{t=t_0+n\tau_s}, \tag{16}$$

we need to approximate the derivative, since we have measurements only every τ_s. The natural approximation would be

$$\dot{s}(n) \approx \frac{s(t_0 + (n+1)\tau_s) - s(t_0 + n\tau_s)}{\tau_s}. \tag{17}$$

With finite τ_s this is just a crude high pass filter of the data, and we are producing a poor representation of the time derivative. When we next try to estimate $\ddot{s}(n)$ with second differences, we are producing an even poorer version of the second derivative, etc. If we examine the formula for the derivatives, we see that at each step we are adding to the information already contained in the measurement $s(n)$ measurements at other times lagged by multiples of the observation time step τ_s. In approximately 1980 both a group at the University of California, Santa Cruz (Packard *et al.*, 1980) and David Ruelle apparently simultaneously and independently introduced the idea of using time-delay coordinates to reconstruct the phase space of an observed dynamical system. The Santa Cruz group (Packard *et al.*, 1980) implemented the method in an attempt to reconstruct the time derivatives of $s(t)$. The main idea is that we really do not need the derivatives to form a coordinate system in which to capture the structure of orbits in phase space, but that we could directly use the lagged variables $s(n+T) = s(t_0 + (n+T)\tau_s)$, where T is some integer to be determined. Then using a collection of time lags to create a vector in d dimensions,

$$\mathbf{y}(n) = [s(n), s(n+T), s(n+2T), \dots, s(n+(d-1)T)], \tag{18}$$

we would have provided the required coordinates. In a nonlinear system, the $s(n+jT)$ are some (unknown) nonlinear combination of the actual physical variables that comprise the source of the measurements. While we do not know the relationship between the physical variables $\mathbf{u}(n)$ and the $s(n+jT)$, for purposes of creating a phase space, it does not matter, since any smooth nonlinear change of variables will act as a coordinate basis for the dynamics. This idea was put on a firmer footing by Mañé and Takens (Mañé, 1981; Takens, 1981) and further refined relatively recently (Sauer *et al.*, 1991).

To see this idea in action we take data from the Lorenz attractor described earlier. We take points from the variable $x(n)$, and in Fig. 7 we plot the points $[x(n), x(n+dT), x(n+2dT)]$ in R^3. Comparing this to Fig. 3, we see that from the complicated time trace in Fig. 1 we have created, by using 3-vectors of time-delayed variables, a geometric object similar in appearance, though somewhat distorted, to the original geometric portrait of the strange attractor. The distortion is not surprising since the coordinates $[x(n), x(n+dT), x(n+2dT)]$ are different from $[x(n), y(n), z(n)]$.

Now how shall we choose the time lag T and the dimension of the space d to go into the timelagged vectors $\mathbf{y}(n)$, Eq. (18)? The arguments of Mañé and Takens suggest that it does not matter what time lag one chooses, and a sufficient condition for d depends on the dimension of the attractor d_A. They argue that if d, which is an integer, is larger than $2d_A$, which can be fractional, then the attractor as seen in the space with lagged coordinates will be smoothly related to the attractor as viewed in the original physical coordinates, which we do not know. In practice, their theorem tells us that, if we have chosen d large enough, physical properties of the attractor that we wish to extract from the measurements will be the same when computed on the representative in lagged coordinates and when computed in the physical coordinates. The procedure of choosing sufficiently large d is formally known as embedding, and any dimension that works is called an embedding dimension d_E. Once one has achieved a large enough $d = d_E$, then any $d \geq d_E$ will also

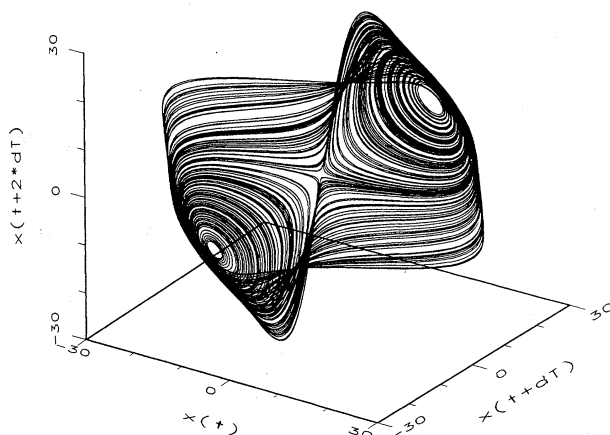

FIG. 7. Lorenz time series $x(t)$ (Fig. 1) embedded in a three-dimensional phase space $(x(t), x(t+dT), x(t+2dT)), dT = 0.2$.

provide an embedding.

Time-delay embedding is certainly not the only method for embedding scalar data in a multivariate dynamical environment. Other methods (Landa and Rosenblum, 1989; Mindlin *et al.*, 1991) are available and have their advantages. Time-delay embedding is, perhaps, the only systematic method for going from scalar data to multidimensional phase space, and it has certainly been the most explored in the literature.

A. Choosing time delays

The statement of the embedding theorem that any time lag will be acceptable is not terribly useful for extracting physics from the data. If we choose T too small, then the coordinates $s(n+jT)$ and $s(n+(j+1)T)$ will be so close to each other in numerical value that we cannot distinguish them from each other. From any practical point of view they have not provided us with two independent coordinates. Similarly, if T is too large, then $s(n+jT)$ and $s(n+(j+1)T)$ are completely independent of each other in a statistical sense, and the projection of an orbit on the attractor is onto two totally unrelated directions. The origin of this statistical independence is the ubiquitous instability in chaotic systems, which results in any small numerical or measurement error's being amplified exponentially in time. A criterion for an intermediate choice is called for, and it cannot come from the embedding theorem itself or considerations based on it, since the theorem works for almost any value of T.

One's first thought might be to consider the values of $s(n)$ as chosen from some unknown distribution. Then computing the *linear autocorrelation function*

$$C_L(\tau)=\frac{\frac{1}{N}\sum_{m=1}^{N}[s(m+\tau)-\bar{s}][s(m)-\bar{s}]}{\frac{1}{N}\sum_{m=1}^{N}[s(m)-\bar{s}]^2} , \quad (19)$$

where

$$\bar{s}=\frac{1}{N}\sum_{m=1}^{N}s(m) ,$$

and looking for that time lag where $C_L(\tau)$ first passes through zero, would give us a good hint of a choice for T. Indeed, this does give a good hint. It tells us, however, about the independence of the coordinates only in a linear fashion. To see this, recall that if we want to know whether two measurements $s(n)$ and $s(n+\tau)$ depend linearly on each other on the average over the observations, we find that their connection, in a least-squares sense, is through the correlation matrix just given.

That is, if we assume that the values of $s(n)$ and $s(n+\tau)$ are connected by

$$s(n+\tau)-\bar{s}=C_L(\tau)[s(n)-\bar{s}] , \quad (20)$$

then minimizing

$$\sum_{n=1}^{N}\{s(n+\tau)-\bar{s}-C_L(\tau)[s(n)-\bar{s}]\}^2 , \quad (21)$$

with respect to $C_L(\tau)$, immediately leads to the definition of $C_L(\tau)$ above.

Choosing T to be the first zero of $C_L(\tau)$ would then, on average over the observations, make $s(n)$ and $s(n+T)$ linearly independent. What this may have to do with their nonlinear dependence or their utility as coordinates for a nonlinear system is not addressed by all this. Since we are looking for a *prescription* for choosing T, and this prescription must come from considerations beyond those in the embedding theorem, linear independence of coordinates may serve, but we prefer another point of view, one that stresses an important aspect of chaotic behavior—namely the viewpoint of information theory (Shaw, 1984; Fraser and Swinney, 1986; Fraser, 1989a, 1989b)—and leads to a nonlinear notion of independence.

B. Average mutual information

In the Introduction we discussed how the idea of information created by chaotic behavior is suggested in a qualitative fashion by the KS entropy and the growth in time of the number of states of a system available to a small piece of phase space as $2^{[th(X)]}$. The same idea can be used to identify how much information one can learn about a measurement at one time from a measurement taken at another time. In a general sort of context, let us imagine that we have two systems, call them A and B, with possible outcomes in making measurements on them a_i and b_k. We consider there to be a probability distribution associated with each system governing the possible outcomes of observations on them. The amount one learns in bits about a measurement of a_i from a measurement of b_k is given by the arguments of information theory (Gallager, 1968) as

$$I_{AB}(a_i,b_k)=\log_2\left[\frac{P_{AB}(a_i,b_k)}{P_A(a_i)P_B(b_k)}\right] , \quad (22)$$

where the probability of observing a out of the set of all A is $P_A(a)$, and the probability of finding b in a measurement of B is $P_B(b)$, and the joint probability of the measurement of a and b is $P_{AB}(a,b)$. This quantity is called the *mutual information* of the two measurements a_i and b_k and is symmetric in how much one learns about b_k from measuring a_i. The *average mutual information* between measurements of any value a_i from system A and b_k from B is the average over all possible measurements of $I_{AB}(a_i,b_k)$,

$$I_{AB}(T)=\sum_{a_i,b_k}P_{AB}(a_i,b_k)I_{AB}(a_i,b_k) . \quad (23)$$

To place this abstract definition into the context of observations from a physical system $s(n)$, we think of the sets of measurements $s(n)$ as the A set and of the mea-

surements a time lag T later, $s(n+T)$, as the B set. The average mutual information between observations at n and $n+T$, namely, the average amount of information about $s(n+T)$ we have when we make an observation of $s(n)$, is then

$$I(T) = \sum_{n=1}^{N} P(s(n), s(n+T)) \log_2 \left[\frac{P(s(n), s(n+t))}{P(s(n))P(s(n+T))} \right] ,$$

(24)

and $I(T) \geq 0$ (Gallager, 1968).

The average mutual information is really a kind of generalization to the nonlinear world from the correlation function in the linear world. It is the average over the data or equivalently the attractor of a special statistic, namely, the mutual information, while the correlation function is the average over a quadratic polynomial statistic. When the measurements of systems A and B are completely independent, $P_{AB}(a,b) = P_a(a)P_b(b)$, and $I_{AB}(a_i, b_k) = 0$.

Evaluating $I(T)$ from data is quite straightforward. To find $P(s(n))$ we need only take the time trace $s(n)$ and project it back onto the s axis. The histogram formed by the frequency with which any given value of s appears gives us $P(s(n))$. If the time series is long and stationary, then $P(s(n+T))$ is the same as $P(s(n))$. The joint distribution comes from counting the number of times a box in the $s(n)$ versus $s(n+T)$ plane is occupied and then normalizing this distribution. All this can be done quite easily and rather fast on modern PC's or workstations.

In his dissertation (Fraser, 1988) Fraser gives a very clever recursive algorithm, which is basically a search algorithm for neighbors in d-dimensional space, for evaluating $I(T)$ and its generalizations to higher dimensions. The more appropriate question to answer when creating d-dimensional vectors $\mathbf{y}(n)$ out of time lags is this: what is the joint mutual information among a set of measurements $s(n), s(n+T), \ldots, s(n+(d-1)T)$ as a function of T? For our purposes of giving a rule for an estimation of a useful time lag T, it will suffice to know $I(T)$ as we have defined it.

Now we have to decide what property of $I(T)$ we should select, in order to establish which among the various values of T we should use in making our data vectors $\mathbf{y}(n)$. If T is too small, the measurements $s(n)$ and $s(n+T)$ tell us so much about one another that we need not make both measurements. If T is large, then $I(T)$ will approach zero, and nothing connects $s(n)$ and $s(n+T)$, so this is not useful. Fraser and Swinney (1986) suggest, *as a prescription*, that we choose that T_m where the first minimum of $I(T)$ occurs as a useful selection of time lag T. As an example of this we show in Figs. 8 and 9 the average mutual information for the Lorenz attractor and for the data from the Carroll-Pecora nonlinear circuit. Each has a first minimum at some T_m, and this minimum can be used as a time lag for phase-space reconstruction. The lag T_m is selected as a time lag
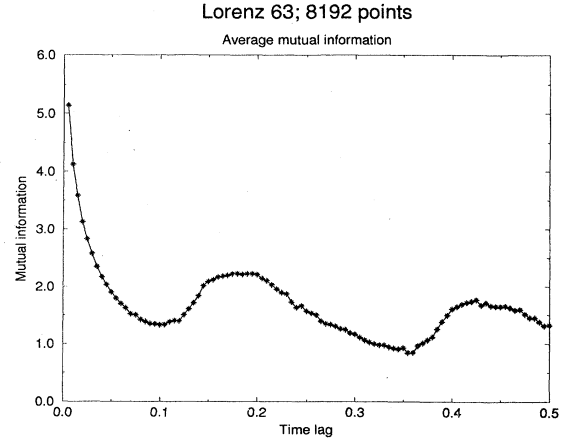
Lorenz 63; 8192 points

FIG. 8. Average mutual information $I(T)$ as a function of time lag T for Lorenz data.

where the measurements are somewhat independent, but not statistically independent:

$$P(s(n), s(n+T_m)) \neq P(s(n))P(s(n+T_m)) .$$

Recognizing that this is a prescription, one may well ask what to suggest if the average mutual information has no minimum. This occurs when one is dealing with maps, as the $I(T)$ curve from $x(n)$ data taken from the Hénon map (Hénon, 1976)

$$x(n+1) = 1 + y(n) - 1.4x(n)^2 ,$$

$$y(n+1) = 0.3x(n)$$

(25)

shows in Fig. 10. Similarly, if one integrates the Lorenz equations with a time step dt in the integrator, which is then increased, the minimum seen in Fig. 8 fills in as dt grows and eventually disappears. This does not mean that $I(T)$ loses its role as a good grounds for selection of T_m, but only that the first minimum criterion needs to be replaced by something representing good sense. Without much grounds beyond intuition, we use $T_m = 1$ or 2 if we
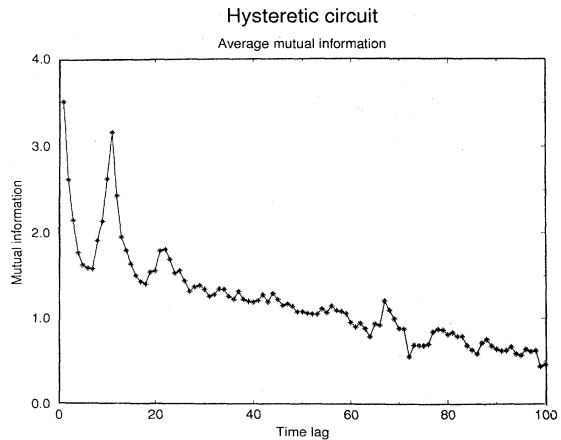
Hysteretic circuit

FIG. 9. Average mutual information for hysteretic circuit data.

Henon map, 8192 points
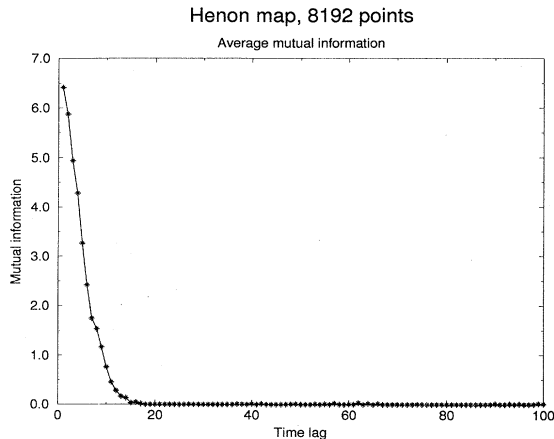
Average mutual information

FIG. 10. Average mutual information for data generated by the Hénon attractor [Eq. (25)].

know the data comes from a map, or choose T_m such that $I(T_m)/I(0) \approx \frac{1}{5}$. This is clearly an attempt to choose a useful T_m in which some nonlinear decorrelation is at work, but not too much.

Since this is prescriptive, one may compare it to the prescription used in linear dynamics of choosing a time lag τ_m such that $C_L(\tau_m) = 0$ for the first time. Why not the second zero or the fifth? Grassberger et al. (1991) scrutinize the use of a simple prescription for choosing T. The authors point out that for high-dimensional data one must use a more sensitive criterion than the minimum average mutual information between $x(n)$ and $x(n+T)$, especially if one allows differing lags in each component of the data vector $\mathbf{y}(n)$. Recognizing, as we have stressed, that the choice of T is prescriptive, we agree with their caution that "we do not believe that there exists a unique optimal choice of delay." Nonetheless, it is useful to have a general rule of thumb as a guide to a delay T that is workable; seeking the optimum is likely to be quite unrewarding.

Liebert and Schuster (1989) have examined the first minimum criterion by adding another criterion involving the so-called correlation integral, which we shall discuss below, and concluded that the T_m of the first minimum is that T_m in which the correlation integral is also numerically well established. This is yet another prescription, and we have no special bent toward any particular prescription, though when the $I(T)$ curve does have minima, we find that to be a practical and useful choice. The main idea is to use criteria based on correlation properties tuned to aspects of chaotic systems, namely, properties that are attuned to chaotic dynamics as information generators, and average mutual information certainly does just that.

C. Choosing the embedding dimension

The embedding theorem of Mañé and Takens rests on topological results discussed sixty years ago (Whitney,

1936). It is really quite independent of dynamics and is a generalization of ideas on regular geometries. The goal of the reconstruction theorem is to provide a Euclidean space R^d large enough so that the set of points of dimension d_A can be unfolded without ambiguity. This means that if two points of the set lie close to each other in some dimension d they should do so because that is a property of the set of points, not of the small value of d in which the set is being viewed. For example, if the attractor is one dimensional and it is viewed in $d = 2$, it may still be that the one-dimensional line crosses itself at isolated points. At these points there is an ambiguity about which points are neighbors of which other points. This ambiguity is completely resolved by viewing the same one-dimensional attractor in $d = 3$, where all possible zero-dimensional or point crossings of the set on itself are resolved. Of course, if one wants to view the set of points in $d = 4$, that is fine too. No further ambiguities are introduced by adding further coordinates to the viewing or embedding space. When all ambiguities are resolved, one says that the space R^d provides an embedding of the attractor.

An equivalent way to look at the embedding theorem is to think of the attractor as comprised of orbits from a system of very high dimension, even infinite dimensional as in the case of delay differential equations or partial differential equations. The attractor, which has finite d_A, lies in a very small part of the whole phase space, and we can hope to provide a projection of the whole space down to a subspace in which the attractor can be faithfully captured. Most of the large space, after all, is quite empty. If this projection takes us down to too small a space, as for example the scalar one-dimensional measurements themselves are almost sure to do, then the evolution as seen in this too small space will be ambiguous because of the presence of incorrect neighboring points. In this too small space the appearance of the orbit will be unduly complicated, and part of the bad name of chaos or irregular behavior comes from looking at the orbits in the wrong space. The appearance of the attractor when fully unfolded in the larger embedding space is much more regular and attractive than when projected down to one dimension as in the scalar measurements. The embedding theorem provides a *sufficient* condition from geometrical considerations alone for choosing a dimension d_E large enough so that the projection is good—i.e., without orbit crossings of dimension zero, one, two, etc.

The *sufficient* integer dimension d_E is not always necessary. For the Lorenz attractor, for example, we have $d_A = 2.06$ (we discuss later how to establish this). This sufficient condition would suggest, since $d_E > 2d_A$, that if we choose to view the Lorenz attractor in $d_E = 5$ we can do so without ambiguity. This is not at all inconsistent with our knowledge of the Lorenz differential equations, which provide three coordinates for an R^3 in which to view the attractor, since we are speaking now about a time-delayed version of one of the Lorenz variables, and these are nonlinear combinations of the origi-

nal $x(t), y(t), z(t)$ which my twist the attractor onto itself in only three dimensions. The embedding theorem guarantees us that, however complicated these twists may be, if the nonlinear transformation is smooth, $d_E = 5$ will untwist them globally in phase space. The theorem is silent on local phase-space properties.

Why should we care if we always work in a dimension d_E that is sufficient to unfold the attractor but may be larger than necessary? Would it not be just fine to do all analysis of the properties of the Lorenz system in $d_E = 5$? From the embedding theorem point of view, the answer is always yes. For the physicist more is needed. Two problems arise with working in dimensions larger than really required by the data and time-delay embedding: (i) many of the computations we shall discuss below for extracting interesting properties from the data, and that is after all the goal of all this, require searches and other operations in R^d whose computational cost rises exponentially with d; and (ii) in the presence of "noise" or other high-dimensional contamination of our observations, the "extra" dimensions are not populated by dynamics, already captured by a smaller dimension, but entirely by the contaminating signal. In too large an embedding space one is unnecessarily spending time working around aspects of a bad representation of the observations which are solely filled with "noise."

This realization has motivated the search for analysis tools that will identify a *necessary* embedding dimension from the data itself. There are four methods we shall discuss: (i) *singular-value decomposition of the sample covariance matrix;* (ii) *"saturation" with dimension of some system invariant;* (iii) *the method of false nearest neighbors,* (iv) *the method of true vector fields.*

1. Singular-value analysis

If our measurements $y(n)$ are composed of the signal from the dynamical system we wish to study plus some contamination from other systems, then in the absence of specific information about the contamination it is plausible to assume it to be rather high dimensional and to assume that it will fill more or less uniformly any few-dimensional space we choose for our considerations. Let us call the embedding dimension necessary to unfold the dynamics we seek d_N. If we work in $d_E > d_N$, then in an heuristic sense $d_E - d_N$ dimensions of the space are being populated by contamination alone. If we think of the observations embedded in d_E as composed of a true signal $y_T(n)$ plus some contamination c: $y(n) = y_T(n) + c(n)$, then the $d_E \times d_E$ sample covariance matrix

$$\mathbf{COV} = \frac{1}{N} \sum_{n=1}^{N} [\mathbf{y}(n) - \mathbf{y}_{av}][\mathbf{y}(n) - \mathbf{y}_{av}]^T , \quad (26)$$

with

$$\mathbf{y}_{av} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}(n) , \quad (27)$$

will, again in an heuristic sense, have d_N eigenvalues arising from the variation of the (slightly contaminated) real signal about its mean and $d_E - d_N$ eigenvalues which represent the "noise." If the contamination is quite high dimensional, it seems plausible to think of it as filling these extra $d_E - d_N$ dimensions in some uniform manner, so perhaps one could expect the unwelcome $d_E - d_N$ eigenvalues, representing the power in the extra dimensions, to be nearly equal. If this were the case, then by looking at the eigenvalues, or equivalently the singular values of **COV**, we might hope to find a "noise floor" at which the eigenvalue spectrum turned over and became flat (Broomhead and King, 1986). There are d_E eigenvalues, and the one where the floor is reached may be taken as d_N.

This analysis can also be carried out *locally* (Broomhead and Jones, 1989), which means that the covariance matrix is over a neighborhood of the N_B nearest neighbors, $\mathbf{y}^{(r)}(n)$ of any given data point $\mathbf{y}(n)$:

$$\mathbf{COV}(n) = \frac{1}{N_B} \sum_{r=1}^{N_B} [\mathbf{y}^{(r)}(n) - \mathbf{y}_{av}(n)][\mathbf{y}^{(r)}(n) - \mathbf{y}_{av}(n)]^T$$

$$\mathbf{y}_{av}(n) = \frac{1}{N_B} \sum_{r=1}^{N_B} \mathbf{y}^{(r)}(n) . \quad (28)$$

The global singular-value analysis has the attractive feature of being easy to implement, but it has the downside of being hard to interpret on occasion. It gives a *linear* hint as to the number of active degrees of freedom, but it can be misleading because it does not distinguish two processes with nearly the same Fourier spectrum (Fraser, 1989a, 1989b) or because of differing computers the anticipated "noise floor" is reached at different numerical levels. Using it as a guide to the physicist, to be looked at along with other tools, can be quite helpful. Local covariance analysis can provide quite useful insights as to the structure of an attractor and, as emphasized by Broomhead and Jones (1989), can be used to distinguish degrees of freedom in a quantitative fashion. The local analysis is also useful in certain signal-separation algorithms when one of the signals to be separated has substantial resemblance to white noise (Sauer, 1991; Cawley and Hsu, 1992).

2. Saturation of system invariants

If the attractor is properly unfolded by choosing a large enough d_E, then any property associated with the attractor which depends on distances between points in the phase space should become independent of the value of the embedding dimension once the necessary d_E has been reached. Increasing d_E beyond this d_N should not affect the value of these properties, and, *in principle,* the appropriate necessary embedding dimension d_N can be established by computing such a property for $d_E = 1, 2, \ldots$ until variation with d_E ceases.

A familiar example of this comes from the average

over the attractor of moments of the number density $n(r,\mathbf{x})$: the number of points on the attractor or equivalently on the orbit within a radius of r of points $\mathbf{x}$ in the phase space, $n(r,\mathbf{x})$, is defined by

$$n(r,\mathbf{x}) = \frac{1}{N}\sum_{k=1}^{N}\theta(r - |\mathbf{y}(k) - \mathbf{x}|) , \qquad (29)$$

with $\theta(u) = 0$, $u < 0$; $\theta(u) = 1$, $u > 0$, and the average over all points of powers of $n(r,\mathbf{x})$ is (Grassberger and Procaccia, 1983a, 1983b)

$$C_q(r) = \frac{1}{M}\sum_{j=1}^{M}[n(r,\mathbf{y}(j))]^{(q-1)} . \qquad (30)$$

We shall show in Sec. V that averages such as these are independent of initial conditions and are thus good candidates for characterizing an attractor. We shall also elaborate on the discussion of the correlation integrals $C_q(r)$ in Sec. V.D. Now we want to concentrate on the fact that this quantity depends on the embedding dimension d_E used to make the vectors $\mathbf{y}(k)$ $= [s(n), s(n+T), \ldots, s(n+(d_E-1)T)]$.

We can evaluate $C_q(r)$ as a function of d_E and determine when the slope of its logarithm as a function of $\log(r)$ becomes independent of d_E. In Fig. 11 we show a log-log plot of $C_2(r)$ for data from the Lorenz attractor as a function of d_E. It is clear that for $d_E = 3$, or perhaps $d_E = 4$, the slope of the function $C_2(r)$ becomes independent of embedding dimension, and thus we can safely choose $d_E = 3$ or $d_E = 4$, which is less than the sufficient condition from the embedding theorem of $d_E = 5$.

While what we have described is quite popular, it does not give any special distinction, from the point of view of the embedding theorem, to $C_q(r)$ as the function whose independence of d_E we should establish. As we shall see in the next section, there is a very large class of functions one could consider, all of which are equivalent from the point of view of embedding.

Indeed, it is our own opinion that determining d_N by looking for the independence of some function on the attractor, whether it be $C_q(r)$ or another choice, is somewhat of an indirect answer to the direct geometric question posed by the embedding theorem: when have we unfolded the attractor by our choice of time-delay coordinates?

FIG. 11. Sequence of correlation integrals $C_2(r)$ for data from the Ikeda map in embedding dimensions $d_E = 1, \ldots, 8$.

3. False nearest neighbors

The third method for determining d_N comes from asking, directly of the data, the basic question addressed in the embedding theorem: when has one eliminated false crossings of the orbit with itself which arose by virtue of having projected the attractor into a too low dimensional space?

Answers to this question have been discussed in various ways. Each of the ways has addressed the problem of determining when points in dimension d are neighbors of one another by virtue of the projection into too low a dimension. By examining this question in dimension one, then dimension two, etc. until there are no incorrect or false neighbors remaining, one should be able to establish, from geometrical considerations alone, a value for the necessary embedding dimension $d_E = d_N$. We describe the implementation of Kennel *et al.* (1992).

In dimension d each vector

$$\mathbf{y}(k) = [s(k), s(k+T), \ldots, s(k+(d-1)T)] \qquad (31)$$

has a nearest neighbor $\mathbf{y}^{NN}(k)$ with nearness in the sense of some distance function. Euclidean distance is natural and works well. The Euclidean distance in dimension d between $\mathbf{y}(k)$ and $\mathbf{y}^{NN}(k)$ we call $R_d(k)$:

$$R_d(k)^2 = [s(k) - s^{NN}(k)]^2 + [s(k+T) - s^{NN}(k+T)]^2 + \cdots + [s(k+T(d-1)) - s^{NN}(k+T(d-1))]^2 . \qquad (32)$$

$R_d(k)$ is presumably small when one has a lot of data, and for a data set with N entries, this distance is more or less of order $1/N^{1/d}$. In dimension $d+1$ this nearest-neighbor distance is changed due to the $(d+1)^{st}$ coordinates $s(k+dT)$ and $s^{NN}(k+dT)$ to

$$R_{d+1}^2(k) = R_d^2(k) + [s(k+dT) - s^{NN}(k+dT)]^2 . \qquad (33)$$

If $R_{d+1}(k)$ is large, we can presume it is because the near neighborliness of the two points being compared is due to
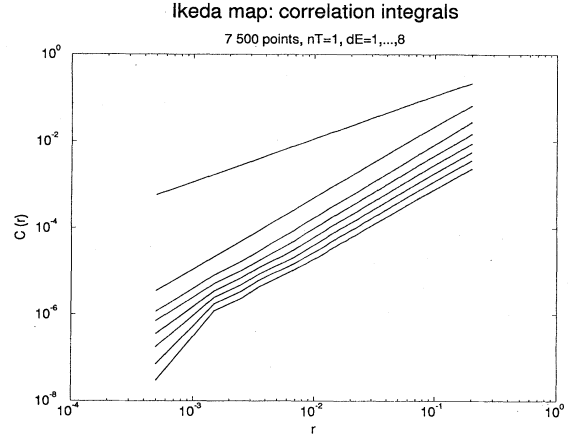
the projection from some higher-dimensional attractor down to dimension d. By going from dimension d to dimension $d+1$, we have "unprojected" these two points away from each other. Some threshold size R_T is required to decide when neighbors are false. Then if

$$\frac{|s(k+Td) - s^{NN}(k+Td)|}{R_d(k)} > R_T , \qquad (34)$$

the nearest neighbors at time point k (or $t_0 + k\tau_s$) are de-

clared false. In practice, for values of R_T in the range $10 \leq R_T \leq 50$ the number of false neighbors identified by this criterion is constant. With such a broad range of independence of R_T one has confidence that this is a workable criterion.

When this criterion is applied to various standard models, as shown in Figs. 12–14, one finds results that are quite sensible. For the Lorenz attractor, one finds that the number of false nearest neighbors drops to zero at $d_N = 3$, while the sufficient condition from the embedding theorem is at $d_E = 5$. For the Hénon map, we find $d_N = 2$, while the embedding theorem only guarantees us success in unfolding the attractor at $d_E = 3$, since $d_A = 1.26$ for this map.

For the Ikeda map (Ikeda, 1979; Hammel *et al.*, 1985) of the complex plane $z(k) = x(k) + iy(k)$ to itself,

$$z(k+1) = p + Bz(k)\exp\left[i\kappa - \frac{i\alpha}{(1+|z(k)|^2)}\right], \qquad (35)$$

which describes the evolution of a laser in a ring cavity with a lossy active medium, we find, with "standard" parameters $p = 1.0$, $B = 0.9$, $\kappa = 0.4$, and $\alpha = 6.0$, where $d_A \approx 1.8$, that we require $d_N = 4$ to eliminate all false neighbors. This is the same dimension that the embedding theorem tells us is sufficient. In a sense this example tells us that the false-nearest-neighbor test is really quite powerful. While one needs only a phase space of dimension two for the original map, the test reveals that the choice of time delays as coordinates for a reconstructed state space twists and folds the attractor so much that in these coordinates these features of the coordinate system require a larger dimension to be undone. Comparing the actual $[x(k), y(k)]$ phase plane with the $[x(k), x(k+1)]$ phase portrait as in Figs. 15 and 16 makes this twisting quite apparent.

The criterion stated so far for false nearest neighbors has a subtle defect. If one applies it to data from a very-high-dimensional random-number generator, such as is
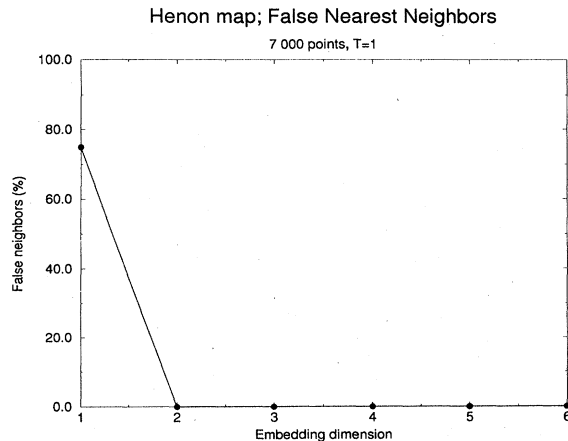

Henon map; False Nearest Neighbors

FIG. 13. Percentage of false nearest neighbors as a function of embedding dimension for Hénon map data.

found on any modern computer, it indicates that this set of observations can be embedded in a small dimension. If one increases the number of points analyzed, the apparent embedding dimension rises. The problem is that when one tries to populate uniformly (as "noise" will try to do) an object in d dimensions with a fixed number of points, the points must move further and further apart as d increases because most of the volume of the object is at large distances. If we had an infinite quantity of data, there would be no problem, but with finite quantities of data eventually all points have "near neighbors" that do not move apart very much as dimension is increased. As it happens, the fact that points are nearest neighbors does not mean they are close on a distance scale set by the approximate size R_A of the attractor. If the nearest neighbor to $y(k)$ is not close, so $R_d(k) \approx R_A$, then the distance $R_{d+1}(k)$ will be about $2R_d(k)$. This means that distant, but nearest, neighbors will be stretched to the extremities of the attractor when they are unfolded from each other, if they are false nearest neighbors.

Lorenz 63; False Nearest Neighbors

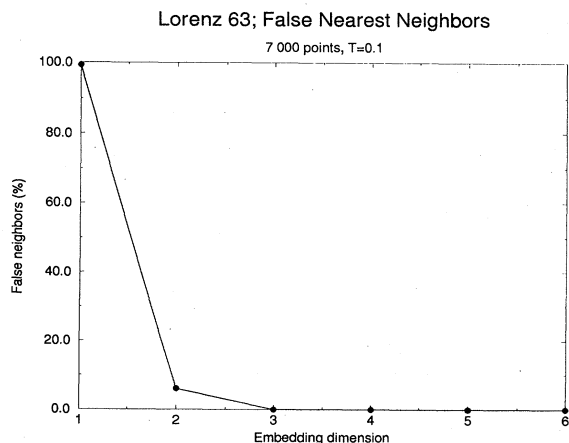FIG. 12. Percentage of false nearest neighbors as a function of embedding dimension for clean Lorenz data.

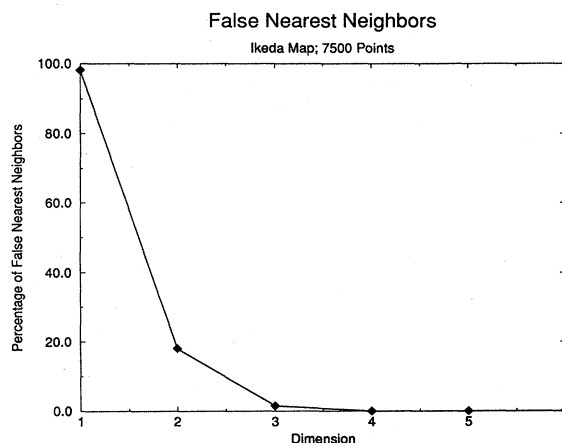False Nearest Neighbors

FIG. 14. Percentage of false nearest neighbors as a function of embedding dimension for data from the Ikeda map.
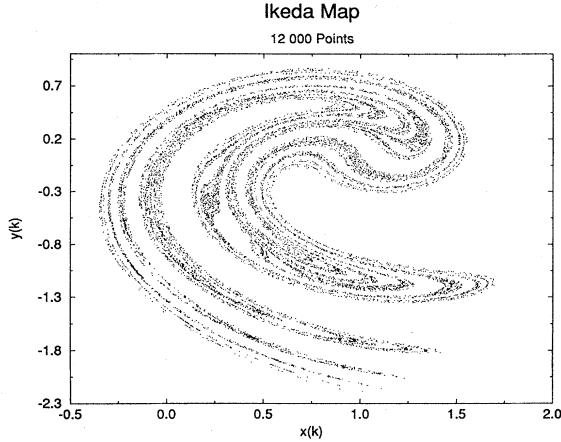
Ikeda Map

12 000 Points

FIG. 15. Ikeda attractor in two-dimensional phase space $(x(k), y(k))$.

This suggests as a second criterion for falseness of nearest neighbors that if

$$\frac{R_{d+1}(k)}{R_A} \geq 2 , \tag{36}$$

then $y(k)$ and its nearest neighbor are false nearest neighbors. As a measure of R_A one may use the rms value of the observations

$$R_A^2 = \frac{1}{N} \sum_{k=1}^{N} [s(k) - \bar{s}]^2$$

$$\bar{s} = \frac{1}{N} \sum_{k=1}^{N} s(k) . \tag{37}$$

The choice of other criteria for the size of the attractor works as well. When nearest neighbors failing either of these two criteria are designated as false, uniform noise from a random-number generator is now identified as high dimensional and scalar data from known low-

Ikeda Map

12 000 Points

FIG. 16. Time series from the Ikeda map embedded in a two-dimensional space $(x(k), x(k+1))$.

dimensional chaotic systems is identified as low dimensional.

This is a good occasion to emphasize a point that we have more or less made implicitly during the discussion. Noise, as commonly viewed, is thought of as completely unpredictable and "random." When a practical characterization of noise is sought, one states that it has a broadband Fourier spectrum and an autocorrelation function that falls rapidly to zero with time. Since the power spectrum and autocorrelation function are just Fourier transforms of one another, these are more or less the same statement. Just these properties are possessed by low-dimensional chaos as well, and chaos cannot be distinguished from noise on the basis of Fourier criteria. Instead, we see in the false-nearest-neighbor method or the other methods indicated for determining embedding dimension a true distinguishing characteristic: noise appears as high-dimensional chaos. Qualitatively they are the same. Quantitatively and practically one may not be able to work with dimensions higher than ten or twenty or whatever. It may be useful, then, in the sense of being able to work with problems in physics, to adopt any of the usual statistical techniques developed for random processes for signals with dimension higher than this threshold.

When one uses the false-nearest-neighbor test on measurements from a dynamical system that have been corrupted with "noise" or with signals from another dynamical system (low- or high-dimensional), there is a very useful robustness in the results. Figure 17 shows the effect of adding uniform random numbers lying in the interval $[-L, L]$ to a $x(n)$ signal from the Lorenz attractor. For the Lorenz attractor $R_A \approx 12$, and the contamination level is given in units of L/R_A in the figure. Until $L/R_A \approx 0.5$ we see a definite indication of low-dimensional signals. When the contamination level is low, the residual percentage of false neighbors gives an indication of the noise level. In any case, when the per-
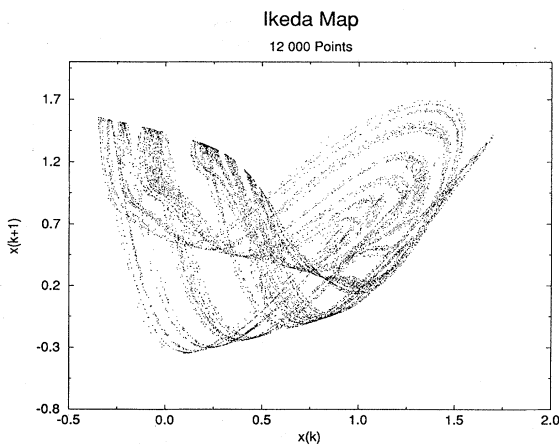
Lorenz63; False Nearest Neighbors

Noise 0.0, 0.005, 0.01,0.05,0.1,0.5,1.0; RMS L/R

FIG. 17. Percentage of false nearest neighbors as a function of embedding dimension for noisy Lorenz data: $\bigcirc$, $\sigma=0.0$; $\square$, $\sigma=0.005$; $\Diamond$, $\sigma=0.01$; $\triangle$, $\sigma=0.05$; $*$, $\sigma=0.1$; $\blacklozenge$, $\sigma=0.5$; $\bullet$, $\sigma=1.0$.

Hysteretic circuit; False Nearest Neighbors

7 000 points, T=6

FIG. 18. Percentage of false nearest neighbors as a function of embedding dimension for hysteretic circuit data.

centage of false nearest neighbors has dropped below $\approx 1\%$, one may choose that dimension as d_N and have some real sense of the error one might make in any subsequent calculations.
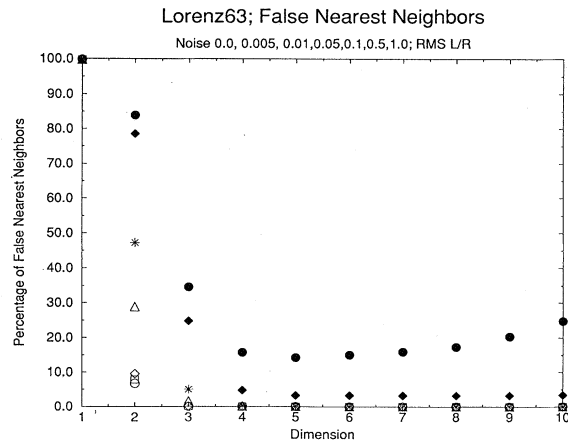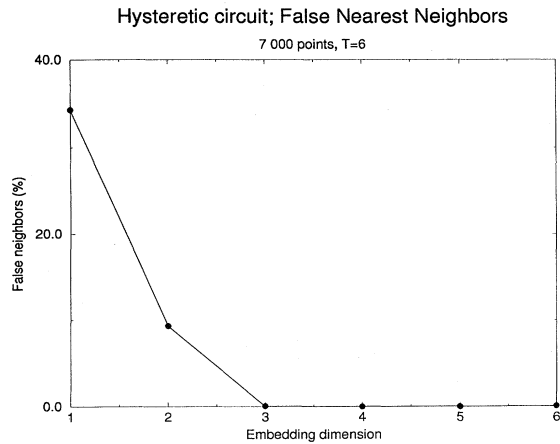
As an illustration of the method applied to real data, we present in Fig. 18 the percentage of false nearest neighbors for data from the hysteretic circuit of Carroll and Pecora. Clearly, choosing $d_N = 3$ would result in errors of less than 1% in computing distances or in prediction schemes. This is also commensurate with the error level in the observations.

4. True vector fields

Another geometrical approach to determining the embedding dimension for observed time series data is given by Kaplan and Glass (1992), who examine the uniqueness of the vector field responsible for the dynamics. If the dynamics is given by the autonomous rule $x \rightarrow F(x)$, and $F(x)$ is smooth, that is, differentiable, then the tangents to the evolution of the system are smoothly and uniquely given throughout the state space. Kaplan and Glass establish the local vector field by carving up phase space into small volumes and identifying where orbits enter and exit the volumes. This defines the local flow under the assumption that the volumes are small enough. "Small enough" is something one determines in practice, but there seems no barrier to making the volumes adequately small.

If one is in too small an embedding dimension, the vector fields in any location will often not be unique, as they will overlap one another because of the projection of the dynamics from a larger embedding space. As one increases the embedding dimension the frequency of overlap will fall to zero and the vector field will have been unfolded. If the dynamics is very high dimensional, then the vector field will not be unfolded until this high dimension has been reached. If this dimension is above

ten, say, then at the present time we can think of dealing with the problem using statistical methods.

To test for the deviation of the vector field from what would be expected for a random motion, taken to be random walks, the statistic

$$\frac{(V_j)^2 - (R^d_{n_j})^2}{1 - (R^d_{n_j})^2} \qquad (38)$$

is considered. The vectors V_j are the average directions of the line segments from entry to exit of the trajectory as it passes through box number j. $R^d_{n_j}$ is the known average displacement per step for a random walk in dimension d after n_j steps. This quantity will be averaged over all boxes.

If the data set is drawn from a random walk, the statistic will be 0, and if it is effectively random by being a high-dimensional deterministic system viewed in too low a dimension, it will be small. For a low-dimensional system viewed in a large enough dimension this will be near unity. The test appears to work quite well even with contaminated data. Indeed, the spirit of this method is quite similar to that of false nearest neighbors. Both are geometric and attempt to unfold a property of attractors from its overlap on the observation axis to its unambiguous state in an appropriately high-dimensional state space.

D. T and d_E

The determination of the appropriate phase space in which to analyze chaotic signals is one of the first tasks, and certainly a primary task, for all who wish to work with observed data in the absence of detailed knowledge of the system dynamics. We have spent some time on the methods that have been devised for this, both because it has been a subject of such intense interest, and because it is important for the physicist to realize that not only are there a variety of reliable techniques available for this job but also the set of tools may be useful in combination and as a complement to each other when the setting for the data changes.

To determine the time lag to be used in an embedding, one may always wish to use something nonlinear, such as average mutual information, but the data may mitigate against that. If one has sampled a map, achieved stroboscopically or taken as a Poincaré section, there is typically no minimum in the average mutual information function. The reason is quite simple: the time between samples τ_s is so long that the orbit has become decorrelated, in an information-theoretic sense. One can see this quite clearly by taking a flow such as the Lorenz model and increasing the time step in the differential equation solver step by step. At first, for a small time step, the average mutual information has a minimum. As the time step increases, the minimum moves in towards zero and, since time steps are only in integer steps, disappears.

What is one to do at this stage?

Typically, when one is handed data, there is nothing to do about resampling it, but if that is an option, take it. If not, one can turn to the autocorrelation function of the time series to find at least an estimate of what one can reliably use for a time delay in state-space reconstruction. While the criterion is linear, it may not be totally misleading to use the first zero crossing of the autocorrelation function as a useful time lag. When the average mutual information does have a first minimum, it is usually more or less the same order, in units of τ_s, as the first zero crossing of the autocorrelation, so one is not likely to be terribly misled by this tactic.

Once a time delay has been agreed upon, the embedding dimension is the next order of business. As we have seen, there are numerous options available, and we shall not repeat them in detail. Our own experience is that it is better to work with algorithms that are geometric rather than derivative from the data. Computing correlation functions $C_q(r)$ not only requires a large data set, it also degrades rapidly when the data are contaminated. If one wishes simply to know that the embedding dimension is low, then geometric methods will suffice. If one wishes to know whether to use dimension d or $d+1$, then geometric methods will allow a way to start the selection, and then after performing some signal separation as indicated below, perhaps fractal dimension estimators can pin down the answer more precisely. In any case, robustness seems to come with methods that do not require precise determination of distances between points on the strange attractor.

V. INVARIANTS OF THE DYNAMICS

Classification of the physical system producing one's observations is a standard problem for physicists. In an earlier section we discussed general matters related to invariants of the dynamics. Now we give more detail.

The basic idea is that one wants to identify quantities that are unchanged when initial conditions on an orbit are altered or when, anywhere along the orbit, perturbations are encountered. As usual we assume that underlying the observations $\mathbf{y}(k)$ there is a dynamical rule $\mathbf{y}(k+1)=\mathbf{F}(\mathbf{y}(k))$. For purposes of defining invariants, we need some notion of invariant distribution or invariant measure on the attractor. This issue is discussed at some length in the earlier review article of Eckmann and Ruelle (1985), and for us it is only necessary to agree that only one of the many possible candidates for invariant distribution seems to be stable under the influence of errors or noise. This is the natural density of points in the phase space, which tells us where the orbit has been, and whose integral over a volume of state space counts the number of points within that volume. The definition of this density is

$$\rho(\mathbf{x})=\frac{1}{N}\sum_{k=1}^{N}\delta^d(\mathbf{x}-\mathbf{y}(k)) \ , \tag{39}$$

in phase space of dimension d.

Now any function on phase space $g(\mathbf{x})$, when integrated with this density, is invariant under the dynamics $\mathbf{x}\rightarrow\mathbf{F}(\mathbf{x})$. Defining

$$\langle g \rangle = \int d^dx \, g(\mathbf{x})\rho(\mathbf{x}) \ ,$$

$$= \frac{1}{N}\sum_{k=1}^{N} g(\mathbf{y}(k)) \ , \tag{40}$$

we see that

$$\int d^dx \, g(\mathbf{F}(\mathbf{x}))\rho(\mathbf{x}) = \frac{1}{N}\sum_{k=1}^{N} g(\mathbf{y}(k+1))$$

$$= \frac{1}{N}[g(\mathbf{y}(N+1))-g(\mathbf{y}(1))]+\langle g \rangle \ , \tag{41}$$

and in the limit $N\rightarrow\infty$, the averages of $g(\mathbf{x})$ and $g(\mathbf{F}(\mathbf{x}))$ are the same. Since $\mathbf{F}(\mathbf{x})$ moves us along the orbit precisely one step, we see that, for long enough data sets, the physicist's meaning of $N\rightarrow\infty$, all $\langle g \rangle$ are unchanged by perturbations within the basin of attraction—all those points in state space which end up on an attractor—of the observations.

Quantities such as $\langle g \rangle$ can now be used to identify the system that produced the observations. The $\langle g \rangle$ are statistical quantities for a deterministic system and are the only quantities that will be the same for any long observation of the same system. Different chaotic orbits will, as we shall make quantitative below, differ essentially everywhere in phase space. The issue, then, is to choose a $g(\mathbf{x})$ that holds some interest for physical questions.

A. Density estimation

Before proceeding any further, we need to discuss the practical issues associated with numerical estimation of the invariant distribution, Eq. (39). Expressing the measure as a sum of delta functions is supremely convenient for defining other invariant quantities, as Eq. (40) demonstrates. But integrals of delta functions presuppose an absolutely continuous, infinitely resolvable state space, which is something most experimentalists are not familiar with.

Any measurement of a tangible variable necessarily imposes a partitioning of state space. The number of distinct, discrete states observable by any apparatus can be no larger than the inverse of the measurement resolution. Each element of this partition can be assigned a label, which we shall denote as i (typically this label might be some expression of the measurement value). In this discretized space, the state-space integral

$$\langle g \rangle = \int d^dx \, g(x)\rho(x) \tag{42}$$

becomes

$$\langle g \rangle = \sum_{i=1}^{N} g(x_i)p(x_i) \ , \tag{43}$$

where the probability of the ith element of the partition is

$$p(x_i) = \frac{\int_{x_i} d^d x \, \rho(x)}{\sum_i \int_{x_i} d^d x \, \rho(x)} = \frac{n(x_i)}{\sum_i n(x_i)} ,$$ (44)

with $n(x_i)$ equal to the number of counts in partition element i. These discrete estimates of the partition probabilities in the state space provide our approximation for the invariant measure. Equation (40) can be interpreted as the sparsely populated version of Eq. (43), resulting from taking so fine a partitioning of phase space that each partition element contains only one point.

This type of partition-based estimator is simply a multidimensional histogram and, as such, does not provide smooth distributions and can suffer from boundary effects. One way of improving upon traditional histograms is to use kernel density estimators, which replace the delta function in Eq. (39) with one's favorite integrable model for a localized, but nonsingular, distribution function. Tight Gaussians are a popular choice for the kernel, but scores of alternatives have been suggested on various grounds of optimality criteria (Silverman, 1986).

Kernel density estimators illuminate an alternative to the traditional *fixed-volume* methods for estimating densities/probabilities. Instead of fixing a uniform grid to partition the phase space and counting the number of observations in each equal-sized partition element, it is possible to base our estimates on the localized properties of the given data set around each individual point. *Fixed-mass* techniques perform a virtual partitioning of the phase space with overlapping elements centered on every point in the data set. The radius of each partition element, $r_{N_B}(x)$, is increased until a prespecified number N_B of neighboring points have been enclosed. In this way we accumulate a large number of overlapping partition elements, all containing the same number of points but having different spatial sizes. The local density, or probability, is then estimated as a function of the inverse of the bin radius:

$$\rho(x) \propto r_{N_B}(x)^{-d_E} ,$$ (45)

where d_E is the embedding dimension of the reconstructed phase space.

It will soon become clear that every time we can identify two broad classes of methods for implementing a type of algorithm, the particular properties of any given data set will dictate a hybrid technique for optimal performance. For example, if there is obvious clustering in the data, it would make sense to have the different clusters lie in different partition elements.

B. Dimensions

Attractor dimension has been the most intensely studied invariant quantity for dynamical systems. Much of the interest of the past decade was spawned by the reali-

zation that attractors associated with chaotic dynamics have a fractional dimension, in contrast to regular, or integrable, systems, which always have an integer dimension. Armed with this knowledge, a legion of researchers embarked on the search for evidence of fractional dimensions in experimental time series, to demonstrate the existence of chaos in the real world.

Since there are already a large number of very good reviews of dimensions and their estimation (Farmer *et al.*, 1983; Paladin and Vulpiani, 1987; Theiler, 1990), we shall not devote much space to a full discussion of all the relevant issues. In particular, we refer the reader to Farmer *et al.* (1983) for a more detailed discussion of the basic theoretical concepts and Theiler (1990) for a comprehensive review of many of the subtleties of estimation. Paladin and Valpiani (1987) provide a good discussion of the inhomogeneity of points distributed on an attractor, stating clearly how one determines quantitative measures of this inhomogeneity and how one looks for effects in the physics of this phenomenon.

Previous sections described the basic mathematical concept of the phase-space dimension as the (integer) number of quantities that need to be specified to fully identify the state of the system at any instant. In the case of ordinary differential equations, or discrete mappings, the phase-space dimension corresponds to the number of equations defining the evolution of each component of the state vectors. In many cases of physical interest, e.g., Navier-Stokes or any system of partial differential equations, this state-space dimension is infinite. However, in dissipative dynamical systems it is frequently the case that the system's asymptotic behavior relaxes onto a small, invariant subset of the full state space. The variety of dimensions about to be introduced describe the structure of these invariant subsets.

Beyond the simplest concept of dimension as the number of coordinates needed to specify a state is the geometrically related concept of how (hyper) volumes scale as a function of a characteristic length parameter:

$$V \propto L^D .$$ (46)

Planar areas scale quadratically with the length of a side, and volumes in real world space go as the cube of the side length. Contingent upon some workable generalization of the idea of volume, we can invert Eq. (46) to "define" a dimension:

$$D = \frac{\log V}{\log L} .$$ (47)

Early techniques for estimating the dimension used a covering of the attractor to calculate V. Consider a partitioning of the d_E-dimensional phase space with an ϵ-sized grid. Count how many of the partition elements contain at least one point from the sample data, and use this value as the measure of V at resolution ϵ. Then refine the phase-space partition by decreasing ϵ. Now count how many of these smaller partition elements are not empty. Continue this procedure until ϵ has spanned

a large range, hopefully at least several orders of magnitude. In theory, in the limit as $\epsilon \to 0$ the ratio in Eq. (47) describes the space-filling properties of the point set being analyzed. It is obvious that the largest value of D calculable with this *box-counting* algorithm is d_E. If the embedding dimension is not large enough to unfold the attractor's structure fully, we shall only be observing a projection of the structure, and much information will be obscured. Therefore the common practice is to repeat the described box-counting calculations in a number of increasing embedding dimensions. For lower-dimensional embeddings, the attractor's projection is expected to "fill" the space, resulting in an estimated fractal dimension equal to the embedding dimension. As the embedding dimension increases through the minimum required for complete unfolding of the geometric structure, the calculated fractal dimension saturates at the proper value. Figure 19 illustrates this basic technique with data from the Hénon map.

According to our geometric intuition, Eq. (46) takes "the amount" of an object to be equivalent to its volume. But this is not the only measure for how much of something there is. Theiler has nicely phrased Eq. (46) in the more general form

$$\text{bulk} \propto \text{size}^{\text{dimension}} . \tag{48}$$

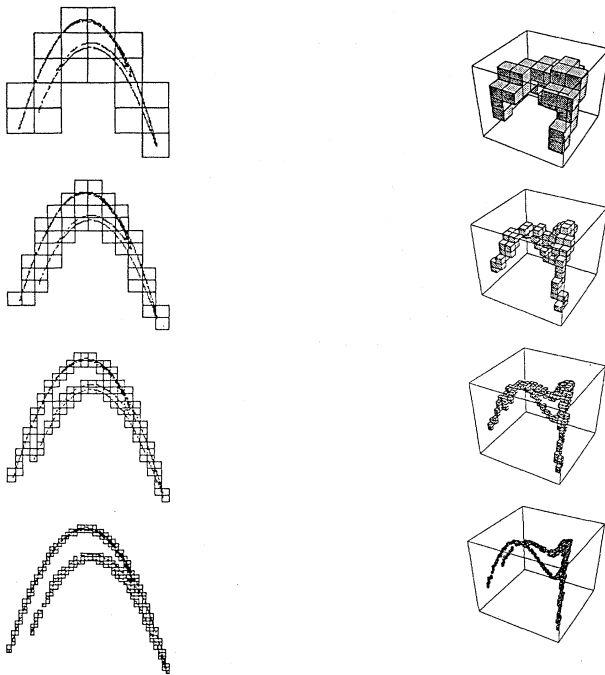This now invites us to consider other measures of bulk.

FIG. 19. Box counting for Hénon attractor in different embedding dimensions $d_E = 2, 3$ and at differing spatial resolutions.

1. Generalized dimensions

Let us now return to our discussion of invariant quantities by considering the case in which bulk is measured in terms of some moment of the invariant distribution. In other words, take $g(x) = \rho^p(x)$ in Eq. (40) and use $\langle g \rangle^{1/p}$ to measure bulk. Putting everything together we arrive at

$$
\begin{aligned}
D &= \lim_{r \to 0} \frac{\log(\text{bulk})}{\log(\text{size})} = \lim_{r \to 0} \frac{\log \langle g \rangle^{1/p}}{\log r} \\
&= \lim_{r \to 0} \frac{\log \left[\rho \sum \rho^p \right]^{1/p}}{\log r} = \lim_{r \to 0} \frac{1}{p} \frac{\log \sum \rho^{p+1}}{\log r} \\
&= \lim_{r \to 0} \frac{1}{q-1} \frac{\log \sum \rho^q}{\log r} = D_q .
\end{aligned}
$$

This definition of the *generalized dimension* D_q provides a whole spectrum of invariant quantities for $-\infty < q < \infty$. We are already familiar with $D_0 = -\log N_0 / \log r$; this is just the capacity estimated by the box-counting method.

Next we have D_1 to consider. The $q \to 1$ limit results in an indeterminate form, but one easy way to find the limit is with L'Hospital's rule:

$$D_1 = \lim_{r \to 0} \frac{1}{r} \frac{\frac{\partial}{\partial q} \left[\log \sum_i p_i^g \right]_{q=1}}{\frac{\partial}{\partial q} [q - 1]_{q=1}} \tag{49}$$

$$= \lim_{r \to 0} \frac{1}{r} \left[\frac{1}{\sum_i p_i^q} \sum_i p_i^q \log p_i \right]_{q=1} \tag{50}$$

$$= \lim_{r \to 0} \frac{\sum_i p_i \log p_i}{\log r} . \tag{51}$$

Note that, when $p_i = 1/N$ for all i, then $D_1 = D_0$. But when the p_i are not all equal $D_1 < D_0$. (In fact, in general $D_i \leq D_j$, if $i > j$.)

The difference between the two measures is explained by how we use our state-space partition. For D_0 the bulk is how many partition elements are nonempty. An element containing one sample contributes as much bulk as an element containing a million samples. D_1 is calculated from a bulk in which each partition element's contribution is proportional to how many samples it contains. So bulk in D_0 is simply the volume, whereas for D_1 bulk is equivalent to the mass.

Because of the functional form of the numerator in Eq. (51), D_1 is commonly called the *information dimension*. That numerator expresses the amount of information needed to specify a state to precision r, and the ratio defining the dimension tells us how fast this required information grows as r is decreased.

Next we have D_2, the so-called *correlation dimension*:

$$D_2 = \lim_{r \to 0} \frac{-\log \sum_i p_i^2}{\log r} . \qquad (52)$$

We see that the numerator constitutes a two-point correlation function, measuring the probability of finding a pair of random points within a given partition element, just as the numerator in the definition of D_1 measures the probability of finding one point in a given element.

Consequently, we can try to estimate the numerical value of that numerator by counting how many pairs of points have a separation distance less than some value r. In 1983 Grassberger and Procaccia (1983a, 1983b) suggested using

$$C_2(r) = \int d^d x \, \rho(\mathbf{x}) n(r, \mathbf{x})$$

$$= \frac{2}{N(N-1)} \sum_{i \neq j}^{N} \theta(r - |\mathbf{y}(j) - \mathbf{y}(i)|) \qquad (53)$$

as a simple and computationally efficient means of estimating $\sum_i p_i^2$. With the introduction of the Grassberger-Procaccia algorithm the floodgates were opened for researchers looking for chaos in experimental data. From fluid mechanics and solid-state physics, to epidemiology and physiology, to meteorology and economics, there was a staggering volume of papers publishing estimates of fractal dimensions.

2. Numerical estimation

Evaluation of the D_q has been the subject of numerous inquiries (Smith, 1988; Ruelle, 1990; Theiler, 1990). Much of the discussion centers on how many data are required to determine the dimension reliably. Generally speaking, large quantities of data are necessary to achieve accurate approximations for the density of points in the different regions of the attractor, and a good signal-to-noise ratio is required to probe the fine structure of any fractal sets in the attractor. Clearly, the required number of points must scale as some function of the dimension being estimated. The following are just a few opinions that can be found in the literature:

● Bai-Lin Hao (1984) points out the most basic limit. To examine a 30-dimensional attractor "the crudest probe requires sampling two points in each direction and that makes $2^{30} \approx 10^9$ points, a value not always reached in the published calculations."

● Theiler (1990) tells us $N \sim \Theta^D$. "Experience indicates that Θ should be of the order of 10, but experience also indicates the need for more experience."

● Leonard Smith (1988) claims that to keep errors below 5 percent one must have $N \geq 42^M$, where M is the largest integer less than the set's dimension.

● On the other hand, Abraham *et al.* (1986) argue that "Increasing N to the point where $L/N^{1/D}$ is less than a characteristic noise length is not likely to provide any more information on the attractor's structure."

● Ruelle (1990) and Essex and Nerenberg (1991) argue that if one estimates a dimension larger than $2 \times \log_{10}(N)$ that one has only counted something related to the statistics of the sample size N. This would mean that, for estimating a dimension d, a data set of *at least* size $10^{d/2}$ is required.

It is not uncommon to see attempts to overcome the limitations imposed by small data sets by measuring the system more frequently. However, as discussed earlier in Sec. IV, this is not an effective tactic. The raw number of points is not what matters; it is the number of trajectory segments, how many different times any particular locale of state space is revisited by an evolving trajectory, that counts.

Identifying the scaling region for estimating the dimension is another significant challenge. Standard practice calls for accumulating the data's statistics and then plotting $\log C(r)$ versus $\log r$ for a set of increasing embedding dimensions. For embedding dimensions smaller than the minimum required for complete unfolding of the attractor, the data will fill the entire accessible state space, and the slope of the plot will equal the embedding dimension. As the embedding dimension increases, the slope of the $\log C(r)$ versus $\log r$ plot should saturate at a value equal to the attractor's dimension. The dimension is defined as the slope of this plot in the $r \to 0$ limit, but this region of the plot is always dominated by noise and the effects of discrete measurement channels. Therefore one hopes to identify a scaling region at intermediate length scales, where a constant slope allows reliable estimation of the dimension by Eq. (47). Working with large quantities of machine-precision computer-generated data, it is not difficult to produce gorgeous plots with clearcut scaling regions spanning many decades. But when dealing with a limited quantity of moderately noisy data, the scaling region may not be identifiable. For example, one of the characteristic signatures of oversampled data is the appearance of a "knee" in the correlation integral plot, separating regions with different scaling behavior. The knee is caused by enhanced numbers of near neighbors at small scales due to the strong correlation between temporally successive points. We refer the reader to the review by Theiler (1990) for a more complete and detailed explanation of the subtleties of dimension estimation, along with a survey of proposed refinements for increasing the efficiency and reliability of the algorithms.

Motivated by the simple fact that many interesting time series are extremely short or noisy, substantial effort has gone into clarifying the statistical limits of the techniques (Smith, 1988; Theiler, 1990). At the other extreme is the question of how large a dimension can be investigated with these techniques. Although many experienced investigators believe the data requirements limit the applicability of these algorithms to systems of dimension 6 or less, several groups claim encouraging results up to 20 dimensions with careful application of the Grassberger-Procaccia algorithm.

In addition to the difficulties in accumulating sufficient statistics for the reliable estimation of dimensions, there is the uncomfortable fact that one can still be tricked by the algorithm's output. A controversial example was pointed out by Osborne and Provenzale (1989), who claimed that time series generated by colored stochastic processes with randomized phase are typically identified as finite dimensional by standard correlation integral methods. In one sense, this can be explained as a consequence of the fact that the data are not stationary, since the correlation length grows with the length of the time series. Although Theiler (1991) identified the sources of their paradoxical results and suggested several procedures for the avoidance of such anomalies, the fact remains that the identification of systems with $1/f^{\alpha}$ power spectra can be a frustrating task. In Sec. VIII we shall see that filtered signals are also susceptible to misinterpretation by standard dimension-estimating algorithms.

More important to attend to is the use to which the D_q are put for establishing that one has a low-dimensional chaotic signal. If this is indeed one's interest, then using the false-nearest-neighbor method or one of the variants will do that with many fewer data and much less computation. If one wishes to see an invariant on the attractor saturate, then choose one that is less computationally intensive. If one really wants the numerical value of D_q, then *caveat emptor*.

With all this warning, we nonetheless quote some familiar D_2, which come from graphs of good clean correlation functions as a function of $\log[r]$. From these we are able to read off for the Hénon map $D_2 \approx 1.26$, for the Lorenz system $D_2 \approx 2.06$, and for the Ikeda map $D_2 \approx 1.8$. For the hysteretic circuit of Pecora and Carroll we see the results in Fig. 20. In earlier sections we have loosely referred to the dimension of the attractor as d_A and made no distinction between the value of D_2 as quoted here and the values of the box-counting D_0 dimension and information dimension D_1 which appear in
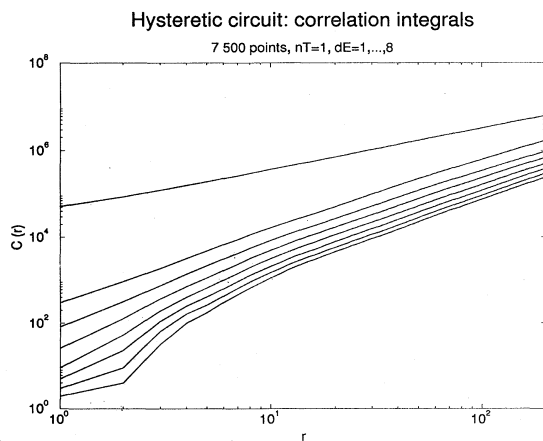
more precise statements. In practice, these dimensions are all numerically so similar that there is no advantage in distinguishing them from a signal analysis point of view. A clear discussion of how and when the various fractal dimensions are similar is given by Beck (1990), who relates this question to the singularity of the invariant distribution on the attractor.

C. A segue—from geometry to dynamics

One frustrating aspect of dimension estimation is the meager quantity of useful information gained for the amount of effort expended. Fractal dimensions are a convenient label for categorizing systems, but they are not very useful for practical applications such as modeling or system control. The dimensions describe how the sample of points along a system orbit tend to be distributed spatially, but there is no information about the dynamic, temporally evolving, structure of the system.

An interesting connection between the fractal dimensions discussed in the previous section and the stability properties of the system dynamics was conjectured by Kaplan and Yorke in 1980.

Assume the state space can be decomposed at each point into local stable and unstable subspaces, with a basis set containing expanding and contracting directions. The rates at which these deformations occur in each direction are called the Lyapunov numbers. With the typical ordering $L_1 > L_2 \ldots > L_u \geq 1 > L_s \ldots > L_d$, and the existence of an attractor requires $\prod_{j=1}^{d} L_j < 1$. A dynamical system is said to be chaotic if at least one of the $L_j > 1$.

Consider a partitioning of the state space with d-dimensional elements of side length ϵ, i.e., the number of partition elements required to cover a unit d-cube is ϵ^{-d}. Along the lines of a box-counting algorithm, let us count the number of partition elements needed to cover our attractor, and call this number $N(\epsilon)$.

Now consider one of these elements of the partition that covers the attractor as a fiducial volume, and let this volume evolve according to the system dynamics. Each axis of the fiducial volume will be scaled by a factor proportional to that direction's Lyapunov number. Directions with a Lyapunov number of magnitude less than 1 will contract, meaning one ϵ width will still be needed to cover these directions. Directions with a Lyapunov number greater than one will be expanded, thereby requiring more ϵ widths to cover the evolved fiducial volume in these directions.

Let us use this information to estimate the number of partition elements needed to cover the attractor if the partition resolution is improved. Take the partition spacing to be $C\epsilon$, where C is some constant. How is $N(C\epsilon)$ related to $N(\epsilon)$? Following the same reasoning as above, we expect that each direction associated with a Lyapunov number less than C will contribute a factor of 1.0, i.e., leave N unchanged. But each direction associated with a Lyapunov number greater than C will contrib-

Hysteretic circuit: correlation integrals

7 500 points, nT=1, dE=1,...,8

FIG. 20. Sequence of correlation integrals $C_2(r)$ for hysteretic circuit data.

ute a factor of L_i/C to N:

$$N(C\epsilon) = N(\epsilon) \prod_{i=1}^{d} \max\left[\frac{L_i}{C}, 1\right] . \tag{54}$$

Recall the basic scaling relation that motivated our box-counting algorithm:

$$V \propto N(\epsilon) \sim \epsilon^{-D} . \tag{55}$$

So

$$\frac{N(C\epsilon)}{N(\epsilon)} = \prod_{i=1}^{d} \max\left[\frac{L_i}{C}, 1\right] = \frac{(C\epsilon)^{-D}}{\epsilon^{-D}} = C^{-D} , \tag{56}$$

yielding yet another dimension measure:

$$D_L = -\frac{\log\left[\prod_{i=1}^{d} \max\left[\frac{L_i}{C}, 1\right]\right]}{\log C} . \tag{57}$$

Now what value should be used for C? Consider using one of the Lyapunov numbers, say L_{K+1}. Then

$$D_L = -\frac{\log\left[\prod_{i=1}^{d} \max\left[\frac{L_i}{L_K+1}, 1\right]\right]}{\log L_{K+1}}$$

$$= -\frac{\log\left[\prod_{i=1}^{K} \frac{L_i}{L_{K+1}}\right]}{\lambda_{K+1}}$$

$$= \frac{K\lambda_{K+1} - \sum_{i=1}^{K} \lambda_i}{\lambda_{K+1}} = K - \frac{\sum_{i=1}^{K} \lambda_i}{\lambda_{K+1}} ,$$

where the λ_i, the logarithms of the L_i, are called the Lyapunov exponents.

Which Lyapunov exponent should be picked for λ_k? In the spirit of the definition of the capacity D_0 we want to choose the value that minimizes D_L. Note that if all the λ_i are less than zero the attractor is a stable fixed point and has dimension zero. If the only non-negative exponents are zero, the attractor is a limit cycle. Multiple null exponents correspond to the number of incommensurate frequencies in a quasiperiodic system, which is also the system's dimension. When at least one exponent is positive, the last equation is to be minimized with respect to λ_k, to provide the most efficient covering of the invariant set.

To find the optimal λ_k consider the difference in D_L calculated with λ_k or λ_{k+1}:

$$\Delta D_L = \left[k+1 - \frac{\sum_{i=1}^{k+1} \lambda_i}{\lambda_{k+2}}\right] - \left[k - \frac{\sum_{i=1}^{k} \lambda_i}{\lambda_{k+1}}\right] = 1 - \sum_{i=1}^{k} \lambda_i \left[\frac{1}{\lambda_{k+2}} - \frac{1}{\lambda_{k+1}}\right] - \frac{\lambda_{k+1}}{\lambda_{k+2}}$$

$$= 1 - \left|\frac{\lambda_{k+1}}{\lambda_{k+2}}\right| + \sum_{i=1}^{k} \lambda_i \left[\frac{|\lambda_{k+1}| - |\lambda_{k+2}|}{|\lambda_{k+1}||\lambda_{k+2}|}\right] = \left[\frac{|\lambda_{k+2}| - |\lambda_{k+1}|}{|\lambda_{k+2}|}\right] \left[1 - \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|}\right] .$$

Assuming that $\lambda_{k+1}, \lambda_{k+2} < 0$, and therefore $|\lambda_{k+1}| < |\lambda_{k+2}|$, we see that the first factor is a positive quantity less than one. The second factor is a negative quantity, and D_L is decreasing, so long as $\sum_{i=1}^{k} \lambda_i > |\lambda_{k+1}|$. Therefore we take

$$D_L = K - \frac{\sum_{i=1}^{K} \lambda_i}{\lambda_{K+1}} , \tag{58}$$

such that $\sum_{i=1}^{K} \lambda_i > 0$, and $\sum_{i=1}^{K+1} \lambda_i < 0$.

An exact correspondence between D_L, constructed from measures of the stability properties of the dynamics, and the dimensions defined in terms of moments of the measure of the limit set is not obvious. The derivation of the quantity is based upon a clever twist in the traditional

description of the capacity D_0. By constructing a refinement of the phase-space partitioning that is custom tailored to the way phase space is deformed by the dynamics, we automatically create an efficient covering for the invariant set, and we know how the population of the covering scales in the $\epsilon \rightarrow 0$ limit. However, that efficient covering is defined in terms of the Lyapunov exponents, which are invariant quantities, defined as averages over the invariant measure. Therefore it seems that D_L might equal D_1.

But which geometric-based dimension is equal to D_L is not very relevant from the time series analyst's viewpoint. Rarely does the real world bestow upon the experimentalist a data set of good enough quality to distinguish among these various dimensions. The important issue is which definition can motivate an efficient and

stable numerical algorithm for dimension estimation. One characteristic the D_q all share is that they become unwieldy and demand too many data as the dimension of the attractor approaches and exceeds ten. Being based on the dynamics rather than the distribution of sample points, the Lyapunov dimension allows us to investigate much higher dimensions so long as we can get reliable estimates for the Lyapunov exponents.

D. Lyapunov spectrum

Dimensions characterize the distribution of points in the state space; Lyapunov exponents describe the action of the dynamics defining the evolution of trajectories. Dimensions only deal with the way measure is distributed throughout the space, whereas Lyapunov exponents examine the structure of the time-ordered points making up a trajectory.

To understand the significance of the spectrum of Lyapunov exponents, consider the effects of the dynamics on a small spherical fiducial hypervolume in the (possibly reconstructed) phase space. Arbitrarily complicated dynamics, like those associated with chaotic systems, can cause the fiducial element to evolve into extremely complex shapes. However, for small enough length scales and short enough time scales the initial effect of the dynamics will be to distort the evolving spheroid into an ellipsoidal shape, with some directions being stretched and others contracted. The primary, longest, axis of this ellipsoid will correspond to the most unstable direction of the flow, and the asymptotic rate of expansion of this axis is what is measured by the largest Lyapunov exponent. More precisely, if the infinitesimal radius of the initial fiducial volume is called $r(0)$, and the length of the ith principal axis at time t is called $l_i(t)$, then the ith Lyapunov exponent can be defined as

$$\lambda_i = \lim_{t \to \infty} \frac{1}{t} \log \frac{l_i(t)}{r(0)} \ . \tag{59}$$

By convention, the Lyapunov exponents are always ordered so that $\lambda_1 > \lambda_2 > \lambda_3 \cdots$.

Equivalently, the Lyapunov exponents can be seen to measure the rate of growth of fiducial subspaces in the phase space. λ_1 measures how quickly linear distances grow—two points initially separated by an infinitesimal distance ϵ will, on average, have their separation grow as $\epsilon e^{\lambda_1 t}$. The two largest principal axes define an area element, and the sum $\lambda_1 + \lambda_2$ determines the rate at which two-dimensional areas grow. In general, the behavior of d-dimensional subspaces is described by the sum of the first d exponents, $\sum_{j=0}^{d} \lambda_j$.

E. Global exponents

The spectrum of Lyapunov exponents is determined by following the evolution of small perturbations to an orbit by the *linearized dynamics* of the system. Suppose we

have established by phase-space reconstruction or direct observation an orbit $y(k); k = 1, 2, \ldots, N$ to which we make a small change at "time" unity: $w(1)$. The evolution of the dynamics will now be

$$y(k+1) + w(k+1) = F(y(k) + w(k)) \ , \tag{60}$$

which will determine the $w(k)$. If the $w(k)$ start small and stay small, then

$$w(k+1) = F(y(k)) - y(k+1) + DF(y(k)) \cdot w(k) + \cdots$$
$$= DF(y(k)) \cdot w(k) \ . \tag{61}$$

The evolution of the perturbation can be written as

$$w(L+1) = DF(y(L)) \cdot DF(y(L-1)) \cdots DF(y(1)) \cdot w(1)$$
$$= DF^L(y(1)) \cdot w(1) \ , \tag{62}$$

which defines our shorthand notation, $DF^L(y(1))$, for the composition of L Jacobian matrices $DF(x)$ along the orbit $y(k)$.

In 1968 Oseledec (1968) proved the important *multiplicative ergodic theorem*, which includes a demonstration that the eigenvalues of the orthogonal matrix $DF^L(x) \cdot [DF^L(x)]^T$ are such that the matrix

$$\lim_{L \to \infty} \{DF^L(x) \cdot [DF^L(x)]^T\}^{1/2L} \tag{63}$$

exists and has eigenvalues $\exp[\lambda_1], \exp[\lambda_2], \ldots, \exp[\lambda_d]$ for a d-dimensional dynamical system which are independent of x for almost all x within the basin of attraction of the attractor. The λ_a are the global Lyapunov exponents. Their independence of where one starts within the basin of attraction means that they are characteristic of the dynamics, not the particular observed orbit. We call them global because the limit $L \to \infty$ means they are, in a sense we shall make more precise, a property of the global aspects of the attractor.

Oseledec also proved the existence of the eigendirections of $DF^L(y(1))$. These are linear-invariant manifolds of the dynamics: points along these eigendirections stay along those directions under action of the dynamics as long as the perturbation stays small. Ruelle (1979) and others extended this to the nonlinear manifolds and to more complicated spaces. These eigendirections depend on where one is on the attractor, but not on where one starts in order to get there. That is, if we want to know the eigendirections of DF^L at some point $y(L)$ and we begin at $y(1)$ and take L steps, or begin at $y(2)$ and take only $L-1$ steps, the eigendirections will be the same as long as L is large enough. In practice, on the order of ten steps are usually required to establish the eigendirections accurately.

The multiplicative ergodic theorem of Oseledec is a statement about the eigenvalues of the dynamics in the tangent space (Eckmann and Ruelle, 1985) to $x \to F(x)$ and is motivated by a statement about linearized dynamics of the mapping. It is also a characterization via the λ_a of nonlinear properties of the system. Basically it is an analysis of the behavior of the nonlinear system in the

neighborhood of an orbit on the strange attractor. By following the tangent-space behavior (the linearized behavior near an orbit) locally around the attractor, we extract statements about the global nonlinear behavior of the system. A compact geometrical object, such as a strange attractor, which has positive Lyapunov exponents cannot arise in a globally linear system. The stretching associated with unstable directions of $DF(y(k))$ locally and the folding associated with the dissipation are the required ingredients for the theorem.

If any of the λ_a are positive, then over the attractor small perturbations will grow exponentially quickly. Positive λ_a are the hallmark of chaotic behavior. If all the λ_a are negative, then a perturbation will decrease to zero exponentially quickly, and all orbits are stable. In a dissipative system, the sum of the λ_a must be negative, and the sum governs the rate at which volumes in phase space shrink to zero. If the underlying dynamics is governed by a differential equation, one of the λ_a will be zero. This corresponds to a perturbation directly along the vector field, and such a perturbation simply moves one along the same orbit on which one started, so nothing happens in the long run. Indeed, one can tell, in principle, if the source is governed by differential equations or by a finite time map by the presence or absence of a zero global Lyapunov exponent.

A Hamiltonian system preserves phase-space volume and respects the other invariants reflecting the symplectic symmetry (Arnol'd, 1978). This leads to two consequences: (i) the sum of all λ_a is zero, and (ii) the λ_a come in pairs, which are equal and opposite. If λ_a appears in the collection of Lyapunov exponents, so does $-\lambda_a$. For Hamiltonian dynamics there are two zero global Lyapunov exponents, one associated with the conservation of energy, the second associated with the fact that the evolution equations are differential.

An interesting result due to Pesin (1977) relates, under fairly general circumstances, the sum of the positive exponents to the KS entropy:[3]

$$h(X) = \sum_{\lambda_a > 0} \lambda_a . \qquad (64)$$

F. Ideal case: Known dynamics

If one knows the vector field, determination of the λ_a is straightforward, though numerically it poses some challenges (Greene and Kim, 1986).[4] The main challenge

---

[3]See also the paper by D. Ruelle (1978), which gives a bound on $h(X)$ in rather general settings.

[4]This paper also presents eigendirections associated with Lyapunov exponents. These are not the same eigendirections mentioned above for DF^L, they are the eigendirections associated with $DF^L(y(1))\cdot[DF^L(y(1))]^T$. The eigendirections of DF^L need not be orthogonal.

comes from the fact that although each $DF(y)$ has eigenvalues more or less like $\exp[\lambda_a]$, the composition of L Jacobians, $DF^L(y)$, has eigenvalues approximately $\exp[\lambda_1 L]$, $\exp[\lambda_2 L]$, ..., $\exp[\lambda_d L]$, and since $\lambda_1 > \lambda_2 > \cdots > \lambda_d$, as L becomes large, the matrix is terribly ill conditioned. The diagonalization of such a matrix poses serious numerical problems. Standard **QR** decomposition routines do not work as well as one would like, but a recursive **QR** decomposition due to Eckmann, Kamphorst, Ruelle, and Ciliberto (Eckmann *et al.*, 1986) does the job.[5] The problem in the direct **QR** decomposition is keeping track of the orientation of the matrices from step to step, so they propose to write each Jacobian in its **QR** decomposition as

$$DF(y(i))\cdot Q(i-1) = Q(i)\cdot R(i) , \qquad (65)$$

where we begin with $Q(0)$ as the identity matrix. $DF(y(1))$ is then decomposed as $Q(1)\cdot R(1)$ as usual with the **QR** decomposition of matrices. Then we write

$$DF(y(2))\cdot Q(1) = Q(2)\cdot R(2) , \qquad (66)$$

so we have for the product

$$DF(y(2))\cdot DF(y(1)) = Q(2)\cdot R(2)\cdot R(1) . \qquad (67)$$

Continuing, we have for $DF^L(y(1))$

$$DF^L(y(1)) = Q(L)\cdot \Pi^L_{k=1} R(k) . \qquad (68)$$

The problem of directions within the product of matrices has been handled step by step; at each step of this recursive procedure no matrix $R(k)$ is much larger than $\exp[\lambda_1]$ and the condition number is more or less $\exp[\lambda_1 - \lambda_d]$, which is quite reasonable for numerical accuracy.

If we have differential equations rather than mappings, or more precisely, if we want to let the step in the map approach zero, we can simultaneously solve the differential equation and the variational equation for the perturbation around the computed orbit. Then a recursive **QR** procedure can be used for determining the Lyapunov exponents. Greene and Kim (1986) use a recursive singular-value decomposition, and the idea of the method is quite the same.

G. Real world: Observations

The problem of determining the Lyapunov exponents when only a scalar observation $s(n)$ is made is another challenge. As always, the first step is reconstruction of the phase space by the embedding techniques described in Sec. IV. Once we have our collection of embedded vectors $y(n)$ there are two general approaches to estimating the exponents. In one case an analytic approach is

---

[5]A similar recursive method was shown to us independently by E. N. Lorenz (private communication, 1990).

adopted when a predictive model is available for supplying values for the Jacobians of the dynamical rules. Lacking such detailed information about the dynamics, one can still try to estimate the larger exponents by tracking the evolution of small sample subspaces through the tangent space.

1. Analytic approach

The problem is to reconstruct the phase space and then estimate numerically the Jacobians $DF(y(k))$ in the neighborhood of each orbit point. Then the determination of the Lyapunov exponents by use of a recursive QR or singular-value decomposition is much the same as before. To estimate the partial derivatives in phase space we evidently must use the information we have about neighbors of each orbit point on the attractor. The idea (Sano and Sawada, 1985; Eckmann *et al.*, 1986; Brown *et al.*, 1990, 1991) is to make local maps from all points in the neighborhood of the point $y(k)$ to their mapped image in the neighborhood of $y(k+1)$. In the early literature (Sano and Sawada, 1985;[6] Eckmann *et al.*, 1986) this map was taken to to be locally linear, but this accurately gives only the largest exponent, since the numerical accuracy of the local Jacobians is not very good and when one makes mistakes in each of the elements of the composition of Jacobians that one is required to diagonalize to determine the λ_a, those errors are exponentially compounded by the ill conditioned nature of the problem (Chatterjee and Hadi, 1988). The problem is that the burden of making the local map *and* the burden of being an accurate Jacobian were both put on the $d_E \times d_E$ matrix in the local linear map. The solution is to allow for a larger class of local neighborhood-to-neighborhood maps, and then extract the Jacobian matrix from that map.

This means that locally in state space, that is, near a point $y(n)$ on the attractor, one approximates the dynamics $x \rightarrow F(x)$ by

$$F_n(x) = \sum_{k=1}^{M} c_n(k) \phi_k(x) , \tag{69}$$

with the $\phi_k(x)$ a set of basis functions chosen for convenience or motivated by the data. The coefficients $c_n(k)$ are then determined by a least-squares fit minimizing the residuals for taking a set of neighbors of $y(n)$ to a set of neighbors of $y(n+1)$. The numerical approximation to the local Jacobian then arises by differentiating this approximate local map.

The burden of making the map and achieving a numerically accurate Jacobian are separated by this approach,

---

[6]This paper also introduces the local-phase-space method for finding $DF(y(n))$, but uses only local linear maps for this purpose. The method is unable to determine all exponents, just the largest exponent.

and quite accurate values for the full Lyapunov spectrum are thus achieved. In the papers of Brown *et al.* (Bryant *et al.*, 1990; Brown *et al.*, 1991), local polynomials were used, and that was the method of Briggs (1990) as well. Various basis functions have been used for this: radial basis functions (Parlitz, 1992), sigmoidal basis functions (called neural networks in popular parlance) (Ellner *et al.*, 1991; McGaffrey *et al.*, 1992), and probably other bases as well.

Typical of the results one can achieve are the exponents shown in Table II for the Lorenz attractor, where it is known that $\lambda_1 = 1.51$, $\lambda_2 = 0.0$, and $\lambda_3 = -22.5$. We achieve high accuracy in establishing these values from data on the $x(n)$ observable only. The table comes from making a local polynomial fit from neighborhood to neighborhood in the phase space of the Lorenz attractor, reconstructed from the observations $x(n)$. From this local map numerical values of the Jacobian are read off and then the eigenvalues of products of Jacobians are found by the recursive method indicated. In the data used for $x(n)$ the parameter values are $\sigma = 4$, $b = 10$, and $r = 45.92$. We do not use the equations of the Lorenz model in determining these values for the λ_a, but from them we can establish a check on the results because the rate of phase-space contraction in the Lorenz model is $\exp[-(\sigma + b + 1)]$, and by the general properties of the Lyapunov exponents this must also be equal to $\exp[\lambda_1 + \lambda_2 + \lambda_3]$. This check is satisfied very accurately by the numerical results.

All these results are quite sensitive to the presence of contamination by other signals. This is quite natural, since the evaluation of $DF(y)$ is quite sensitive to distances in phase space, and inaccurate determination of distances leads to inaccurate determination of DF. In particular, achieving an accurate value of $DF(y(k))$ is a task in extrapolating from measurements at finite distances in phase space to values on the orbit, i.e., zero distance, and that is just where noise or contamination has the most effect. The effect of noise is shown in Fig. 21,
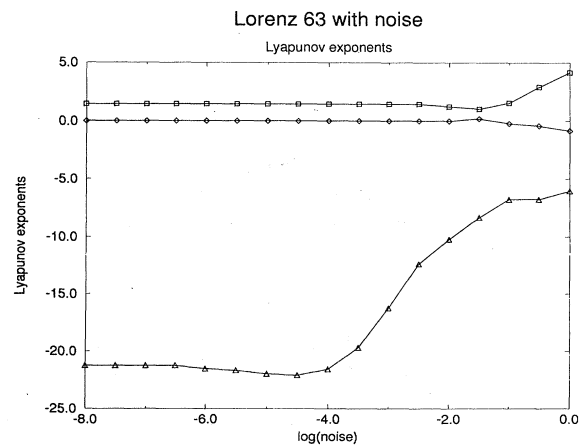

FIG. 21. Global Lyapunov exponents for Lorenz attractor in the presence of noise.

TABLE II. Global Lyapunov exponents computed from $x(t)$ data for the Lorenz attractor. The order of the neighborhood-to-neighborhood polynomial map is varied. The correct values for these exponents are $\lambda_1 = 1.50$, $\lambda_2 = 0.0$, and $\lambda_3 = -22.5$.

Order of polynomial used in local fit	Global Lyapunov exponents for the Lorenz system		
	λ_1	λ_2	λ_3
Linear	1.549	−0.094 70	−14.31
Quadratic	1.519	−0.026 47	−20.26
Cubic	1.505	−0.005 695	−22.59
Quartic	1.502	−0.002 847	−22.63
Quintic	1.502	−0.000 387	−22.40

where the three Lyapunov exponents for the Lorenz system are evaluated from $x(n)$ data for various levels of uniform random numbers added to the data. The sensitivity is clearly seen here.

2. Trajectory tracing method

As an alternative to finding numerical values for the local Jacobians of the dynamics along the orbit, one can attempt to follow small orbit differences (Wolf *et al.*, 1985) and small areas defined by two orbits, and then small subvolumes defined by three or more orbits to sequentially extract values for the Lyapunov exponents. These methods, while popular, almost always give accurate values for only the largest Lyapunov exponent λ_1 due to inevitable numerical errors. Nonetheless, we describe this method and alert the reader to its sensitivity.

Assume we have a large collection of experimental points from a long time series. What space our data is in is not important, whether it be the natural space of the defining dynamics or an embedding by time delays or whatever. We want to find points in our sample library that have close neighbors from different temporal segments of the library. Call these points $x(t)$ and $y(t)$. We then keep track of how they deviate under dynamical evolution until they become separated by a distance greater than some threshold and/or another trajectory segment passes closer to the presently monitored one. When either of these conditions is met, the ratio of

$$\frac{1}{n}\frac{\|y(n+t)-x(n+t)\|}{\|y(t)-x(t)\|} \tag{70}$$

is calculated and a new pair of neighboring points is observed. The largest Lyapunov exponent is estimated by the log of the long-term average of the above quantity. Strictly speaking, we want the asymptotic value as $n \to \infty$. In practice, we are happy to be able to trace a trajectory in this manner for several dozen characteristic time steps.

If a large sample of observation points is available, we may adopt a numerical trajectory-tracing technique motivated by an alternative, but equivalent, definition of the largest Lyapunov exponent:

$$\lambda_{\max} = \lim_{n \to \infty} \frac{1}{n}\log\|\mathbf{DF}^L\mathbf{u}\|, \tag{71}$$

where $\mathbf{u}$ is a nearly randomly oriented unit vector perturbation in the tangent space. We say *nearly* because if $\mathbf{u}$ should just happen to fall precisely along the stable manifold, it would remain within that invariant set and not relax onto the most unstable direction.

Our task is to find points in our sample library that are very close together and watch how trajectories specified by following the points separate. In locating the initial neighboring points we must not consider points that are from the same temporal segment of the library. Let us denote the starting point of our reference orbit as $r(0)=\mathbf{y}(i)$ and call its nearest spatial neighbor $v_1(0)=\mathbf{y}(j)$. We require that i and j differ by some minimum decorrelation length.

Now consider following the evolving trajectories emanating from these two points. Initially the points are separated by some distance $\|\Delta_0\|$, and we want to see how this distance grows under dynamical evolution. Therefore we continue to calculate a series of separation distances

$$\|\Delta_k\| = \|r(k)-v_1(k)\| = \|\mathbf{y}(i+k)-\mathbf{y}(j+k)\| \tag{72}$$

until we find a value of Δ_k exceeding a present threshold value. Typically this threshold will be some intermediate distance between the smallest resolvable distances and the overall size of the attractor as a whole. $\|\Delta_0\|$ is assumed to start off at the smallest length scale, but when $\|\Delta_k\|$ has grown to some sizable fraction of the attractor's extent we expect other nonlinear mechanisms of enfoldment to dominate the further evolution of Δ.

Therefore we attempt to locate another point $v_2(0)$ in our sample library close to $r(k)$, lying as near as possible to the line segment connecting $r(k)$ and $v_1(k)$. Things can get tricky here as we try to trade off between nearness to $r(k)$ and nearness to the separation segment from $r(k)$ to $v(k)$. Locating $v_2(0)$ close to $|r(k)-v_1(k)|$ is important, because in addition to being stretched by the unstable components of the nonlinear dynamics, Δ is rotated into alignment with the unstable manifold. In order to get an accurate estimate of the expansion properties of the dynamics, we want to keep all future Δ's

aligned with the unstable direction.

When we reinitialize our neighboring trajectory with $v_2(0)$, we store the final value of Δ_k for s_1 along with its terminating value of $k = k_1^{\max}$. Then we iterate the procedure, following $r(k)$ further, with its new neighbor v_2, monitoring the separation:

$$\Delta_{k_1^{\max}+k} = \|r(k_1^{\max}+k) - v_2(k)\| \ . \tag{73}$$

We continue this procedure for as long as the data or computing resources allow. Then we calculate the ratio

$$\Delta = \frac{\Delta_{k_1^{\max}}}{\Delta_0} \frac{\Delta_{k_2^{\max}}}{\Delta_{k_1^{\max}+1}} \cdots \frac{\Delta_{k_M^{\max}}}{\Delta_{k_{M-1}^{\max}+1}} \tag{74}$$

and estimate

$$\lambda_1 = \frac{1}{k_M^{\max}} \log\Delta \ . \tag{75}$$

To determine the second Lyapunov exponent we repeat the above tracing procedure, but monitor the evolution of area elements rather than separation lengths. Consider the same starting points used in the above example, $r(0)$ and $v_1(0)$. We want to augment this pair with another neighboring point $w_1(0)$, which is close to $r(0)$ while being in a direction orthogonal to δ_0, the separation between $r(0)$ and $v_1(0)$. Call the separation vector between $r(0)$ and $w_1(0)$ γ_0. Then δ_0 and γ_0 define a fiducial area element, which we can try to follow under dynamical evolution. We again follow $r(k)$, $v_1(k)$, and now $w_1(k)$ also, at each point calculating the area spanned by δ_k and γ_k.

However, following this two-dimensional element introduces further complications because, in general, γ_k will tend to become aligned with δ_k along the unstable direction. Therefore, in addition to reinitializing v and w when the spanned area grows above a threshold value, we must carefully reorient w to keep it as orthogonal to v as the data resources permit.

Clearly the trajectory tracing method requires a substantial quantity of data to facilitate the aligned updating procedure. If we are unable to locate nearby update points in the required directions, our exponents will be underestimated, since the maximal expansion rates will not be observed. Although the basic idea easily generalizes to higher-dimensional subspace elements for estimating the rest of the exponents, the precision and quantity of data required, especially in larger dimensions, makes this technique impractical for anything beyond the largest couple of exponents.

3. Spurious exponents

One difficulty associated with working in reconstructed phase spaces is that artifacts arising from inefficiencies of representation inherent to the *ad hoc* embedding can be annoying. A good example of such an annoyance is the presence of spurious Lyapunov exponents when the dimension of the reconstructed space exceeds the dimension of the dynamics' natural space d_N.

Analysis of a dynamical system in a d-dimensional space necessarily produces d Lyapunov exponents. And Takens' theorem tells us that the embedding space may have to be expanded to $d = 2d_N + 1$. So it is possible, even in the ideal noise-free case, to have $d_N + 1$ spurious exponents in addition to the d_N real ones.

Takens' theorem assures us that d_N of the exponents we calculate in a large enough embedding space will be the same as those we could calculate from a complete set of measurements in the system's natural space. But little was known about the behavior of the spurious exponents until recent work clarified the robustness of the numerical estimation techniques.

First clues were provided by Brown *et al.* (Bryant *et al.*, 1990; Brown *et al.*, 1991) when they investigated the use of higher-order polynomial interpolants for modeling the dynamics in the reconstructed phase space. Identifying the directions associated with each computed Lyapunov exponent, they demonstrated that very few sample points could be found along the directions of the spurious exponents. Essentially, the directions of the real exponents were thickly populated, indicating there was actual dynamical structure along these directions. The directions of the artificial exponents showed no structure beyond a thickness attributable to noise and numerical error.

Recently another method has been introduced for discriminating between real and spurious exponents. Based on the idea that most of the deterministic dynamical systems of physical interest are invertible, we first perform the typical analysis on the given time series, and then we reverse the time series, relearn the reversed dynamics, and proceed to calculate exponents for this backward data.

Under time reversal, stable directions become unstable, and vice versa, no real exponents are expected to change sign. On the other hand, the spurious exponents, being artifacts of the geometry imposed by our *ad hoc* embedding, will not, in general, behave so nicely. In fact, experience seems to show that the artificial exponents vary strongly under time reversal, making identification easy.

H. Local exponents from known dynamics and from observations

Global Lyapunov exponents tell us how sensitive the dynamics is *on the average* over the attractor. A positive Lyapunov exponent assures us that orbits will be exponentially sensitive to initial conditions or to perturbations or to roundoff errors. No information is gained from them about the local behavior on an attractor. Just as the inhomogeneity of the attractor leads to interesting characterization of powers of the number density, so it is natural to ask about the local behavior of instabilities in various parts of state space.

A question of more substantial physical interest is this: if we make a perturbation to an orbit near the phase-space point $\mathbf{y}$, how does this perturbation grow or shrink in a finite number of time steps? This question is directly relevant to real predictability in the system, since the local Lyapunov exponents vary significantly over the attractor, so there may be some parts of phase space where prediction is much more possible than in other parts. For example, if we are investigating the ability to predict weather in a coupled ocean-atmospheric model, we are much more interested in making predictions six weeks ahead of a specified state-space location than 10^6 weeks (20 000 years) ahead. Global Lyapunov exponents may suffice for the latter, but local Lyapunov exponents are required for the former. Local Lyapunov exponents are defined directly from the Oseledec matrix (Abarbanel *et al.*, 1991),

$$\mathbf{OSL}(\mathbf{y},L)=\mathbf{DF}^L(\mathbf{y})\cdot[\mathbf{DF}^L(\mathbf{y})]^T , \tag{76}$$

which has eigenvalues that behave approximately as $\exp[2L\lambda_a(\mathbf{y},L)]$. The $\lambda_a(\mathbf{y},L)$ are the local Lyapunov exponents. From the multiplicative ergodic theorem of Oseledec we know

$$\lim_{L\to\infty} \lambda_a(\mathbf{y},L)=\lambda_a , \tag{77}$$

the global Lyapunov exponents. It is the variation in both phase-space location $\mathbf{y}$ and length of time L that is of interest.

First of all, the $\lambda_a(\mathbf{y},L)$ are strongly dependent on the particular orbit that we examine, just as are all functions that we evaluate on a chaotic orbit. So we must ask about averages over orbits with the invariant density. For this purpose we define an average local Lyapunov exponent $\bar{\lambda}_a(L)$,

$$\bar{\lambda}_a(L)= \int d^d y\, \rho(\mathbf{y})\lambda_a(\mathbf{y},L)$$

$$=\frac{1}{N} \sum_{k=1}^{N} \lambda_a(\mathbf{y}(k),L) . \tag{78}$$

For large L, since $\lambda_a(\mathbf{y},L)$ becomes independent of L, within the basin of attraction of the attractor defined by the observations we have

$$\lim_{L\to\infty} \bar{\lambda}_a(L)=\lambda_a . \tag{79}$$

The variations around the mean value are also of some importance, and we define the moments

$$[\sigma_a(p,L)]^p= \int d^d y\, \rho(\mathbf{y})[\lambda_a(\mathbf{y},L)-\bar{\lambda}_a(L)]^p$$

$$=\frac{1}{N} \sum_{k=1}^{N} [\lambda_a(\mathbf{y}(k),L)-\bar{\lambda}_a(L)]^p . \tag{80}$$

Two of the moments are trivial: $\sigma_a(0,L)=1$ and $\sigma_a(1,L)=0$. The moments for $p\geq 2$ tell us how the local exponents vary around the attractor, though they clearly average this information over the whole attractor. The direct local information resides in the $\lambda_a(\mathbf{y},L)$. Each of

the $\sigma_a(p,L)\to 0$ for $L\to\infty$ and $p\geq 2$, again by the Oseledec multiplicative ergodic theorem.

The behavior of the $\lambda_a(\mathbf{y},L)$ can also be captured in a probability density for values of each exponent, for a fixed L. The probability density for the ath exponent to lie in the interval $[u,u+du]$ is after L steps along the orbit,

$$P_a(u,L)=\frac{1}{N} \sum_{k=1}^{N} \delta(\lambda_a(\mathbf{y}(k),L)-u) , \tag{81}$$

and the moments above follow in the usual way.

The behavior of the $\bar{\lambda}_a(L)$ in model systems is quite striking. For the familiar test chaotic systems such as the Hénon map, the Ikeda map, the Lorenz attractor, and others, we have for large L

$$\bar{\lambda}_a(L)\approx\lambda_a+\frac{c_a}{L^{\nu_a}}+\frac{c_a'}{L}+\cdots , \tag{82}$$

where c_a and c_a' are constants, ν_a is a scaling exponent found numerically to be in the range $0.6\leq\nu_a\leq 0.85$ for many systems, and the final term proportional to L^{-1} comes from the lack of complete independence of the local Lyapunov exponents on the choice of coordinate system. A demonstration of this is found in Abarbanel *et al.* (1992)]. This also shows the independence of the global exponents from the coordinate system when $L\to\infty$.

All moments vanish for large L and do so at the rate

$$\sigma_a(p,L)\approx\frac{c_a''}{L^{\xi_a}}+\frac{c_a'''}{L} , \tag{83}$$

where c_a'' and c_a''' are constants, ξ_a is a scaling index in approximately the same numerical range as the ν_a, and the last term comes from the coordinate system dependence.
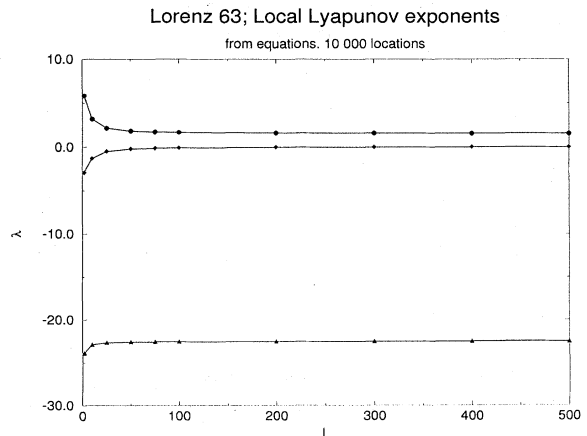
A typical example can be seen in Fig. 22, where the

FIG. 22. Average local Lyapunov exponents for Lorenz attractor calculated from the Lorenz equations (11) as a function of the composition length L. Averages are over 10 000 initial conditions.
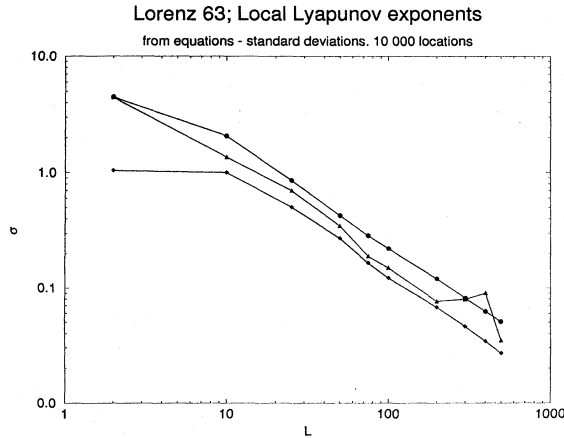
FIG. 23. Standard deviations from the average for local Lyapunov exponents calculated from the Lorenz equations (11) as a function of the composition length L. Averages are over 10 000 initial conditions.

three $\overline{\lambda}_a(L)$ for the Lorenz attractor are shown, and in Fig. 23, where the three $\sigma_a(2,L)$ for the same system are shown. We also present in Fig. 24 the distribution of exponents $P_1(u,L)$ for the Lorentz attractor for $L=2$ and $L=50$.

The scaling indices ν_a and ξ_a are characteristic of the dynamics and can be used, just as are the D_q or the K_q or the λ_a, for classifying the source of signals. It is important to evaluate them from data. The method is really a straightforward extension of the techniques outlined above for evaluating the global Lyapunov exponents from data—with one small exception.

As before we compute the recursive **QR** decomposition of

$$DF^L(\mathbf{y}) \cdot [DF^L(\mathbf{y})]^T = \mathbf{Q}_1(2L)\mathbf{R}_1(2L)\mathbf{R}_1(2L-1) \cdots \mathbf{R}_1(1).$$

$$(84)$$

If $\mathbf{Q}_1(2L)$ were the identity matrix, the eigenvalues of $OSL(\mathbf{y},L)$ would all lie in the "upper triangular" part of the **QR** decomposition. In general, $\mathbf{Q}_1(2L)$ is not the identity. We follow the construction in Stoer and Burlisch (1980) and make a similarity transformation to

$$\mathbf{Q}_1^T(2L)\mathbf{Q}_1(2L)\mathbf{R}_1(2L)\mathbf{R}_1(2L-1) \cdots \mathbf{R}_1(1)\mathbf{Q}_1(2L)$$

$$= \mathbf{R}_1(2L)\mathbf{R}_1(2L-1) \cdots \mathbf{R}_1(1)\mathbf{Q}_1(2L) . \quad (85)$$

Now we repeat the **QR** decomposition,

$$DF^L(\mathbf{y}) \cdot [DF^L(\mathbf{y})]^T$$

$$= \mathbf{Q}_2(2L)\mathbf{R}_2(2L)\mathbf{R}_2(2L-1) \cdots \mathbf{R}_2(1) , \quad (86)$$

which has the same eigenvalues as the original $OSL(\mathbf{y},L)$. We continue this K times to reach

$$\mathbf{Q}_K(2L)\mathbf{R}_K(2L)\mathbf{R}_K(2L-1) \cdots \mathbf{R}_K(1) , \quad (87)$$

which still has the same eigenvalues as $OSL(\mathbf{y},L)$. As K increases, $\mathbf{Q}_K(2L)$ converges rapidly to the identity matrix (Stoer and Burlisch, 1980).

When $\mathbf{Q}_K(2L)$ is the identity to desired accuracy, the final form of $OSL(\mathbf{y},L)$ is upper triangular to desired accuracy, and one can read off the local Lyapunov exponents, $\lambda_a(\mathbf{y},L)$, from the diagonal elements of the $\mathbf{R}_n(k)$'s:

$$\lambda_a(\mathbf{y},L) = \frac{1}{2L\tau_s} \sum_{j=1}^{2L} \log[R_K(j)_{aa}] , \quad (88)$$

since the eigenvalues of a product of upper triangular matrices are the product of the eigenvalues of the individual matrices. The rapid rate of convergence of $\mathbf{Q}_K(2L)$ to the identity means only a few steps, $K \approx 3$ or so, are
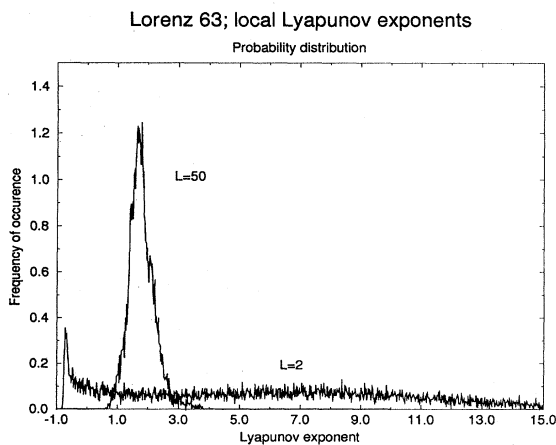

FIG. 24. The distribution of the largest Lyapunov exponent for the Lorenz model for composition lengths $L=2$ and $L=50$. Each distribution is normalized to unity. Note that when $L=2$, this *largest* exponent can be negative.
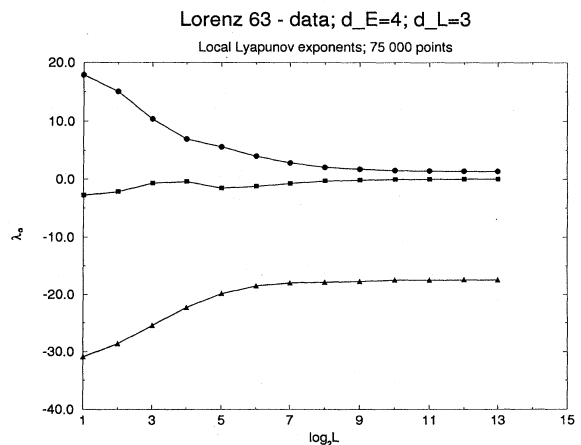

FIG. 25. Average local Lyapunov exponents calculated directly from the time series $x(t)$ generated by the Lorenz equations (11). 75 000 data points are used. The embedding dimension is $d_E=4$, and the order of local polynomials is taken to be $d_L=3$; 1500 initial conditions are used.
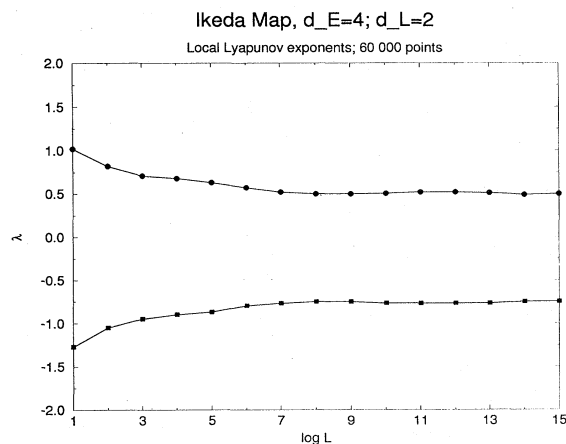
Ikeda Map, d_E=4; d_L=2
Local Lyapunov exponents; 60 000 points

FIG. 26. Average local Lyapunov exponents for data from the Ikeda map; 60 000 points are used. $d_E = 2$, $d_L = 2$; 1500 initial conditions.

ever needed to assure that $Q_K(2L)$ differs from the identity matrix by a part in 10^{-5} or 10^{-6}.

In Figs. 25 and 26 we show the local Lyapunov exponents determined from observations on the $x(n)$ data from the Lorenz equations and from $\text{Re}[z(n)]$ from the Ikeda map. In the former case we used $d_E = 4$ and used local polynomials of dimension $d_L = 3$ for the neighborhood-to-neighborhood maps, while in the latter, $d_E = 4$ and $d_L = 2$. We shall discuss below how we know $d_L = 2$ for the Ikeda map.

I. Topological invariants

An ideal strange attractor, which never has a coincidence between stable and unstable directions, has embedded within it a dense set of unstable periodic orbits (Devany and Nitecki, 1979; Auerbach *et al.*, 1987; Cvitanović, 1988; Melvin and Tufillaro, 1991). Such a *hyperbolic* attractor is rarely, if ever, found in physical settings, but the approximation of real strange attractors by this idealization can be quite rewarding. Indeed, in many examples of strange attractors, the Hénon and Ikeda maps used as examples throughout this article seem to possess numerically many of the properties of the "ideal" strange attractor. The topological properties of the linkings among these unstable periodic orbits characterizes classes of dynamics, and these topological classifications remain true even when the parameters of the vector field governing the dynamics are changed. So when one goes from stable periodic behavior to quasiperiodic behavior to chaos, the classification by integers or ratios of integers characteristic of topological invariants remains the same. This is in direct contrast to the various invariant properties we have discussed heretofore. It is clear that, if one were able to extract these topological invariants from experimental observations, they could be used to exclude incorrect dynamics and to point to classes of dynamics that could be further investigated, say, by computing dimensions or Lyapunov exponents. In that sense such geometric quantities can be a critical tool in selecting possible vector fields that could describe the observations.

If the dynamics can be embedded in R^3, then the periodic orbits are closed curves, which can be classified by the way they are knotted and linked with each other (Birman and Williams, 1983a, 1983b). This is the case that has been studied in detail in the literature (Mindlin *et al.*, 1990; Flepp *et al.*, 1991; Mindlin *et al.*, 1991; Tufillaro *et al.*, 1991; Papoff *et al.*, 1992) and applied to experimental data. Whether one will be able, in a practical way, to extend the existing literature to more than three dimensions (Mindlin *et al.*, 1991) remains to be seen, but the applications to physical systems that happen to fall into this category have already proven quite interesting.

The first task, clearly, is to identify within a chaotic time series the unstable periodic orbits. To do this one does not require an embedding (Mindlin *et al.*, 1991), though to proceed further with topological classification an embedding is required. Basically one seeks points in the time series, embedded or not, which come within a specified distance of one another after a fixed elapsed time—namely, the period (Auerbach *et al.*, 1987; Mindlin *et al.*, 1990) of the unstable orbit. To succeed in this one must be slightly lucky in that the Lyapunov number associated with this unstable orbit had better be close to unity or in one period the orbit will have so departed from the unstable periodic structure in state space that one will never be able to identify the unstable periodic orbit. If some orbits can be found by this procedure, then one can proceed with the classification.

Using time-delay embedding may not always work well for the purpose of classifying attractors by their topology, for in three dimensions there may be self intersections of the flow; i.e., $d_E = 3$ may be too low for time-delay embedding. If the local dimension of the dynamics is three, one may be able to follow the orbits in $d = 4$ or higher, but this has not yet been done. In much of the reported work, embeddings other than time-delay embeddings are utilized, though there is no systematic procedure available. One useful embedding is to take the scalar data $s(n)$ and form

$$y_1(n) = \sum_{k=1}^{n} [s(k)e^{-(n-k)\Delta}],$$

$$y_2(n) = s(n),$$

$$y_3(n) = s(n) - s(n-1).$$

(89)

This is a perfectly fine embedding, and in the study of experiments on chemical reactions (Mindlin *et al.*, 1991) and laser physics (Papoff *et al.*, 1992) it has served to unfold the unstable periodic orbits in three dimensions. One may not always be able to accomplish this, but when it is possible, tools are then available for further analysis.

In particular, by analyzing the linking of low-order

periodic orbits and the torsion—the angle (in units of π) through which the flow twists in one period in the neighborhood of an unstable periodic orbit—of such low-order orbits, one is able to establish a *template* or classification scheme which fully characterizes all unstable periodic orbits and other topological properties of the attractor (Mindlin *et al.*, 1991). Once the template has been established one may verify it by examining the linkage of higher-order unstable periodic orbits, if they are available. It is in the template that one finds possible deviation from the rigorous mathematics (Birman and Williams, 1983a, 1983b) available for hyperbolic strange attractors. In the template one may find periodic orbits without counterpart in the actual flow. The gain for the analysis of experimental data is such that this is a small price to pay. Indeed, as is often the case, precise mathematics not quite applicable to a physical setting may serve as a warning rather than forbidding attempted application.

It appears (Gilmore, 1992) that this procedure is rather robust against contamination, since noise first destroys one's ability to identify high-order unstable periodic orbits. These, happily, are not needed in establishing the template. In the presence of noise at the level of nearly 60% of the signal in one case (Mindlin *et al.*, 1991), the low-order unstable periodic orbits were still visible. If one could perform some sort of signal separation, even as crude as local averaging on the time series, higher-order noise levels could perhaps be handled.

This sequence of steps, from identifying the unstable periodic orbits within a strange attractor to creating a template, has been carried out in the analysis of a laser physics experiment (Papoff *et al.*, 1992) and a chemical reaction experiment (Mindlin *et al.*, 1991).

In the case of the NMR laser studied by Flepp *et al.* (1991), it was established that a term in addition to the conventional Bloch-Kirchhoff equations is required by the properties of the unstable periodic orbits in the data. This is a powerful result and underlines the importance of topological tools based on unstable periodic orbits when they are available. Indeed, we would say that if one is able to extend the existing analysis methodology to dimensions greater than three, this topological approach should always be among the first tools one applies to any chaotic data set. It will then serve as a critical filter for proposed model equations, regardless of whatever ability one has to evaluate invariants depending on distances such as dimensions and Lyapunov exponents. The utility of this approach and its robustness in the presence of contamination serves as another item in our list of techniques that rest on geometry and may prove increasingly important when real, noisy data are confronted. There is no substitute, eventually, for knowing Lyapunov exponents of a system if prediction or control is one's goal, but real data may require sharper instruments for its study before these dynamical quantities can be reliably extracted. The topology of the unstable periodic orbits may well provide that instrument.

VI. NONLINEAR MODEL BUILDING: PREDICTION IN CHAOS

The tasks of establishing the phase space for observations and of classifying the source of the measurements can be enough in themselves for many purposes. Moving on to the next step, *building models and predicting future behavior of the system,* is the problem that holds the most interest for physicists. In some sense the problem of prediction is both the simplest and the hardest problem among all those we discuss in this review. The basic idea is quite straightforward: since we have information on the temporal evolution of orbits $y(k)$ and these orbits lie on a compact attractor in phase space, each orbit has near it a whole neighborhood of points in phase space which also evolve under the dynamics to new points. We can combine this knowledge of the evolution of whole neighborhoods of phase space to enhance our ability to predict in time by building local or global maps with parameters **a**: $y \rightarrow F(y,a)$, which evolve each $y(k) \rightarrow y(k+1)$. Using the information about how neighbors evolve, we utilize phase-space information to construct the map, and then we can use the map either to extend the evolution of the last points in our observations forward in time or to interpolate any new phase-space point near the attractor forward (or backward, if the map is invertible) in time.

All that said, the problem reduces to determining the parameters **a** in the map, given a class of functional forms which we choose in one way or another. As discussed by Rissanen (1989), there is no algorithmic way to determine the functional form. This is actually good news or all physics would be reduced to fitting data in an algorithmic fashion without regard to the interpretation of the data.

Suppose, from guessing or some physical reasoning, we have chosen the functional form of our map. If we are working with local dynamics, then local polynomials or other natural basis functions would be a good choice, since any global form can be approximated arbitrarily well by the collection of such local maps, if the dynamics is smooth enough. Choosing the parameters now requires some form of cost function or metric which measures the quality of the fit to the data and establishes how well we do in matching $y(k+1)$ as observed with $F(y(k),a)$. Call this the local deterministic error, $\epsilon_D(k)$,

$$\epsilon_D(k) = y(k+1) - F(y(k),a) . \tag{90}$$

The cost function for this error we shall call $W(\epsilon)$. If the map $F(y,a)$ we are constructing is *local,* then for each neighbor of $y(k)$, $y^{(r)}(k) r = 1, 2, \ldots, N_B$,

$$\epsilon_D^{(r)}(k) = y(r, k+1) - F(y^{(r)}(k),a) .$$

$y(r, k+1)$ is the point in phase space to which the neighbor $y^{(r)}(k)$ evolves; it is not necessarily the rth nearest neighbor of $y(k+1)$. For a least-squares metric the local neighborhood-to-neighborhood cost function would be

$$W(\epsilon,k)=\sum_{r=1}^{N_B} |\epsilon_D^{(r)}(k)|^2 / \sum_{r=1}^{N_B} [\mathbf{y}(k)-\langle\mathbf{y}(k)\rangle]^2 , \quad (91)$$

and the parameters determined by minimizing $W(\epsilon,k)$ will depend on the "time" associated with $\mathbf{y}(k)$: $\mathbf{a}(k)$. The normalization is arbitrary, but the one shown uses the size of the attractor to scale the deterministic error.

If the fit is to be a global least-squares determination of

a, then

$$W(\epsilon)=\sum_{k=1}^{N} |\epsilon_D(k)|^2 / \sum_{r=1}^{N} [\mathbf{y}(k)-\langle\mathbf{y}(k)\rangle]^2 , \quad (92)$$

and so forth. An interesting option for modeling is to require the **a** to be selected by maximizing the average mutual information between $\mathbf{y}(r,k+1)$ over the neighborhood and the $\mathbf{F}(\mathbf{y}^{(r)}(k),\mathbf{a})$; i.e., maximize

$$I(\mathbf{y}(r,k+1),\mathbf{F}(\mathbf{y}^{(r)}(k),\mathbf{a}))=\sum_{r=1}^{N_B} P[\mathbf{y}(r,k+1),\mathbf{F}(\mathbf{y}^{(r)}(k),\mathbf{a})]\log_2\left[\frac{P[\mathbf{y}(r,k+1),\mathbf{F}(\mathbf{y}^{(r)}(k),\mathbf{a})]}{P[\mathbf{y}(r,k+1)]P[\mathbf{F}(\mathbf{y}^{(r)}(k),\mathbf{a})]}\right] , \quad (93)$$

with respect to the $\mathbf{a}(k)$ at each k.

With one exception, to be noted shortly, this general outline is followed by all of the authors who have written about the subject of creating maps and then predicting the evolution of new points in the state space of the observed system. The exception is a representation of the mapping function taking $\mathbf{x}\rightarrow\mathbf{F}(\mathbf{x})$ in an expansion in orthonormal functions $\phi_a(\mathbf{x})$

$$F_\alpha(\mathbf{x})=\sum_{a=1}^{M} c_\alpha(a)\phi_a(\mathbf{x}) , \quad (94)$$

where the functions are chosen to be orthonormal with respect to the observed invariant density of the data

$$\rho(\mathbf{x})=\frac{1}{N}\sum_{j=1}^{N} \delta^d(\mathbf{x}-\mathbf{y}(j)) , \quad (95)$$

as

$$\int d^d x\, \rho(\mathbf{x})\phi_a(\mathbf{x})\phi_b(\mathbf{x})=\delta_{ab} . \quad (96)$$

The form of $\rho(\mathbf{x})$ allows one to determine the expansion coefficients $c(a)$ without any fitting (Giona *et al.*, 1991; Brown, 1992), as we shall see below.

All other mapping functions, local or global, are of the form of Eq. (96) with a structure for the functions $\phi_a(\mathbf{x})$ that is preassigned and not dependent on the data from the observed system.

In what follows we give a short survey of the methods developed for the prediction in chaos. There are many different methods, and we shall follow the traditional classification of them as local and global models. By definition, local models vary from point to point (or from neighborhood to neighborhood) in the phase space. Global models are constructed once and for all in the whole phase space (at least, the part occupied by the data stream). It should be noted, however, that the dividing line is becoming harder and harder to define. Models using *radial basis functions* or *neutral nets* carry the features of both. They are usually used as global functional forms, but they clearly demonstrate localized behavior. In the next subsections we start with the classic local and global models in order to explain the main ideas and then discuss the newer techniques. First we take up the topic of choosing the appropriate dimension for our models.

A. The dimension of models

Before discussing model making, it is useful to consider how to choose the appropriate dimension of the models when we have observed only scalar data $s(n)$. When we reconstruct phase space using time delays, we, at best, arrive at a necessary dimension d_E for unfolding the system attractor. As the example of the Ikeda map shows quite clearly, d_E may be larger than the local dimension of the dynamics, d_L. For the Ikeda map, $d_E=4$, while the local dimension, that is, the dimension of the actual dynamics, is $d_L=2$. Recall that this difference comes from the choice of coordinate system and has nothing to do with the dynamical content of the observations. While we could, in the Ikeda map again, make four-dimensional models without any special theoretical penalty, making four-dimensional local models would require unnecessary work and unwanted susceptibility to "noise" or other contamination of various dimensions higher than d_E.

A method we can use for choosing the dimension of models is to examine the local average Lyapunov exponents $\bar{\lambda}_a(L)$. If we compute these exponents in dimension d_E, we, of course, have d_E of them. d_E-d_L are totally geometric and have nothing to do with the dynamics required to describe the observations. If we compute the $\bar{\lambda}_a(L)$, reading the d_E dimensional data forward in time, and then do precisely the same, reading the data backward in time, the true exponents of the system will change sign (Eckmann and Ruelle, 1985; Bryant *et al.*, 1990; Brown *et al.*, 1991; Parlitz, 1992; Abarbanel and Sushchik, 1993). The false d_E-d_L exponents will not change sign and will be clearly identified. If one performs this calculation keeping d_E fixed, so distances between points on the attractor are properly evaluated, but changing the dimension of the local model of the neighborhood-to-neighborhood dynamics from $d_L=d_E$ to $d_L=d_E-1$ to $d_L=d_E-2$, etc. until there are no remaining false local Lyapunov exponents, then the correct value of d_L will be that for which the false exponents are first absent. At that local dimension, if it is desired, one may return to the computation of the local exponents using more data and higher-accuracy data to evaluate the local Lyapunov exponents with as much pre-
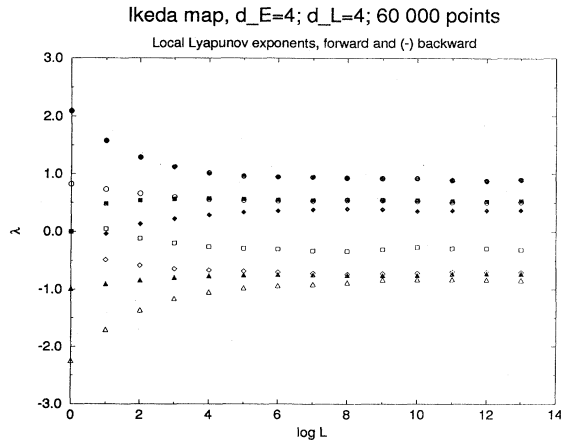
Ikeda map, d_E=4; d_L=4; 60 000 points

Local Lyapunov exponents, forward and (-) backward

FIG. 27. Local Lyapunov exponents calculated from data generated by the Ikeda map, $d_E = 4$, $d_L = 4$. Forward exponents are marked by open symbols, minus backward exponents by solid symbols. Clearly, there are two true exponents and two spurious ones.

cision as desired. The advantage to using *local* Lyapunov exponents is simply that one has more points of direct comparison between forward exponents and negative backward exponents than just the comparison of global exponents. Further, if the data come from a differential equation, local exponents are able to identify true and false versions of the zero exponent, which as a global exponent will not change sign under time reversal.

In Fig. 27 are shown the average local Lyapunov exponents for the Ikeda map with $d_E = 4$, which is larger than $d_L = 2$ but is the dimension suggested by the global false-nearest-neighbor algorithm. Forward and minus backward $\bar{\lambda}_a(L)$ are displayed. Evidently only two exponents satisfy the condition of equality of forward and minus backward average local Lyapunov exponents. Decreasing d_L from four to three and then to two, we observe (but do not display) the same phenomenon: only two exponents satisfy the required time-reversal condition.

B. Local modeling

We now assume that our data are embedded in an appropriate phase space, and we have determined the dimension of the model. The problem now is to reconstruct the deterministic rule underlying the data.

We start our discussion with the simplest and earliest nonlinear method of local forecasting, which was suggested by E. Lorenz (1969). Let us propose to predict the value of $\mathbf{y}(k+1)$ knowing a long time series of $\mathbf{y}(j)$ for $j \leq k$. In the *"method of analogs"* we find the nearest neighbor to the current value of $\mathbf{y}(k)$ say, $\mathbf{y}(m)$ and then assume that the $\mathbf{y}(m+1)$ is the predicted value for $\mathbf{y}(k+1)$. This is pure persistence, and is not much of a model. In the classification by Farmer and Sidorowich (1988) it is called a *first-order approximation*. Clearly, the

quality of this prediction can be improved in different ways. The first thing one can do is to take a collection of near neighbors of the point $\mathbf{y}(k)$ and take an averaged value of their images (Pikovsky, 1986) as the prediction. Further improvement of the same idea is to make a least-squares fit of a local linear map of $\mathbf{y}(k)$ into $\mathbf{y}(k+1)$ using $N_B > d_L$ near neighbors for a given metric $\|\cdot\|$,

$$W = \sum_{r=1}^{N_B} \|\mathbf{y}(r,k+1) - \mathbf{a}^{(r)}\mathbf{y}^{(r)}(k) - \mathbf{b}^{(r)}\|^2 . \tag{97}$$

This represents a *second-order approximation*. The sum in Eq. (97) can be weighted in order to provide a larger contribution from close points. More sophisticated local prediction schemes (Farmer and Sidorowich, 1987) may represent the map in the form of a higher-order polynomial whose coefficients must be fit using near neighbors. In this case one can expect better results for more long-term forecasts. However, the use of high-order polynomials requires many free parameters to be determined, namely, $(m + d_L)!/(m!d_L!) \approx d_L^m$, where d_L is the local dimension and m is the degree of the polynomial. It makes these models less practical for local forecasting if the embedding dimension and/or the order of the polynomial are too high.

If one wishes to predict the value of $y(k+K)$, i.e., K steps ahead, one can use a *direct method*, which finds the coefficients of the polynomial map $\mathbf{F}^K(\mathbf{y})$ by direct minimization of

$$W^{(K)} \equiv \sum_{r=1}^{N_B} \|\mathbf{y}(r,K+k) - \mathbf{F}^K(\mathbf{y}^{(r)}(k))\|^2 . \tag{98}$$

One particularly useful property of local polynomial forecasts is that we can estimate how the quality of the forecasts scales as a function of prediction interval K, size of training set N, dimension of attractor d_A, and largest Lyapunov exponent λ_1.

Assume we accumulate a data set $y_n = y(x_n)$ from observing a continuously evolving system. In other words, we know y varies continuously, but we can only make measurements at the x_n. We might try to model approximately the intervening values of the system variable y by interpolating a fit to the measured data. Let $F(x_n)$ denote the true dynamics driving the system's evolution, and consider its Taylor expansion:

$$F(x_n + \delta) = F(x_n) + DF(x_n)\delta + \frac{1}{2}D^2F(x_n)\delta^2$$
$$+ \frac{1}{6}D^3F(x_n)\delta^3 + \cdots . \tag{99}$$

If we construct a local polynomial model of degree m, our leading-order expectation for the forecast error is the δ^{m+1} term in Eq. (99). Recall one of our basic definitions of dimension, and rearrange it to express the distance variable as

$$L = V^{1/D} . \tag{100}$$

If we make the admittedly faulty assumption that the points are evenly distributed through the state space,

then $V = \rho N$, which implies

$$\frac{L}{\rho^{1/D}} = N^{1/D} \approx \delta \ . \tag{101}$$

This expression for the average distance between neighboring points can be used to replace δ in Eq. (99).

Next there are the derivative factors in each term of Eq. (99). As we have seen in Sec. V, the Lyapunov exponents can be viewed as measuring the average derivative of the dynamics over the attractor. Discrepancies between our forecasts and the actual points on the evolving trajectory would grow much in the same way that initially nearby points would separate under the dynamics, and this expansion rate is measured by the largest Lyapunov exponent:

$$\langle DF(x) \rangle \propto \exp(\lambda_1 K) \ . \tag{102}$$

Although nontrivial to justify (Farmer and Sidorowich, 1988), the assumption that higher-order derivatives also average in a similar manner,

$$\langle D^n F(x) \rangle \propto \exp(n\lambda_1 K) \ , \tag{103}$$

provides a useful model that agrees well with numerical results from ideal calculations:

$$\langle E_{\mathrm{rms}} \rangle \propto N^{-(m+1)/d_L} \exp[(m+1)\lambda_1 K] \ . \tag{104}$$

Having a limited number of data points, we must balance two opposing factors. Higher-order polynomials potentially promise higher accuracy, but also require more points in the neighborhood (and, therefore, larger neighborhoods). That in turn makes the local mapping more complex.

Frequently it is desirable to attempt long-term forecasts. One straightforward approach is to use the already described approximation techniques to construct a predictive model, $F_K(x)$, specifically for the long-term interval:

$$F_K(x_n) \approx x(n+K) \ . \tag{105}$$

In this case we construct a model that predicts over the full time interval in a single step. Any implementation of the supervised learning algorithm would select a sample of domain points from the neighborhood of the point of interest, and the range sample would be all those neighbors' post-images the full time K later.

An alternative approach is to split the interval K into n subintervals of length $K_{\mathrm{step}} = K/n$ and concatenate n short-term forecasts to build up the full prediction. The original domain sample remains the same, but the range sample is now found, on average, to be much more closely packed, since the prediction interval is much shorter. The initial forecast produces a new state, which is used as the basis for selecting a new domain sample, which in turn supplies a new range sample for the next predictive model fit, and so on.

This iterative, or recursive, forecasting technique can demonstrate substantial advantages over the direct

method. For an appropriate range of base forecasting time steps, K_{step}, the prediction error scales differently,

$$E \propto N^{-(m+1)/d_L} \exp[\lambda_1 n K_{\mathrm{step}}] \ , \tag{106}$$

and implies higher accuracy due to the absence of the factor $(m+1)$ in the second exponent.

Experience indicates that the scaling behavior of forecasting errors strongly depends on some of the modeling parameters. If the time series is amenable to the sort of analysis we are prescribing, there will be a range of values for K_{step} where Eqs. (104) and (106) hold. This range typically centers on the optimal value for the time delay used in phase-space reconstruction (Sec. IV).

If data are collected by sampling a continuous signal at a very high rate, forecasts can be made over intervals shorter than the delay time used in phase-space reconstruction, with performance limited by the signal-to-noise ratio of the data. However, using the shortest possible composition time in an iterative forecasting scheme is not a wise strategy, since the large number of forecasts that need to be concatenated compound the errors at each step and can cause the recursive technique to perform worse than the direct methods.

Different local prediction methods were tested by Farmer and Sidorowich (1988) and Casdagli (1989) for a variety of low-dimensional chaotic systems. In Fig. 28 we show the comparison of different orders of direct and iterative approximations for the Ikeda map. As expected, for first-order (constant) approximations there is no difference between direct and iterative scalings. For linear and quadratic predictors the iterative procedure is clearly superior, in accordance with the scaling expectations. Figure 29 illustrates the difference between direct and iterated forecast performance of local linear models for hysteretic circuit data of Pecora and Carroll.
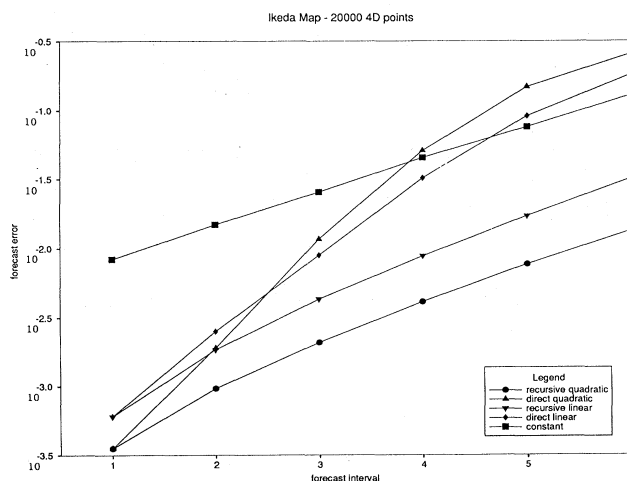

FIG. 28. Prediction error as a function of prediction time for the Ikeda map for different prediction techniques.
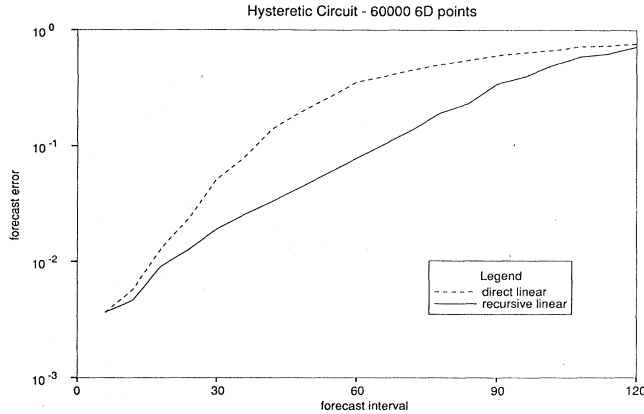
FIG. 29. Direct and iterative linear forecast error for hysteretic circuit data.

C. Global modeling

The collection of local polynomial (or other basis function) maps form some global model. The shortcomings of such a global model are its discontinuities and extremely large number of adjustable parameters. The latter is clearly a penalty for high accuracy. At the same time, it would be nice to have a relatively simple continuous model describing the whole collection of data. A number of solely global models has been invented which present a closed functional representation of the dynamics in the whole phase space (or, at least, on the whole attractor).

Each method uses some expansion of the dynamical vector field $F(x)$ in a set of basis functions in R^d. The first such global method that comes into mind is to use polynomials again. Their advantage in local modeling where least-squares fitting works well, is now reduced by the extremely large number of data points and the need to use rather high-order polynomials. There is an attractive approach to finding a polynomial representation of a global map. This *measure-based functional reconstruction* (Giona *et al.*, 1991; Brown, 1992) uses orthogonal polynomials whose weights are determined by the invariant density on the attractor.

The method eliminates the problem of multiparameter optimization. Finding the coefficients of the polynomials and the coefficients of the function $F(y)$ requires only the computation of moments of data points in phase space. It works as follows:

We introduce polynomials $\phi_a(x)$ on R^d which are orthogonal with respect to the natural invariant density on the attractor,

$$\int d^dx \rho(x)\phi_a(x)\phi_b(x)=\delta_{ab} , \quad (107)$$

and which are determined by a conventional Gram-Schmidt procedure starting from

$$\phi_1(x)=1 . \quad (108)$$

The vector field $F(x)$, which evolves data points $y(k+1)=F(y(k))$, is approximated in Mth order as

$$F_M(x)= \sum_{a=1}^{M} c(a)\phi_a(x) , \quad (109)$$

and the coefficients $c(a)$ are determined via

$$\begin{aligned} c(a) &= \int d^dx\ F(x)\phi_a(x)\rho(x) \\ &= \frac{1}{N}\sum_{k=1}^{N} F(y(k))\phi_a(y(k)) \\ &= \frac{1}{N}\sum_{k=1}^{N} y(k+1)\phi_a(y(k)) . \end{aligned} \quad (110)$$

This demonstrates the power of the method directly. Once we have determined the orthogonal polynomials $\phi_a(x)$ from the data, which we can do before computations are initiated, the evaluation of the vector field at any order we wish is quite easy: only sums over powers of the data with themselves are required.

Furthermore, the form of the sums involved allows one to establish the vector field from a given set of data and adaptively improve it as new data are measured. The best aspect of the method, however, may be robustness against contamination of the data (Brown, 1992). There is no least-squares parameter search involved, so no distances in state space need be evaluated. The geometric nature of the method does not rely on accurately determining distances and is thus not so sensitive to "noise" which spoils such distance evaluations.

It is possible that the two general forms of global modeling we have touched on can be further improved by working with rational polynomials rather than simple Taylor series. Rational polynomials have a greater radius of convergence, and thus one may expect them to be more useful when the data set is small and perhaps even allow the extrapolation of the model from the immediate vicinity of the measured attractor out into other regions of phase space.

D. In between local and global modeling

As we already pointed out, the present direction of nonlinear modeling combines features of local and global models. Consider, for example, the method of *radial basis functions* (Powell, 1981, 1985, 1987), which, as Casdagli (1989) notes, "is a global interpolation technique with good localization properties." In this method a local predictor $F(y)$ is sought in the form

$$F(y)= \sum_{n=1}^{N_c} \lambda_n \Phi(\|y-y_n\|) , \quad (111)$$

where Φ is some smooth function in R^1 and $\|\cdot\|$ is the Euclidean norm. The coefficients λ_n are chosen to satisfy, as usual,

$$F(y(k))=y(k+1) . \quad (112)$$

Depending on the number of points N_c used for reconstruction, this method can be considered as local (small $N_c \ll N$) or global ($N_c \approx N$).

One can choose different $\Phi(\|\mathbf{x}\|)$ as the radial basis function. For example, $\Phi(r) = (r^2 + c^2)^{-\beta}$ works well for $\beta > -1$ and $\beta \neq 0$. Moreover, if one adds a sum of polynomials to the sum of radial basis functions, then even increasing functions $\Phi(r)$ provide good localization properties. However, for a large number of points N_c this method is computationally as expensive as the usual least-squares fit. Numerical tests carried out by Casdagli (1989) show that for a small number of data points radial basis predictors do a better job than polynomial models, whereas for a larger quantity of data ($N \geq 10^4$) the local polynomial models seem to be superior.

A modification of the radial-basis-function method is the so-called kernel density estimation (Silverman, 1986). This method allows one to estimate a smooth probability distribution from discrete data points. Each point is associated with its *kernel* [a smooth function $K(\|\mathbf{y} - \mathbf{y}_i\|)$], which typically decays with distance but sometimes can even increase. Then one can compute a probability distribution

$$p(\mathbf{y}) = \sum_i K(\|\mathbf{y} - \mathbf{y}_i\|)$$

or a conditional probability distribution

$$p_c(\mathbf{y}|\mathbf{z}) = \sum_i K(\|\mathbf{y} - \mathbf{y}_{i+1}\|) K(\|\mathbf{z} - \mathbf{y}_i\|) \ .$$

Kernel density estimation can then be used for conditional forecasting by the rule

$$\bar{\mathbf{y}}(k+1) = \int d\mathbf{x}\, p_c(\mathbf{x}|\mathbf{y}(k))\mathbf{x} \ . \tag{113}$$

Kernel density estimation usually provides the same accuracy as the first-order local predictors. It has the advantage of being quite stable even with noisy data (Casdagli *et al.*, 1991).

Again, computing the conditional probability distribution, one can impose weights in order to attach more value to the points close to the starting point of prediction (both in time and in phase space). In fact, this leads to a class of models that are hybrids of local and global ones. Moreover, it allows one to construct a model that not only possesses good predicting properties but also preserves important invariants of the dynamics. The prediction model by Abarbanel, Brown, and Kadtke (1989) belongs to this class. It parametrizes the mapping in the form

$$\mathbf{F}(\mathbf{y}, \mathbf{a}) = \sum_{k=1}^{N-1} \mathbf{y}(k+1) g(\mathbf{y}, \mathbf{y}(k); \mathbf{a}) \ , \tag{114}$$

where the function $g(\mathbf{y}, \mathbf{y}(k); \mathbf{a})$ is the analog of the kernel function, that is, it is near 1 for $\mathbf{y} = \mathbf{y}(k)$ and vanishes rapidly for nonzero $\|\mathbf{y} - \mathbf{y}(k)\|$. $\mathbf{a}$ is a set of parameters. The cost function to be minimized is chosen as

$$C(\mathbf{X}, \mathbf{a}) = \frac{\sum_{k=L}^{N-1} \|\mathbf{y}(k+1) - \sum_{n=1}^{L} X_n \mathbf{F}^n(\mathbf{y}(k-n+1), \mathbf{a})\|^2}{\sum_{k=1}^{N} \|\mathbf{y}(k)\|^2} , \tag{115}$$

where X_n, $n = 1, \ldots, L$ is a set of weights attached to the sequence of points $\mathbf{y}(k - n + 1)$ *all of which are to be mapped* into $\mathbf{y}(k+1)$ by maps $\mathbf{F}^n(\mathbf{y})$. It is clear that then $\mathbf{F}(\mathbf{y}(k), \mathbf{a})$ will be close to $\mathbf{y}(k+1)$, as it provides an excellent interpolation function in phase space and in time. The free parameters $\mathbf{a}$ in the kernel function $\mathbf{g}$ and the specific choice of the cost function allow one to predict forward accurately a given number $L > 1$ of time steps and satisfy additional constraints imposed by the dynamics—significant Lyapunov exponents and moments of the invariant density distribution, as determined by the data, will be reproduced in this model. The preservation of Lyapunov exponents does not follow automatically from the accuracy of the prediction. Indeed, it is shown by Abarbanel *et al.* (1989) that models able to predict with great accuracy can have all negative Lyapunov exponents, even when the data are chaotic.

At first sight very different ideas are exploited in *neural networks* when they are used as nonlinear models for prediction (Cowan and Sharp, 1988). Neural networks are a kind of model utilizing interconnected elementary units (neurons) with a specific architecture. After a *training* (or *learning*) procedure to establish the interconnections among the units, we have just a nonlinear model of our usual sort. In fact, this is just a nonlinear functional model composed of sums of sigmoid functions (Lapedes and Farber, 1987) instead of sums of radial basis functions or polynomials as in the previous methods. There are many different modifications of neural nets. Consider, for example, a typical variant of *feed-forward nets*. A standard one includes four levels of units (Fig. 30): input units, output units, and two hidden levels. Each unit collects data from a previous level or levels in the form of a weighted sum, transforms this sum, and produces input values for the next layer. Thus each neuron performs the operation

$$X_i^{\text{out}} = g\left[\sum_j T_{ij} X_j^{\text{in}} + \theta_i\right] \tag{116}$$

where T_{ij} are weights, θ_i are thresholds, and $g(x)$ is a nonlinear function typically of a sigmoidal form such as

$$g(x) = \tfrac{1}{2}[1 + \tanh(x)] \ . \tag{117}$$

For the purpose of forecasting in d_L-dimensional phase space, the input level should consist of d_L neurons, and we feed them by the d_L delayed values of the time series $\{\mathbf{y}(k), \mathbf{y}(k-1), \ldots, \mathbf{y}(k - d_L + 1)\}$. The predicted value $\mathbf{y}(k+p)$ appears at the output level after nonlinear transformation at the two hidden levels:
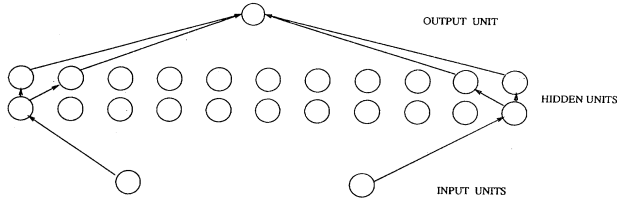
FIG. 30. Sketch of a feed-forward neural network.

$$\mathbf{y}(k+p)$$
$$=\mathbf{F}(\mathbf{y}(k),\mathbf{y}(k-1),\ldots,\mathbf{y}(k-d_L+1),\{T_{ij}\},\{\theta_i\})\ , \tag{118}$$

and the learning procedure consist of minimizing the error,

$$E=\sum_{p=1}^{N_s}\|\mathbf{y}_{\text{data}}(k+p)-\mathbf{y}(k+p)\|\ , \tag{119}$$

where N_s is the number of data in a training set. Depending on what sets of data are used for the training, the neural net eventually can be considered as a local or as a global model. In the first case only $N_c \ll N$ near neighbors are used for feeding the net, while in the second one uses all the data. The fitted parameters are here the weights $\{T_{ij}\}$ and the thresholds $\{\theta_i\}$. Unfortunately, this learning appears to be a kind of nonlinear least-squares problem, which takes much longer than linear fitting to run on regular computers. Nevertheless, neural networks present great opportunities for parallel implementation, and in the future that may become the most useful technique for forecasting.

Neutral nets are quite useful when we have a simple dynamics, and we can hope that a net with an *a priori* architecture will work well. Then the only thing to do is to train this net, i.e., to adjust the weights of connections and thresholds. Adjustment of the architecture itself is a secondary (and very often omitted) procedure for standard neural network modeling. The approach taken by *genetic learning algorithms* (Goldberg, 1989) represent a quite opposite point of view: This starts with finding an architecture based on the data set. Sorting out different architectures for a complex dynamics in a high-dimensional phase space is a very difficult task. To accomplish this, genetic algorithms utilize Darwinian evolution expressed in terms of variables (data), mathematical conditions (architecture), and quality of fit (criterion for sorting out the better architecture). Specifically, if we have a set of pairs $(\mathbf{x}_p,\mathbf{y}_p)$ where each $\mathbf{x}_p$ is mapped into $\mathbf{y}_p$ in some d_L-dimensional space, then we try to learn the set of conditions (*genes*) imposed on $\{\mathbf{x}_p\}$ under which the *quality of fit* associated with $\{\mathbf{y}_p\}$ is extremal. In particular, we can seek conditions determining the subspaces of the whole phase space for which the distribution of $\{\mathbf{y}_p\}$ is sharpest (Packard, 1990). The search for the appropriate set of conditions is performed by a kind of natural selection. We start with a number of random sets

(*genomes*) of conditions (*genes*), and after each generation some fraction of the worst sets of *genes* is discarded and some mutations are produced on the rest of the conditions. The important feature of genetic algorithms is that they allow an effective search in a very high-dimensional phase space. Conditions learned after many generations determine the part of the phase space (or orbit of a dynamical system) for which prediction is possible. In this regard we can see genetic algorithms as a global, adaptive search method, which allows searches in both parameter space and a predetermined, possibly large space of functional representations of the dynamics $\mathbf{x}\to\mathbf{F}(\mathbf{x})$. While there is really only limited development of this approach to nonlinear forecasting at this time, it clearly holds promise as an effective procedure for high-dimensional and complex systems.

VII. SIGNAL SEPARATION—"NOISE" REDUCTION

Now we address some of the problems connected with and some of the solutions to the very first task indicated in Table I, namely, *given observations contaminated by other sources, how do we clean up the signal of interest so we can perform an analysis for Lyapunov exponents, dimensions, model building, etc.?* In linear analysis the problem concerns the extraction of sharp, narrowband linear signals from broadband "noise." This is best done in the Fourier domain, but that is not the working space of the nonlinear analyst. Instead signals need to be separated from one another directly in the time domain. To succeed in signal separation, we need to characterize one or all of the superposed signals in some fashion that allows us to differentiate them. This, of course, is what we do in linear problems as well. If the observed signal $s(n)$ is a sum of the signal we want, call it $s_1(n)$, and other signals $s_2(n),s_3(n)\ldots$,

$$s(n)=s_1(n)+s_2(n)+\cdots\ , \tag{120}$$

then we must identify some distinguishing characteristic of $s_1(n)$ which either the individual $s_i(n)$, $i>1$, do not possess or perhaps the sum does not possess.

The most natural of these distinguishing properties is that $s_1(n)$ or its reconstructed phase-space version,

$$\mathbf{y}_1(n)=[s_1(n),s_1(n+T_1),\ldots,s_1(n+T_1(d_1-1))]\ , \tag{121}$$

satisfies a dynamical rule: $\mathbf{y}_1(k+1)=\mathbf{F}_1(\mathbf{y}_1(k))$ different from any dynamical rules associated with $s_2(n),\ldots$.

There are three cases we can identify:

- *We know the dynamics* $\mathbf{y}_1\to\mathbf{F}_1(\mathbf{y}_1)$.
- *We have observed some clean signal* $\mathbf{y}_R(k)$ *from the chaotic system, and we can use the statistics of the chaotic signal to distinguish it.*
- *We know nothing about a clean signal from the dynamics or about the dynamics itself. This is the "blind" case.*

It is quite helpful as we discuss work on each of these cases to recall what it is we are attempting to achieve by the signal separation. Just for discussion, let us take the observation to be the sum of two signals: $s(n)=s_1(n)+s_2(n)$. If we are in the first case indicated above, we want to know $s_1(n)$ with full knowledge of $\mathbf{y}_1 \rightarrow \mathbf{F}_1(\mathbf{y}_1)$. The information we use about the signal $s_1(n)$ is that it satisfies this dynamics in the reconstructed phase space, while $s_2(n)$ and its reconstruction does not. This is not really enough, since any initial condition $\mathbf{y}_1(1)$ iterated through the dynamics satisfies the dynamics by definition. However, it will have nothing, in detail, to do with the desired sequence $s_1(1),s_1(2), \ldots$ which enters the observations. This is because of the intrinsic instabilities in chaotic systems, so two different initial conditions diverge exponentially rapidly from each other as they move along the attractor. To the dynamics, then, we must add some cost function or quality function or accuracy function, which tells us that we are extracting from the observations $s(n)$ the *particular* orbit that lies "closest" to the particular sequence $s_1(1),s_1(2), \ldots$ which was observed in contaminated form. If we are not interested in that particular sequence but only in properties of the dynamics, we need do nothing with the observations, since we have the dynamics to begin with and can iterate any initial condition to find out what we want.

Similarly in the second case, we shall be interested in the particular sequence that was contaminated or masked during observation. In this case, too, since we have a clean reference orbit of the system, any general or statistical question about the system is best answered by that reference orbit rather than by attempting to extract from contaminated data some less useful approximate orbit.

In the third instance, we are trying to learn both the dynamics and the signal at the same time and with little *a priori* knowledge. Here it may be enough to learn the dynamics by unraveling a deterministic part of the observation in whatever embedding dimension we work in. That is, suppose the observations are composed of a chaotic signal $s_1(k)$, which we can capture in a five-dimensional embedding space combined with a two-hundred-dimensional contamination $s_2(k)$. If we work in five dimensions, then the second component will look nondeterministic, since with respect to that signal there will be so many false neighbors that the direction in which the orbit moves at essentially all points of the low-dimensional phase space will look random. Thus the distinguishing characteristic of the first signal will be its relatively high degree of determinism compared to the second.

A. Knowing the dynamics: manifold decomposition

In this section we assume that the dynamics of a signal are given to us: $\mathbf{y}_2 \rightarrow \mathbf{F}_1(\mathbf{y}_1)$ in d_1-dimensional space. We either observe a contaminated d_1-dimensional signal or, from a scalar measurement $s(k)=s_1(k)+s_2(k)$, we reconstruct the d_1-dimensional space. We call the obser-

vations $\mathbf{y}_O(k)$ in d_1 dimensions, and we wish to extract from those measurements the best estimate $\mathbf{y}_E(k)$ of the particular sequence $\mathbf{y}_1(k)$ which entered the observations. The error we make in this estimation $\mathbf{y}_1(k)-\mathbf{y}_E(k)=\epsilon_A(k)$ we call the *absolute error*. It is some function of ϵ_A that we wish to minimize. We know the dynamics, so we can work with the deterministic error $\epsilon_D(k)=\mathbf{y}_E(k+1)-\mathbf{F}_1(\mathbf{y}_E(k))$, which is the amount by which our estimate fails to satisfy the dynamics.

A natural starting point is to ask that the square of the absolute error $|\epsilon_A(k)|^2$ be minimized subject to $\epsilon_D(k)=0$; that is, we seek to be close to the true orbit $\mathbf{y}_1(k)$ constraining corrections to our observations by the requirement that all estimated orbits satisfy the known dynamics as accurately as possible. Using Lagrange multipliers $\mathbf{z}(k)$, this means we want to minimize

$$\frac{1}{2}\sum_{k=1}^{N}|\mathbf{y}_1(k)-\mathbf{y}_E(k)|^2$$
$$+\sum_{k=1}^{N-1}\mathbf{z}(k)\cdot[\mathbf{y}_E(k+1)-\mathbf{F}_1(\mathbf{y}_E(k))] \qquad (122)$$

with respect to the $\mathbf{y}_E(m)$ and the $\mathbf{z}(m)$.

The variational problem established by this is

$$\mathbf{y}_E(k+1)=\mathbf{F}_1(\mathbf{y}_E(k)) \ ,$$
$$\mathbf{y}_1(k)-\mathbf{y}_E(k)=\mathbf{z}(k-1)-\mathbf{z}(k)\cdot\mathbf{DF}_1(\mathbf{y}_E(k)) \ , \qquad (123)$$

which we must solve for the Lagrange multipliers and the estimated orbit $\mathbf{y}_E(k)$. The Lagrange multipliers are not of special interest, and the critical issue is estimating the original orbit. If we wished, we could attempt to solve the entire problem by linearizing in the neighborhood of the observed orbit $\mathbf{y}_O(k)$ and using some singular-value method for diagonalizing the $N \times N$ matrix, where N is the number of observation points (Farmer and Sidorowich, 1991). Instead we seek to satisfy the dynamics recursively by defining a sequence of estimates $\mathbf{y}_E(k,p)$ by

$$\mathbf{y}_E(k,p+1)=\mathbf{y}_E(k,p)+\Delta(k,p) \ , \qquad (124)$$

where at the zeroth step we shall choose $\mathbf{y}_E(k,0)=\mathbf{y}_O(k)$, since that is all we know from observations. We shall try to let the increments $\Delta(k,p)$ always be "small," so we can work to first order in them. If the $\Delta(k,p)$ start small and remain small, we can approximate the solution to the dynamics as follows:

$$\Delta(k+1,p)+\mathbf{y}_E(k+1,p)=\mathbf{F}_1(\mathbf{y}_E(k,p)+\Delta(k,p)) \ , \qquad (125)$$

$$\Delta(k+1,p)=\mathbf{F}_1(\mathbf{y}_E(k,p)+\Delta(k,p))-\mathbf{y}_E(k+1,p)$$
$$\approx \mathbf{DF}_1(\mathbf{y}_E(k,p))\cdot\Delta(k,p)+\epsilon_D(k,p) \ ,$$

where $\epsilon_D(k,p)=\mathbf{F}_1(\mathbf{y}_E(k,p))-\mathbf{y}_E(k+1,p)$ is the deterministic error, namely, the amount by which the estimate at the pth step fails to satisfy the dynamics.

We satisfy this linear mapping for $\Delta(k,p)$ by using the known value for $\mathbf{y}_E(k,p)$. To find the first correction to the starting estimate $\mathbf{y}_E(k,0)=\mathbf{y}_O(k)$, we iterate the mapping for $\Delta(k,0)$ and then add the result to $\mathbf{y}_E(k,0)$ to arrive at $\mathbf{y}_E(k,1)=\mathbf{y}_E(k,0)+\Delta(k,0)$. The formal solution for $\Delta(k,p)$ can be written down in terms of $\Delta(1,p)$. This approach suffers from numerical instability lying in the positive eigenvalues of the composition of the Jacobian matrices $\mathbf{DF}_1(\mathbf{y}_E(k,p))$. These, of course, are the same instabilities that lead to positive Lyapunov exponents and chaos itself.

To deal with these numerical instabilities, we introduce an observation due to Hammel (1990): since we know the map $\mathbf{x}\rightarrow\mathbf{F}_1(\mathbf{x})$, we can identify the linear stable and unstable invariant manifolds throughout the phase space occupied by the attractor. These linear manifolds lie along the eigendirections of the composition of Jacobian matrices entering the recursive determination of an estimated orbit $\mathbf{y}_E(k)$. If we decompose the linear problem Eq. (125) along the linear stable and unstable manifolds and then iterate the resulting maps forward along the stable directions and backward along the unstable directions, we shall have a numerically stable algorithm in the sense that a small $\Delta(1,p)$ along a stable direction will remain small as we move forward in time. Similarly, a small $\Delta(N,p)$ at the final point will remain small as we iterate backward in time.

Each iteration of the linear map for the $\Delta(k,p)$ will fail to satisfy the condition $\Delta(k,p)=0$, which would be the completely deterministic orbit, because the movement toward $\Delta(k,p)=0$ exponentially rapidly along both stable and unstable directions is bumped about by the deterministic error at each stage. If that deterministic error grows too large, then we can expect the linearized attempt to find the deterministic orbit closest to the observations to fail. This deterministic error is a measure at each iteration of the signal-to-noise level in the estimate relative to the original signal. If this level is too large, the driving of the linear system by the deterministic error will move the $\Delta(k,p)$ away from zero and keep them there.

The main mode of failure of this procedure, which at its heart is a Newton-Raphson method for solving the nonlinear map, arises when the stable and unstable manifolds are nearly parallel. In practice they are never precisely parallel because of numerical roundoff, but if they become so close to parallel that the forward-backward procedure cannot correctly distinguish between stable and unstable directions, then there will be "leakage" from one manifold to another and this error will be magnified exponentially rapidly until the basic stability of the manifold decomposition iteration can catch up with this error.

In Fig. 31 we show the result of iterating this procedure for 15 times when uniform noise is added to the Hénon map at the level of 2.1% of the rms size of the map amplitude. One can see the regions where the stable and unstable manifolds become almost parallel, called
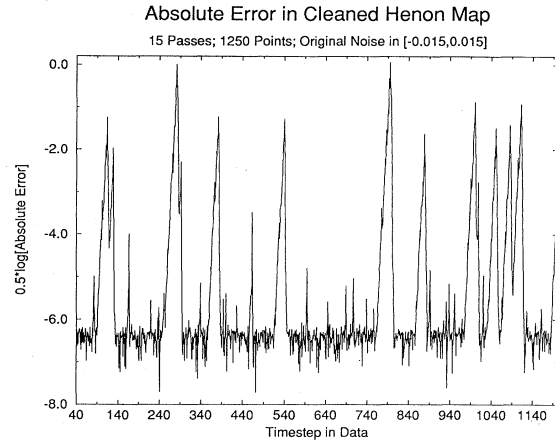

FIG. 31. Absolute error in use of the manifold decomposition method for cleaning Hénon map data contaminated with uniform noise at a level of 2.4% of the rms amplitude of the chaos. Fifteen passes through the algorithm were used. Note the large excursions due to homoclinic tangencies.

homoclinic tangencies in a precise description of the process, and then the orbit recovers. What we show in Fig. 31 is the *absolute error* $\|\mathbf{y}_E(k,p)-\mathbf{y}_1(k)\|$ as a function of k. A plot of $\epsilon_D(k)$ shows no peaks and simply displays noise at the level of machine precision.

The "glitches" caused by the homoclinic tangencies can be cured in at least two ways (Abarbanel *et al.*, 1993) that we are aware of. First, one can go back to the full least-squares problem stated above and in a region around the tangency do a full diagonalization of the matrices that enter. One can know where these tangencies will occur because the Jacobians, which we know analytically since we know $\mathbf{F}_1(\mathbf{x})$, are evaluated at the present best estimate $\mathbf{y}_E(k,p)$. The alternate solution is to scale the step $\Delta(k,p)$ we take in updating the estimated trajectory by a small factor in all regions where a homoclinic tangency might occur. One can do this automatically by putting in a scale factor proportional to the angle between stable and unstable directions or by doing it uniformly across the orbit. The penalty one pays is a larger number of iterations to achieve the same absolute error, but the glitches are avoided.

In either of these cases one arrives at an estimate of the true orbit $\mathbf{y}_1(k)$ which is accurate to the roundoff error of one's machine. So, if one has an orbit of a nonlinear dynamical system that in itself is of some importance, but it is observed to be contaminated by another signal, one can extract that orbit extremely accurately if the dynamics are known. The limitations on the method come from the relative amplitudes of the contamination and the chaotic signal. In practice, when the contamination is 10% or so of the chaos, the method works unreliably. The other important application of the method would be using the chaotic signal as a mask for some signal of interest which has no relation to the dynamics itself. Then

one is able to estimate the chaotic signal contained in the observations extremely accurately and by subtraction expose the signal of interest. Since this application would be most attractive when the amplitude of the signal of interest is much less than the masking chaos, the method should perform quite well.

B. Knowing a signal: probabilistic cleaning

Now we discuss the case in which we are ignorant of the exact dynamics, but we are provided with a clean signal from the system of interest, which was measured at some earlier time. Clearly, we can use any of the techniques from Sec. VI to estimate the dynamics, and then use the resulting models in the algorithms discussed just above. But now we focus on what can be done by using the reference signal $\mathbf{y}_R(k)$ to establish only the statistical properties of the strange attractor. Later one receives another signal from the same system $\mathbf{y}_C(k)$ along with some contamination $\mathbf{z}(k)$. The observed signal is

$\mathbf{s}(k) = \mathbf{y}_C(k) + \mathbf{z}(k)$. We have no *a priori* knowledge of the contamination $\mathbf{z}(k)$ except that is is independent of the clean signal $\mathbf{y}_C(k)$. Except for coming from the same dynamics, $\mathbf{y}_R$ and $\mathbf{y}_C$ are totally uncorrelated because of the chaos itself.

The probabilistic cleaning procedure (Marteau and Abarbanel, 1991) seeks a maximum conditional probability that, having observed the sequence $\mathbf{s}(1), \ldots, \mathbf{s}(N)$, we shall find the real sequence to be $\mathbf{y}_C(1), \ldots, \mathbf{y}_C(N)$. To determine the maximum over the estimated values of the $\mathbf{y}_C$ we write this conditional probability in the form

$$P(\{\mathbf{y}_C\}|\{\mathbf{s}\}) = \frac{P(\{\mathbf{s}\}|\{\mathbf{y}_C\})P(\{\mathbf{y}_C\})}{P(\{\mathbf{s}\})} \tag{126}$$

and note that only the entries in the numerator enter in the maximization over estimated clean orbits.

The probability of the sequence of clean observations can be estimated using the fact that $\mathbf{y}_C(k+1) = \mathbf{F}(\mathbf{y}_C(k))$ so it is first-order Markov. This allows us to write

$$P(\mathbf{y}_C(1), \mathbf{y}_C(2), \ldots, \mathbf{y}_C(m)) = P(\mathbf{y}_C(m)|\mathbf{y}_C(m-1))P(\mathbf{y}_C(1), \mathbf{y}_C(2), \ldots, \mathbf{y}_C(m-1)) \tag{127}$$

and use kernel density estimates of the Markov transition probability $P(\mathbf{y}_C(m)|\mathbf{y}_C(m-1))$ via

$$P(\mathbf{y}_C(m)|\mathbf{y}_C(m-1)) \propto \frac{\displaystyle\sum_{j',j=1}^{N} K(\mathbf{y}_C(m) - \mathbf{y}_R(j))K(\mathbf{y}_C(m-1) - \mathbf{y}_R(j'))}{\displaystyle\sum_{n=1}^{N} K(\mathbf{y}_C(m-1) - \mathbf{y}_R(n))} . \tag{128}$$

The conditional probability $P(\{\mathbf{s}\}|\{\mathbf{y}_C\})$ is estimated by assuming that the contamination $\mathbf{z}$ is independent of the $\mathbf{y}_C$, so

$$P(\mathbf{s}(j); \ j=1, \ m|\mathbf{y}_C(k); \ k=1, m) = \prod_{k=1}^{m-1} P_z(\mathbf{y}_C(k) - \mathbf{s}(k))$$

$$= P(\mathbf{s}(j); j=1, m-1|\mathbf{y}_C(k); k=1, m-1)P_z(\mathbf{y}_C(m) - \mathbf{s}(m)) , \tag{129}$$

with $P_z(\mathbf{x})$ the probability distribution of the contaminants.

These forms for the probabilities lead to a recursive procedure for searching phase space in the vicinity of the observations s. One searches for adjustments to the observations s which maximize the conditional probability $P(\{\mathbf{y}_C\}|\{\mathbf{s}\})$. The procedure is performed recursively along the observed orbit, with adjustments made at the end of each pass through the sequence. Then the observations are replaced by the estimated orbit, the size of the search space is shrunk down toward the orbit by a constant scale factor, and the procedure is repeated. This search procedure is similar to a standard *maximum à posteriori* (Van Trees, 1968) technique with two alterations: (1) the probabilities required in the search are not assumed beforehand, but rather are estimated from knowledge of the reference orbit $\{\mathbf{y}_R(k)\}$, and (2) in the search procedure the initial volume in state space, in which a search is made for adjustments to the observations, is systematically shrunk.

The nonlinear dynamics enters through the reference orbit and the use of statistical quantities of a deterministic system. The method is cruder than the manifold decomposition algorithm, but it also requires much less knowledge of the dynamical system. It works when the signal-to-noise ratio is as low as 0 dB for extracting an initial chaotic signal contaminated by iid (independently and identically distributed) noise. For a signal of interest buried in chaos, it has proven effective when the signal-to-chaos ratio is as low as -20 dB, and even smaller signals will work.

In Fig. 32 we show first a sample of the Lorenz attractor contaminated by uniform noise with a signal (Lorenz chaos)-to-noise ratio of 0.4 dB, and then in Fig. 33 we show the signal recovered by the probabilistic cleaning method after 35 passes through the algorithm. The signal-to-noise ratio in the final signal, relative to the known input clean signal, is 12.6 dB for a gain against noise of 12.2 dB. When the initial signal-to-noise ratio is higher, gains of about 30 to 40 dB with this method are
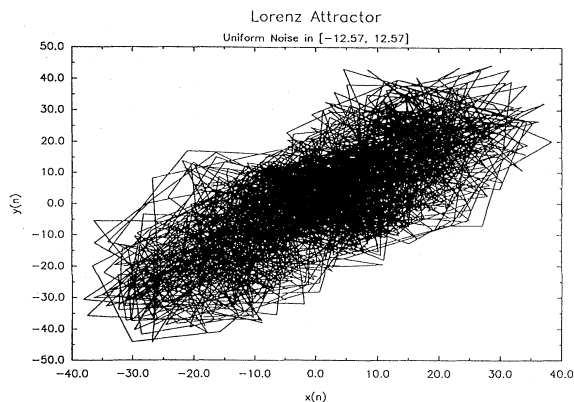
FIG. 32. Noisy Lorenz attractor data. Contamination is uniform noise in the interval $[-12.57, 12.57]$; this gives a signal-to-noise ratio of 0.4 dB.
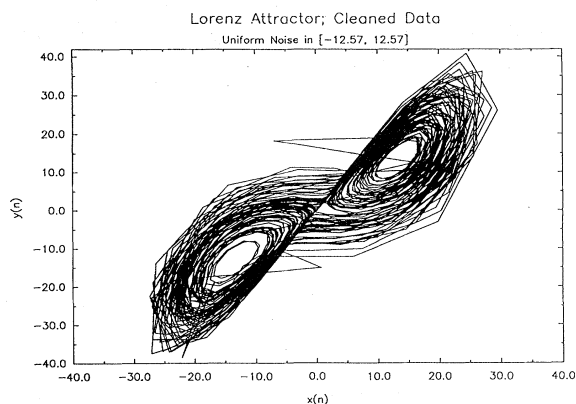

FIG. 33. The cleaned version of the data in Fig. 32. Thirty-five passes of the probabilistic cleaning algorithm are used. The final signal-to-noise ratio, comparing with the known clean data, is 12.6 dB.
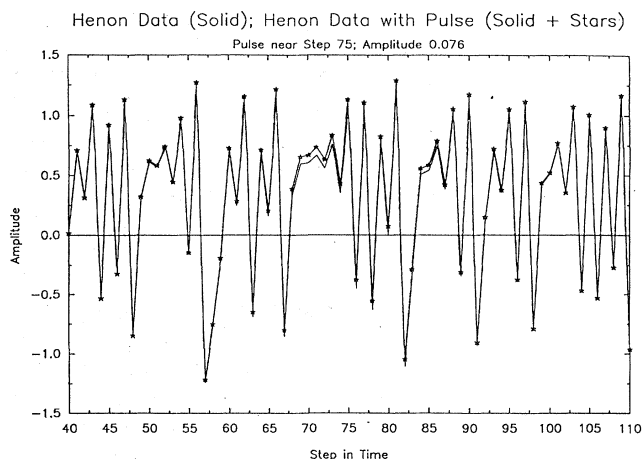

FIG. 34. A pulse added to data from the Hénon map. Signal-to-noise ratio is -20 dB. Pulse is in the neighborhood of time step 75.
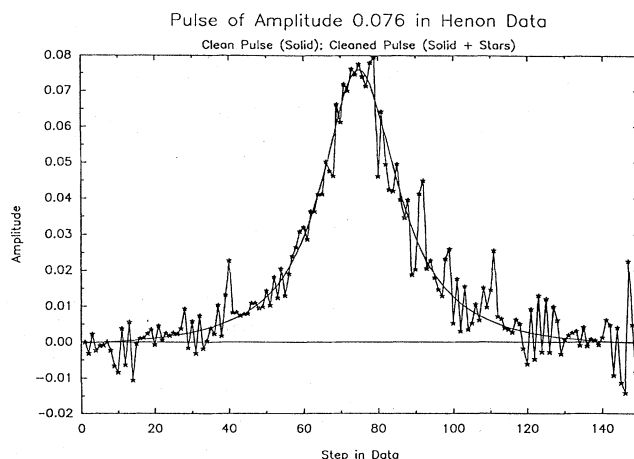

FIG. 35. Pulse recovered from contamination by data from the Hénon map. Signal-to-noise ratio is now $+20.8$ dB.

easily achieved.

In Fig. 34 we show a pulse added to the Hénon map with an initial pulse-to-chaos ratio of -20 dB. This pulse was then extracted from the Hénon chaos to give the result shown in Fig. 35, which has a signal-to-noise ratio of 20.8 dB relative to the known initial pulse. This method can clearly extract small signals out of chaos when a good reference orbit is available.

The probabilistic cleaning method and the manifold decomposition method both use features of the chaos which are simply not seen in a traditional Fourier treatment of signal-separation problems. In conventional approaches to noise reduction the "noise" is considered unstructured stuff in state space with some presumed Fourier spectrum. Here we have seen the power of knowing that chaotic signals possess structure in state space, even though their Fourier spectra are continuous, broadband, and appear as noise to the linear observer. We are confident that this general observation will prove the basis for numerous applications of nonlinear dynamics to signal-based problems in data analysis and signal synthesis or communications.

C. Knowing very little

When we have no foreknowledge of the signal, the dynamics, or the contaminant, we must proceed with a number of assumptions, explicit or implicit. We cannot expect any method to be perfectly general, of course, but within some broad domain of problems we can expect to succeed in separating signals with some success.

In the work that has been done in this area there have been two general strategies:

● *Make local polynomial maps* using neighborhood-to-neighborhood information. Then adjust each point to conform to the map determined by a collection of points; i.e., the whole neighborhood. In some broad sense a local map is determined with some kind of averaging over domains of state space, and after that this averaged dy-

namics is used to realign individual points to a better deterministic map. The work of Kostelich and Yorke (1991) and Farmer and Sidorowich (1991, 1988) is of this kind.

● *Use linear filters locally or globally* and then declare the filtered data to be a better "clean" orbit. Only moving-average or finite-impulse-response filters will be allowed, or we will have altered the state-space structure of the dynamics itself. Many papers have treated this issue. Those of Pikovsky (1986), Landa and Rosenblum (1989), Sauer (1991), Schreiber and Grassberger (1991), and Cawley and Hsu (1992) all fall in this class.

The blind cleaning proceeds in the following fashion. Select neighborhoods around every data point $y(n)$ consisting of N_B points $y^{(r)}(n)$, $r = 1, 2, \ldots, N_B$. Using the $y^{(r)}(n)$ and the $y(r, n + 1)$ into which they map, form a local polynomial map

$$y(r, n + 1) = g(y^{(r)}(n), a)$$
$$= A + B \cdot y^{(r)}(n) + C \cdot y^{(r)}(n) y^{(r)}(n) + \cdots ,$$

(130)

where the coefficients are determined by a local least-squares minimization of the residuals in the local map. Now, focusing on successive points along the orbit $y(n)$ and $y(n + 1)$, find small adjustments $\delta y(n)$ and $\delta y(n + 1)$ which minimize

$$\| y(n + 1) + \delta y(n + 1) - g(y(n) + \delta y(n)) \|^2 \qquad (131)$$

at each point on the orbit. The adjusted points $y(k) + \delta y(k)$ are taken as the cleaned orbit.

In the published work one finds the use only of local linear maps, but the polynomial generalization seems clear. Indeed, it seems to us that use of local-measure-based orthogonal polynomials $\phi_a(x)$ would be precisely in order here as a way of capturing, in a relatively contamination-robust fashion, a good representation of the local dynamics, which would then be used to adjust orbit points step by step. Using only local linear maps is in effect a local linear filter of the data, which does not explicitly use any features of the underlying dynamics, such as details of its manifold structure or its statistics, to separate the signal from the contamination. We suspect that the order of signal separation is limited to $O(1/\sqrt{N_B})$ as a kind of central limit result. This is actually quite sufficient for some purposes, though it is unlikely to allow for separation of signals that have significant overlap in their Fourier spectrum or for separation of signals to recover a useful version of the original, uncontaminated signal as one would desire for a communications application. The method of Schreiber and Grassberger (1991) injects an implicit knowledge of dynamics into their choice of linear filters. They utilize an embedding with information both forward and backward in time from the observation point. Thus, as they argue, the exponential contraction backward and forward along the invariant manifolds in the dynamics can

significantly improve the removal of the contamination.

The second approach to separating signals using no special knowledge of the dynamics rests on the intuition that a separation in the singular values of the local or global sample covariance matrix of the data will occur between the signal, which is presumed to dominate the larger singular values, and the "noise," which is presumed to dominate the smaller singular values. The reasoning is more or less that suggested in an earlier section: the "noise" contributes equally to all singular values, since the singular values of the sample covariance matrix measures how many data, in a least-squares sense, lie in the eigendirection associated with that singular value. If the data are embedded in a rather large dimension $> d_E$, the singular values with the largest index are populated only by the "noise," while the lower-index singular values are governed by the data plus some contamination. While this may be true of white noise, it is quite unlikely for "red" noise, which has most of its power at low frequencies. Nonetheless, given this limitation, the method may work well.

The idea is to form the sample covariance matrix in dimension $d > d_E$ and then project the data (locally or globally) onto the first d_S singular values. This is then taken as new data, and a new time-delay embedding is made. The procedure is continued until the final version of the data is "clean enough." Landa and Rosenblum (1989) project onto the direction corresponding to the largest singular value and argue that the amount of "noise reduction" is proportional to d_E / d multiplied together the number of times the procedure is repeated. Cawley and Hsu (1992) do much the same thing, with interesting twists on how the "new" scalar data are achieved, but locally. The projection onto the singular directions is a linear filter of the data. A combination of local filters, each different, makes for a global nonlinear filter, and in that sense the method of Cawley and Hsu represents a step forward from the earlier work of Landa and Rosenblum. Pikovsky (1986) also does a form of local averaging, but does not use the local singular-value structure in it.

All in all one can say that there is much exploration yet to be done on the properties and limitations of blind signal separation. The hope is certainly that a method, perhaps based on invariant-measure orthogonal polynomials, that utilizes the local nonlinear structure of the data will have some general applicability and validity.

VIII. LINEARLY FILTERED SIGNALS

An important footnote to the discussion of signal separation is to be found in the issue of the effect of filtering by linear filters by convolution on a chaotic signal. All signals are observed through the filter of some measuring instrument. The instrument has its own characteristic response time and bandpass properties. Many filters can be represented as a convolution of the chaotic signal $s(n)$ with some kernel $h(n)$ so that the actual measurement is

$$S(n) = \sum_j h(j) s(n-j) . \tag{132}$$

The question of what happens to the properties of the signal under this filtering has been addressed in many places (Sakai and Tokumaru, 1980; Badii *et al.*, 1988; Mitschke *et al.*, 1988; Chennaoui *et al.*, 1990; Isabelle *et al.*, 1991) with the following basic results. If the filter is a so-called moving-average or finite-impulse-response (FIR) filter, that is, a kernel of the form

$$S(n) = \sum_{j=0}^{M} h(j) s(n-j) , \tag{133}$$

then the dimension and other characteristics of the attractor are unchanged. This is more or less obvious, since this filter is simply a linear change of coordinates from the original $s(k)$, to $S(k)$. By the embedding theorem, invariant properties of the system seen in the reconstructed phase space should be unaltered.

In the case of an autoregressive or infinite-impulse-response (IIR) filter, the story is different. A model of this kind of filter is

$$\mathbf{s}(n+1) = \mathbf{F}(s(n)) ,$$
$$S(n+1) = aS(n) + g(\mathbf{s}(n)) . \tag{134}$$

The first equation is just the dynamics of the system. The second produces from the signal $\mathbf{s}(n)$ an observation $S(n)$ that has its own (*linear*) dynamics $[S(n+1) = aS(n)]$ plus a nonlinear version of $\mathbf{s}(n)$. By adding another degree of freedom to the original dynamics, it is no wonder that one can change the properties of the observed chaotic attractor. One practical implication of this, pointed out by Badii *et al.* (1988), is that low-pass filtering one's data, a quite common practice in experimental setups, can cause an anomalous increase in estimated dimension.

It is rather succinctly pointed out by Isabelle *et al.* (1992) that the criterion for a change in the Lyapunov dimension is the nonconvergence of the sum

$$\sum_{k=0}^{\infty} |h(k)| \exp[k\lambda_N] , \tag{135}$$

where λ_N is the smallest Lyapunov exponent of the total system. A FIR filter always converges, while an IIR filter may fail to converge and thus change properties of the apparent attractor.

The nonlinear dynamics reason for this is that, if the additional dynamics characterizing the filter results in Lyapunov exponents that are large and negative, their effect on the original system is essentially unnoticed because small deviations from the original system are rapidly damped out. If the additional exponent decreases in magnitude, eventually it gets within the range of the original exponents and affects the properties of the original attractor because its effects on orbits becomes both measurable and numerically substantial.

In essence then, the issue of linear filters on a chaotic signal is rather straightforward, since linear systems are

rather well understood. The results stated should act as a sharp warning, however, since numerical issues associated with the clearly stated theory may lead to severe misunderstandings. Nonetheless, from the theoretical point of view, this is a well understood area.

IX. CONTROL AND SYNCHRONIZATION OF CHAOTIC SYSTEMS

A. Controlling chaos

1. Using nonlinear dynamics

In this part of our review we present a method that has been demonstrated in experiments for moving a chaotic system from irregular to periodic, regular behavior. The method has two elements: (1) an application of relatively standard control-theory methods, beginning with a linearization around the orbits, and (2) an innovative idea on how to utilize properties of the *nonlinear* dynamics to achieve linearized control in an exponentially rapid fashion. One of the keys to the method is detailed knowledge of the phase space and dynamics of the system. This can be gleaned from observations on the system in the manner we have outlined throughout this review. An attractive feature of the method is its reliance on quite small changes in parameter values to achieve control once the system has been maneuvered into the appropriate linearized regime by the nonlinear aspects of the problem.

We assume that the experiment is driven by some outside force and the data are in the form of a time series taken once every period of the drive. We call such a data collection technique stroboscopic. An equally useful assumption is that there is some type of fundamental frequency in the nondriven dynamics. We would then use one period of this frequency for our stroboscopic time. Next we shall assume an adjustable parameter which we call p. Let the data vectors on the stroboscopic surface of a section be given by $\mathbf{v}(n)$ for $n = 1, 2, \ldots, N$.

Embedded within a hyperbolic strange attractor is a dense set of unstable periodic orbits (Devany and Nitecki, 1979). The presence of these unstable periodic orbits and the absence of any stable ones is a characteristic sign of chaotic motion. The goal of the control algorithm is to move the system from chaotic behavior to motion along a selected periodic orbit. The orbit is *unstable* in the original, uncontrolled system but clearly *stable* in the controlled system. The reason for this apparent paradox is that the controlled system has a larger phase space, since the parameter that was fixed in the original dynamics is now changed as a function of time. In the new dynamics, the periodic orbit of the original system becomes stable.

The first step in solving this problem is obvious. We must find the locations of the periodic orbits. We discuss how this can be accomplished for the two cases intro-

duced above (Auerbach *et al.*, 1987). In the first case the system is not driven, but there is some fundamental frequency. We choose some, typically small, positive number ϵ. For each data vector $\mathbf{y}(n)$, we find the smallest time index $k > n$ such that

$$\|\mathbf{y}(n) - \mathbf{y}(k)\| < \epsilon \ .$$

We define the period of this point by $M = k - n$. As an example, experimental data collected from the Belousov-Zhabotinski (BZ) chemical reaction (Roux *et al.*, 1983; Coffman *et al.*, 1987) indicate that there is a fundamental periodic orbit $\mathbf{y}_M(1) \mapsto \mathbf{y}_M(2) \mapsto \cdots \mapsto \mathbf{y}_M(M) = \mathbf{y}_M(1)$ near $M = 125$. Our stroboscopic vectors are $\mathbf{v}(m) = \mathbf{y}(mM)$ for $m = 1, 2, \ldots$.

When the system is driven, the procedure is even easier. We again look for recurrent points. Typically for driven systems with data acquired once every period of the drive we will find $M = 1$. There is only one major difference between this system and the previous one. For the previous system the data vectors between $\mathbf{y}_M(1)$ and $\mathbf{y}_M(M) = \mathbf{y}_M(1)$ give the trajectory covered by the periodic orbit. For data that are taken only once every period of the drive we usually do not have the details of trajectory of the periodic orbit with the lowest period.

The evolution of orbits is taken to be governed by a map $\mathbf{v}(n+1) = \mathbf{F}(\mathbf{v}(n), p)$ with some parameters that are at our disposal. The exact phase-space location of the periodic orbit $\mathbf{v}_M \mapsto \mathbf{v}_M$ is a function of the parameter p. We choose the origin of values of p to be on the orbit we have found.

We want to vary p in the neighborhood of $p = 0$, when the orbit of the undriven system comes "close" to the $\mathbf{v}_M(0) = \mathbf{v}_M(p = 0)$. Now, we must either wait until the experimental chaotic orbit comes near this or use some technique to force it into this region (Shinbrot *et al.*, 1990, 1992). Once $\mathbf{v}$ comes near $\mathbf{v}_M(0)$ the control is given by the following steps, illustrated in Fig. 36 (Ditto *et al.*, 1990).

The part of the figure on the left indicates the situation just prior to the control mechanism's being activated. The unstable fixed point is indicated by $\mathbf{v}_M(0)$. Using the techniques we described in the previous section, we evaluate $\mathbf{DF}[\mathbf{v}_M(0)]$, the Jacobian matrix of the mapping
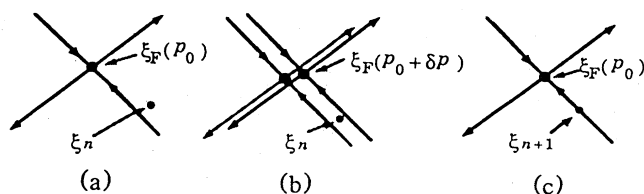

FIG. 36. Sketch of control algorithm invented by Ott *et al.* (1990): (a) if the nth iterate ξ_n is near the fixed point $\xi_F(p_0)$ then (b) one changes the parameter p from p_0 to $p_0 + \delta p$ to move the fixed point so as to (c) force the next iterate onto the stable manifold of the undisturbed fixed point (from Ditto *et al.*, 1990).

at the point $\mathbf{v}_M(0)$. We also determine the directions of the stable and unstable eigenvectors of the Jacobian. The location of the phase-space trajectory at time n is indicated in this figure by the vector $\mathbf{v}(n)$.

The control mechanism is an instantaneous change in the parameter from $p = 0$ to $p = \bar{p}$. For $p = \bar{p}$ the cross shown in solid lines on the left figure becomes the cross shown in dashed lines in the center figure (Ditto *et al.*, 1990). This change is nothing more than the small change in the location of the periodic orbit after a change in the parameter, i.e., $\mathbf{v}_M(0)$ moves to $\mathbf{v}_M(\bar{p})$. We have indicated the location of the cross prior to the change by a solid line. What is important is that the change in the parameter has created a situation that will force the trajectory of $\mathbf{v}(n)$ in the direction of the stable manifold of $\mathbf{v}_M(0)$. In fact, one chooses the magnitude of the parameter $\bar{p}$ so that on the next time around the orbit $\mathbf{v}(n+1)$ is on the stable manifold of $\mathbf{v}_M(0)$. One then instantaneously moves p from $\bar{p}$ back to $p = 0$. From now on the trajectory will march along the stable manifold (of the uncontrolled system) until it reaches the periodic point $\mathbf{v}_M(0)$.

The hard part of the control process is determining the value of $\bar{p}$. When $p = \bar{p}$ the dynamics near $\mathbf{v}_M(\bar{p})$ can be approximated by the linear map

$$\mathbf{v}(n+1) = \mathbf{v}_M(\bar{p}) + \mathbf{DF}_{\bar{p}} \cdot [\mathbf{v}(n) - \mathbf{v}_M(\bar{p})] \ . \tag{136}$$

We must choose $\bar{p}$ so that in one period of the drive the phase-space trajectory from $\mathbf{v}(n)$ to $\mathbf{v}(n+1)$ lands on the stable manifold of $\mathbf{v}_M(0)$. To choose this value of p define the vector $\mathbf{g}$ via the following rule:

$$\mathbf{g} = \frac{\partial \mathbf{v}_M}{\partial p} \bigg|_{p = p_*}$$

$$\simeq \frac{\mathbf{v}_M(\bar{p})}{\bar{p}} \ .$$

To simplify the calculations we shall confine ourselves to $d = 2$ dimensions, though the method is readily generalized to higher dimensions. Since $\bar{p}$ is small, we assume $\mathbf{DF}_{\bar{p}}$ can be approximated by $\mathbf{DF}_{p=0} \equiv \mathbf{DF}$. Using the techniques described above one can find $\mathbf{DF}$ from the data vectors $\mathbf{v}(n)$. Furthermore, one can write $\mathbf{DF}$ as

$$\mathbf{DF} = \Lambda_u \hat{\mathbf{u}} \mathbf{f}_u + \Lambda_s \hat{\mathbf{s}} \mathbf{f}_s \ , \tag{137}$$

where the Λ_s and Λ_u are the stable and unstable eigenvalues, and $\hat{\mathbf{s}}$ and $\hat{\mathbf{u}}$ are the stable and unstable eigenvectors. The vectors $\mathbf{f}_s$ and $\mathbf{f}_u$ are the stable and unstable contravariant eigenvectors defined by $\mathbf{f}_u \cdot \hat{\mathbf{u}} = 1$, $\mathbf{f}_u \cdot \hat{\mathbf{s}} = 0$, $\mathbf{f}_s \cdot \hat{\mathbf{s}} = 1$, and $\mathbf{f}_s \cdot \hat{\mathbf{u}} = 0$.

Equations (136) and (137) and the vector $\mathbf{g}$ are all that is necessary to define the control mechanism. The motion along the stable manifold of $\mathbf{v}_M$ brings the original orbit very rapidly to the vicinity of the unstable periodic orbit $\mathbf{v}_M$. Once the orbit is in this vicinity, small changes in p proportional to $\mathbf{v} - \mathbf{v}_M$ are made. This linearization defines a new linear map near $\mathbf{v}_M$, and we

require the eigenvalues of this new linear problem to be within the unit circle. The requirement on the change in parameter necessary to control the oscillations is given by

$$\bar{p} = \frac{\Lambda_u}{\Lambda_u - 1} \left[\frac{\mathbf{v}(n) \cdot \mathbf{f_u}}{\mathbf{g} \cdot \mathbf{f_u}} \right] . \tag{138}$$

This type of control mechanism has been experimentally verified in an experiment on magnetoelastic ribbons (Ditto *et al.*, 1990). The device involves a ribbon whose Young's modulus changes under the influence of small magnetic fields. An inverted ribbon is placed in the Earth's gravitational field so that initially it buckles due to gravity. By adding a vertical oscillating magnetic field, one creates a driven chaotic oscillator. For this experiment the force driving the oscillation is gravity, while the control parameter is the magnetic field strength. The results of this experiment are shown in Figs. 37 and 38.

In Fig. 37 the iteration number represents passes through the stroboscopic surface of a section, which is one period of the drive. The gray portion for iterates less than approximately 2350 represents the chaotic motion of the ribbon before the control is initiated. For an iterate near 2350, the motion of the ribbon passes near an unstable period-one orbit, and the control is initiated. After initiating control the motion of the ribbon remains near the periodic orbit. In the figure this is indicated by the fact that the location of the orbit on the surface of the section is constant near 3.4 for all subsequent intersections. The period of the driving of this magnetic field is typically 0.85 Hz; thus the period-one orbit has period $1/0.85$ Hz~ 1.2 sec. The experimenters have been able to confine the motion of the ribbon to the unstable orbit for as long as 64 hours. Clearly the control mechanisms works.

Figure 38 is an admirable piece of showing off on the part of the researchers. First they control near a period-one orbit. Then they change their minds and control near period-two orbits. Then they change their minds again and control back on the original period-one orbit. Control has also been demonstrated near period-four orbits.
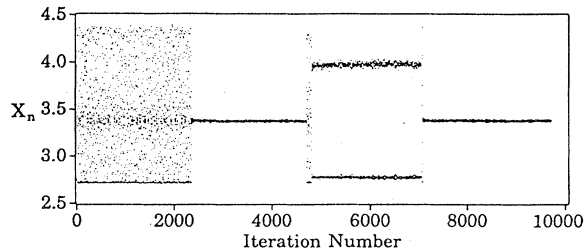
FIG. 38. Time series of voltages from the magnetoelastic ribbon. System is switched from no control to controlling the (unstable) period-1 orbit (fixed point) at iteration 2360, to controlling period-2 orbit at $n = 4800$, and back to the fixed point at $n = 7100$.

2. More traditional methods

It is also possible to influence the dynamics of a chaotic system using traditional techniques from control theory (Cremers and Hübler, 1986; Eisenhammer *et al.*, 1991; Singer and Bau, 1991; Singer *et al.*, 1991). To illustrate this approach we imagine that the uncontrolled dynamics is given by the equation

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}[\mathbf{y}, t] \tag{139}$$

and that $\mathbf{y}^*(t)$ represents the trajectory we are interested in controlling (stabilizing or, possibly, destabilizing).

Traditional control techniques implement a feedback mechanism to manipulate the dynamics. The feedback is an externally applied force we denote as $\mathbf{H}[\mathbf{y}(t) - \mathbf{y}^*(t), t; \rho]$. The vector ρ contains adjustable parameters, which are often called the feedback gains. When the control mechanism is implemented the dynamics changes from Eq. (139) to

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{F}[\mathbf{y}(t), t] + \mathbf{H}[\mathbf{y}(t) - \mathbf{y}^*(t), t; \rho] . \tag{140}$$

The purpose of $\mathbf{H}$ is clear from Eq. (140). $\mathbf{H}$ is a force that gives nudges to orbits. If $\mathbf{y}^*$ is unstable without the control, and the trajectory $\mathbf{y}$ moves away from $\mathbf{y}^*$, then $\mathbf{H}$ can push it back towards $\mathbf{y}^*$ and thus stabilize an undriven, unstable orbit.

This type of control has been implemented in a laboratory thermofluid experiment. The experimental device consisted of a torus heated from below and cooled from above. This thermosyphon implements convection, behaving in a similar fashion to the original 1963 Lorenz equations. The feedback was implemented by controlling the current flow into the heater as applied to the lower part of the torus. The control part of the heating was $\kappa(\Delta T - \Delta T_0)$ where ΔT is the actual temperature difference between two selected points and ΔT_0 is the desired temperature difference. The value of κ was much smaller than the constant heat component applied to the bottom part of the torus (Bau and Singer, 1992).

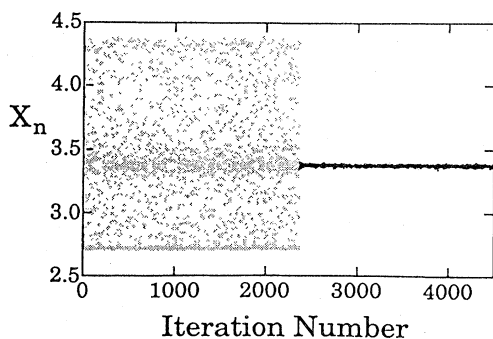Some of the results of this experiment are shown in

FIG. 37. Time series of the voltages stroboscopically sampled from the sensor attached to magnetoelastic ribbon (Ditto *et al.*, 1990). Control was initiated after iteration 2350.
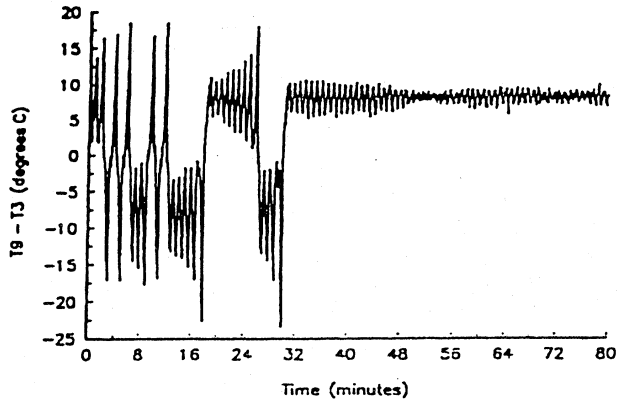
FIG. 39. Experimentally measured temperature oscillations as a function of time. The controller was activated 33 min after the start of the experiment (Bau and Singer, 1992).

Figs. 39 and 40. Figure 39 shows the temperature profile ΔT as a function of time. The control is implemented at time equal to 33 minutes. The control mechanism causes the dynamics of the temperature to stabilize about a particular value that indicates steady laminar flow, corresponding to counterclockwise rotation of the fluid. Figure 40 is showing off in a manner similar to the magnetoelastic ribbon example. In this case the control is used to change from steady counterclockwise motion to steady clockwise motion.

It is important to note that the two main types of control we have discussed are not equivalent. The first type involves changing parameters of the system to achieve control. It also used a knowledge of the location and function of the stable and unstable manifolds of the dynamics. The dynamics in this control technique is given by

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}[\mathbf{y};\rho] + \frac{\partial \mathbf{F}}{\partial \rho} \cdot \delta\rho \ .$$

In this equation there is no need for a new class of *external* forcings as in the one found in Eq. (140). Instead the
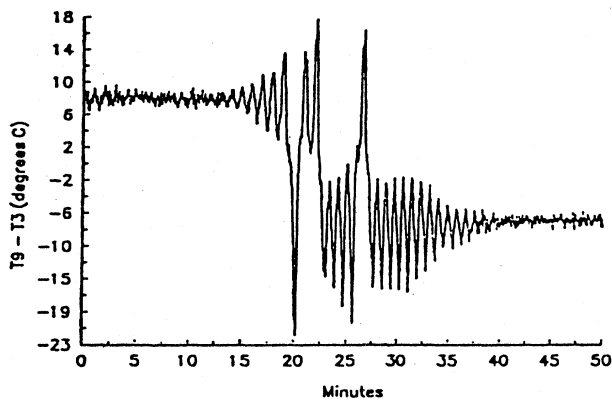

FIG. 40. Switching the direction of the flow in a thermoconvection experiment from clockwise to counterclockwise using the controller (Bau and Singer, 1992).

role of external forces is to effect changes in the parameters of the system rather than the form of the system. The main distinctions are the use of the nonlinear dynamics of the system in the first method and the important fact that only infinitesimal external forces are required to change parameters. This latter point comes from the use in the method of the exponential approach along stable manifolds.

3. Targeting

The final type of control that we shall discuss differs from both of the two previous cases. As with many things, the control mechanisms work well when they work. After the system moves into a region of the phase space that we wish to control, we are able to implement the control with great success. However, we have to wait until the dynamics approaches the desired region before control can be initiated. This can often take a long time. What would be very useful is some technique to persuade the dynamics to approach the control region quickly. In other words, we need a net that can be thrown over the phase space in order to capture the trajectory. Once it is in the net we can drag it back to our control cage and lock up its behavior.

There are two different techniques for directing trajectories towards specific regions of phase space. They differ in detail and implementation, but the *targeting* goal is the same. The first technique is useful when the target point is on the attractor given by the data. When the target is not on the attractor given by the data, the second technique is required.

The first targeting technique involves both forward and backward iterations of small regions of the phase space. Assume that the data are given on a surface of a section by $\mathbf{v}(n)$, $n = 1, 2, \ldots, N$. For purposes of illustration also assume that the dynamics is in two phase-space dimensions and is invertible. Let the initial point be given by $\mathbf{v}_i$ and the target point by $\mathbf{v}_T$. Without altering the dynamics, the time necessary for a typical point on the attractor to enter a ball of size ϵ_t centered on $\mathbf{v}_T$ scales like $\tau \sim 1/\mu(\epsilon_t)$, where μ is the natural measure on the attractor. This measure scales with d_1, the information dimension of the attractor (Badii and Politi, 1985), so $\tau \sim (1/\epsilon_t)^{d_1}$. Obviously for small ϵ_t the time it takes for a trajectory to wander into the target region can be quite long. This phenomenon is evident in Fig. 41, where it took thousands of iterations before the trajectory was near enough to the period-one target to implement control. Other examples of even longer duration exist (Shinbrot *et al.*, 1990).

Clearly this is not acceptable, and happily it can be improved upon. Assume that the dynamics on the stroboscopic surface is given by

$$\mathbf{v}(n+1) = \mathbf{F}[\mathbf{v}(n);p] \ ,$$

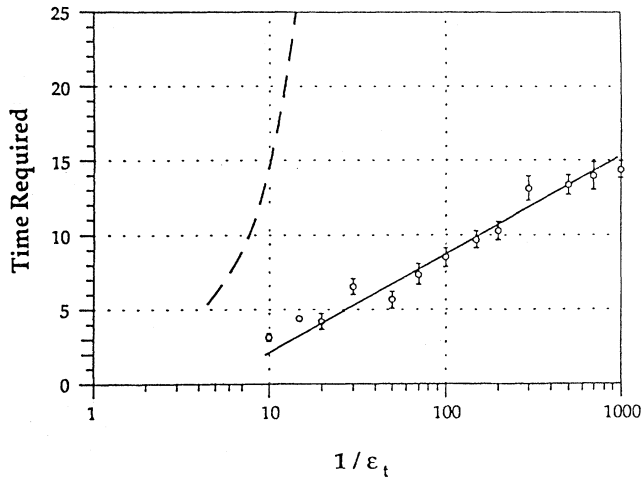where we have explicitly indicated the parameter depen-

FIG. 41. Average time required to reach a typical target neighborhood from a typical initial position as a function of the inverse neighborhood size: dashed line, no control; circles and solid line, with control (from Shinbrot *et al.*, 1990).

dence of $\mathbf{F}$ by p. For now we shall use only one parameter, but that is not necessary. The change in one iteration on the surface due to a change in the parameter is

$$\delta \mathbf{v}(n+1) = \hat{\mathbf{v}}(n+1) - \mathbf{v}(n+1)$$

$$= \frac{\partial \mathbf{F}[\mathbf{v}(n)]}{\partial p} \delta p .$$

If we choose δp small, then $\hat{\mathbf{v}}(n+1) - \mathbf{v}(n+1)$ sweeps out a small line segment through the point $\mathbf{v}(n+1)$. After m steps through the unperturbed map $\mathbf{F}[\mathbf{v};p]$, the line segment will have grown to a length $\hat{\mathbf{v}}(n+m) - \mathbf{v}(n+M) \sim \exp[m\lambda_1] \delta \mathbf{v}(n+1)$. Eventually the length will grow to the same order as the size of the attractor, which we shall conveniently call 1. The number of iterations necessary for this to happen is $m_1 = -\log[\delta \mathbf{v}(n+1)]/\lambda_1$. We now imagine iterating the target region ϵ_T backwards in time. This process will transform the target region into an elongated ellipse. The length of the ellipse will grow like $\exp[\lambda_2 m]$. After a total of m_2 iterations the ellipse will have been stretched to the point where its length is the same order as the size of the attractor.

Generically the line formed by the forward iterations of $\delta \mathbf{v}(n+1)$ will intersect the ellipse formed by the backward iterations of the region ϵ_T. For some value of the parameter, call it p_1, in the range $p + \delta p$ the end of the line segment will be *inside* the backward images of ϵ_T. By changing the parameter at time n from the value p to the value p_1, we can guarantee that the trajectory will land inside the target region. The number of iterations required for this to happen is

$$\tau = m_1 + m_2$$

$$= -\frac{\ln[\delta \mathbf{v}(n+1)]}{\lambda_1} - \frac{\ln[\epsilon_T]}{\lambda_2} .$$

The targeting strategy above utilizes only one change in the parameter. After the initial perturbation, the parameter value is returned to p from p_1. It also assumes that one has complete knowledge of the dynamics. In the presence of noise, or incomplete knowledge of, the dynamics the above procedure will not work without modifications. The modifications are very small. Instead of only perturbing the dynamics once at time n, one should perturb the dynamics at each iteration through the surface. Therefore the entire forward-backward calculation should be done at $\mathbf{v}(n)$, $\mathbf{v}(n+1)$, $\mathbf{v}(n+2)$, etc. until the target region is reached.

The forward-backward technique has been demonstrated to be effective in computer experiments (Shinbrot *et al.*, 1990). More impressive is experimental implementation of these techniques. For the experimental situation only the forward portion of the targeting method was used. This is equivalent to setting $m_2 = 0$ and waiting for the stretched line segment to contact ϵ_T. The results indicate that for the magnetoelastic ribbon experiment without targeting it required about 500 iterations to reach a typically chosen ϵ_T. With control the number of iterations typically was reduced to 20 (Shinbrot *et al.*, 1992). The power of using the nonlinear dynamics of the process is quite apparent.

The second type of targeting is much more ambitious. The basic problem is still the same. Assume that initially the trajectory is located at $\mathbf{y}_i$. How can we control the system so that it quickly arrives at the location $\mathbf{y}_f$? Notice that the locations $\mathbf{y}_i$ and $\mathbf{y}_f$ are legitimate states of the dynamics for some value of the parameters. Other than that they are arbitrary. The type of control that accomplishes this broad category of behavior has the drawback of requiring a complete model of the dynamics for all values of the parameter.

The first step in the control technique involves using cell maps to determine the dynamics of the entire phase space for a range of parameter values, $\rho^{(j)}$ where $j = 1, 2, \ldots, N_p$ (Hsu, 1987). Each different value of the control parameter $\rho^{(j)}$ implies a different dynamical system for Eq. (139) and hence different trajectories in its corresponding cell maps. The dynamics may even experience bifurcations without a conceptual change in the control technique. Once the cell maps have been established, one searches through the possible trajectories. A final path is chosen by using pieces of the possible trajectories provided by the cell maps. Thus the entire controlled trajectory can be given by the sequence of cells $S = (S^{(0)}, S^{(1)}, \ldots, S^{(l)})$. Each of the individual $S^{(j)}$'s is a true trajectory of the dynamical system $\dot{\mathbf{y}} = \mathbf{F}[\mathbf{y}, t; \rho^{(j)}]$. The criteria for determining the $S^{(j)}$'s is that S be as short as possible.

The details of the construction of the cell maps and the search algorithm for determining the shortest S from the large number of possible values can be found in the references (Bradley, 1991, 1992). Once the final phase-space destination is achieved, standard control techniques, as above, can be employed to ensure that $\mathbf{y}$ stays near $\mathbf{y}_f$.

B. Synchronization and chaotic driving

Many physical phenomena can be modeled as an underlying dynamical system driven by external forces. The most often studied types of driving are periodic and steady forcing. A down-to-earth example is given by a system that is driven with a period half that of the natural oscillation. This models pumping a playground swing.

A type of driving that is new to our field of inquiry is chaotic driving (Kowalski *et al.*, 1990; Pecora and Carroll, 1990, 1991a, 1991b). Under this type of scenario one has a dynamical system that is driven by the output of another dynamical system. If the second dynamical system, the one that is providing the drive, is exhibiting chaotic motion, then we say that the first system is undergoing chaotic driving. In Fig. 42 we schematically illustrate chaotic driving. The first dynamical system, the one that is doing the driving, we call the driving system. The second dynamical system, the one that is being driven we call the response system.

Imagine dividing the two dynamical systems into three parts. The degrees of freedom of the driving system are $\mathbf{u}(t)$ and $\mathbf{v}(t)$ while those of the response system are called $\mathbf{w}(t)$. The vector fields for the driving and response are $\mathbf{g}$, $\mathbf{f}$, and $\mathbf{h}$, and the equations of motion are

$$\frac{d\mathbf{u}}{dt} = \mathbf{g}[\mathbf{u};\mathbf{v}] \ ,$$

$$\frac{d\mathbf{v}}{dt} = \mathbf{f}[\mathbf{u};\mathbf{v}] \ , \qquad (141)$$

$$\frac{d\mathbf{w}}{dt} = \mathbf{h}[\mathbf{w};\mathbf{v}] \ .$$

The first and second equations of motion involve the driving system. We have split it into two parts. The first part represents the variables in the driving system that are not involved in driving the response. The second part represents the variables that are actually involved in the driving of the second dynamical system. Finally, the third equation represents the response system. We shall assume that the combined system has n equations of motion, while partitions $\mathbf{g}$, $\mathbf{f}$, and $\mathbf{h}$ have k, m, and l equations of motion, respectively.

A simple familiar example is provided by a damped driven pendulum: $\ddot{\theta} + \gamma\dot{\theta} + \sin(\theta) = \cos(\omega t)$. This system can be written in the form of Eqs. (141) through the identification of variables: $\theta = w_1$, $\dot{\theta} = w_2$, $\phi = v$, and $\dot{\phi} = u$. The result is written
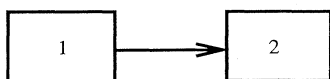

FIG. 42. Sketch of chaotic driving: chaotic generator 1 drives another chaotic system 2.

$$\dot{w}_1 = w_2 \ ,$$

$$\dot{w}_2 = -\gamma w_2 - \sin(w_1) + v \ ,$$

$$\dot{v} = u \ ,$$

$$\dot{u} = -\omega^2 v \ .$$

In this example $m = 1$, $k = 1$, and $l = 2$. The driving system is a simple harmonic oscillator and the response is the damped pendulum.

In chaotic driving we replace the simple harmonic oscillator with a chaotic system. When the chaotic system is identical to the response system the most dramatic effects occur. Under these conditions *the chaotic response system may synchronize with the chaotic driving system.* Synchronization means here that the two systems follow the same trajectory in phase space (Afraimovich *et al.*, 1986). This occurs despite the fact that the response system has a different initial condition from the drive system. Under normal chaotic circumstances a different initial condition would imply that the two trajectories would diverge exponentially on the attractor. Thus the fact that chaotic driving can cause two systems to follow the same trajectory for a long time is quite an attention getter.

When does synchronization occur? That is, when will two nearby initial conditions in $\mathbf{w}$ converge onto the same orbit? The condition that determines whether or not a driven dynamical system will synchronize is obtained by examining the values of the *conditional Lyapunov exponents.* To determine the conditional Lyapunov exponents, as well as to define them, we must simultaneously solve Eqs. (141) and the variational equations of the response system,

$$\frac{d(\Delta\mathbf{w})}{dt} = \mathbf{DF}_{\mathbf{w}}[\mathbf{v},\mathbf{w}] \cdot \Delta\mathbf{w} \ , \qquad (142)$$

where $\Delta\mathbf{w} = \mathbf{w}'(t) - \mathbf{w}(t)$ and

$$(\mathbf{DF}_{\mathbf{w}})_{\alpha,\beta} = \frac{\partial h_\alpha}{\partial w_\beta} \ .$$

The eigenvalues $\exp[L\hat{\lambda}]$ of compositions of $\mathbf{DF}_{\mathbf{w}}$ are the conditional Lyapunov exponents.

For periodic trajectories in $\mathbf{w}$, understanding the behavior of Eqs. (141) and (142) and its influence on $\Delta\mathbf{w}$ is a straightforward application of Floquet theory. But we are interested in chaotic instead of periodic trajectories. A moment's thought convinces us that determining the behavior of $\Delta\mathbf{w}$ is equivalent to calculating the Lyapunov exponents associated with Eqs. (141) and (142). Numerical experiments have determined that the Lyapunov exponents associated with Eqs. (141) *and* (142) are not the same as the Lyapunov exponents calculated only from Eq. (141) alone. Nor are they a subset of those calculated from Eq. (141) alone. The first of these facts is obvious. The variational equations associated with Eq. (141) have an $n \times n$ Jacobian and will produce n Lyapunov exponents. On the other hand, Eq. (142) has an $l \times l$ Jacobian and will produce l conditional

TABLE III. Conditional Lyapunov exponents for synchronized Lorenz attractors. With z as the drive, one of the conditional exponents is positive, so this will not synchronize.

System	Drive	Response	Conditional Lyapunov exponents
Lorenz $\sigma = 16$ $b = 4$ $r = 45.92$	x	(y,z)	$(-2.50, -2.5)$
	y	(x,z)	$(-2.95, -16.0)$
	z	(x,y)	$(0.007\,89, -17.0)$

Lyapunov exponents.

The set of l Lyapunov exponents associated with Eqs. (141) and (142) are the conditional Lyapunov exponents. If all l conditional exponents are less than zero, then the subsystem, $\mathbf{w}$, will synchronize [i.e., $\mathbf{w}'(t) \to \mathbf{w}(t)$ as $t \to \infty$]. If we denote the conditional Lyapunov exponents as $0 > \hat{\lambda}_1 \le \hat{\lambda}_2 \le \cdots \le \hat{\lambda}_l$ then the convergence rate is given by $\exp[\hat{\lambda}_1 t]$. Numerical experiments on the Lorenz system are given in Table III. As the table shows, for the parameter values used one can implement either x or y as the drive. The corresponding subsystems $[(y,z)$ in the case of x drive and (x,z) in the case of y drive] have two negative conditional Lyapunov exponents. The third possible subsystem has z as the driver and possesses a small *positive* conditional Lyapunov exponent. The positive conditional Lyapunov exponent implies that this type of drive/response system will not synchronize.

One can cascade several dynamical systems together (Carroll and Pecora, 1992; Oppenheim *et al.*, 1992) and achieve synchronization in all of them by using only one drive. An example of this type of system is given in Fig. 43. Here we show a system of three cascaded Lorenz systems. The x coordinate of the first is used to drive the second. Similarly, the y coordinate of the second is used to drive the third. One could extend this cascade indefinitely without changing the results. Moreover, since the second and third systems respond to the drive of the first system there is really only one drive, namely, the x coordinate of the first system.
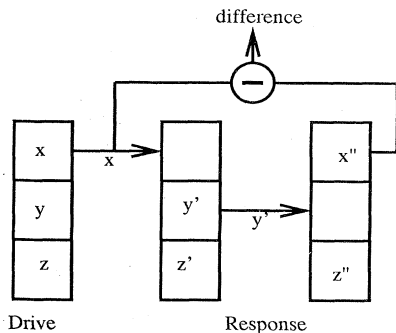

FIG. 43. Sketch of cascaded chaotic systems.

X. SPATIO-TEMPORAL CHAOS

As we remarked in the introduction, the study of spatio-temporal chaos is much further from a full understanding than solely temporal chaotic behavior. The situation is changing quite quickly, and while we shall merely touch upon our own expectation of where this area will go, it seems certain that the developments will bring the analysis of spatio-temporal observations to the same stage as that of temporal chaos, reviewed in this article. So what we include now is less a review than, we hope, a preview to indicate how the subject might develop from the results reviewed throughout this article.

In the last few years increasing attention has naturally been focused on spatio-temporal chaos in various experimental setups such as Rayleigh-Bénard convection (Manneville, 1990), Taylor-Couette flow (Anderek and Hayot, 1992), and surface excitations of the Faraday sort (Gollub and Ramshankar, 1991). On the theoretical side there have been extensive studies of cellular automata which provide coarse graining in the dynamical variables as well as in the independent variables of space and time, coupled map lattices (Keeler and Farmer, 1986), which study the interaction between known chaotic maps coupled together on spatial lattices, and model physical equations such as the Kuramoto-Sivashinsky and various Ginzburg-Landau equations (Kuramoto, 1984). Discussion of those studies as well as relevant mechanisms of transition to chaos go far beyond our scope. In that regard, however, we recommend the review by Cross and Hohenberg (1993), which reports on studies in which the instabilities of linearized dynamics can be traced as the origin of observed space-time patterns in a variety of experiments.

Our present aim is to outline the approach to analysis of data from spatio-temporal systems following the philosophy and methods developed for the pure temporal cases in previous sections. In other words, we need to develop machinery that allows us to calculate meaningful characteristics identifying the spatio-temporal behavior and to build models for prediction of behavior of the spatio-temporal system—and to do all of this working with data.

The main difference (and difficulty) which comes with spatial degrees of freedom is the very high (or even, typically, infinite) dimension of the system. Of course, one may take only one observable from one space location and try to reconstruct the dynamics in the usual way. The idea would be that since the system is totally coupled, by the embedding theorem, *any* measurement should reveal the full structure of all degrees of freedom in interaction. This standard approach of treating a spatio-temporal system as though it were purely temporal, while possible in principle, is practically infeasible with existing or projected computational capability. It seems much more promising to account for the spatial extension explicitly at both the phase-space reconstruction and model-making stages of the problem. We have

seen in the preceding sections that a scalar time series $s(n)$ represents the projection of a multidimensional geometric object onto the observation axis. In the case of spatially extended systems, we deal with similar scalar measurements labeled by space as well as time.

To be concrete, we restrict our discussion to one space dimension, call it x, and time. The amplitude which is measured we call $a(x,t)$. The first step towards the reconstruction of the spatio-temporal phase space is to consider just space series rather than temporal ones (Afraimovich *et al.*, 1992). This approach is quite legitimate if we study the "frozen" spatial disorder resulting, for example, from the evolution of a gradient spatio-temporal system of the form

$$\frac{\partial a(x,t)}{\partial t} = -\frac{\delta F[a]}{\delta a}, \qquad (143)$$

where $F(a)$ is some free-energy functional. For this kind of dynamics, it is easy to show that any initial state

$a(x,0)$ evolves to a steady state $a_\infty(x)$. For this pure space series all the machinery developed for time series can be directly reformulated by substituting x for t. In this way one can calculate fractal dimension of the "space" attractor, evaluate spatial Lyapunov exponents, make models, etc.

A richer situation arises when we have spatio-temporal dynamics and measure the spatio-temporal series $a(x,t)$. The first thing we want to deal with is the correct phase space for this form of measurement. Recall that we reconstruct the phase space in a temporal case using time-delay coordinates. Determination of the time delay T in the temporal case requires that the measurements $x(t)$ and $x(t+T)$ were independent in an information-theoretic sense. In the case of space as well as time, the natural modification of this procedure is to seek a *space delay* at which observations become slightly uncorrelated (in an information-theoretic sense again). In this way we arrive at the idea of a "matrix" phase space composed of M_x (spatial) by M_t (temporal) *matrices* formed from space and time delays of the observations $a(x,t)$:

$$\begin{bmatrix} a(i,n) & \cdots & a(i,n+(M_t-1)T) \\ a(i+D,n) & \cdots & a(i+D,n+(M_t-1)T) \\ \vdots & \vdots & \vdots \\ a(i+(M_x-1)D,n) & \cdots & a(i+(M_x-1)D,n+(M_t-1)T) \end{bmatrix},$$

where $a(i,n) = a(x,t)$ at $a(x_0+iD, t_0+nT)$, and the time delay T is an integer multiple of the sampling time τ_s, while the space delay is an integer multiple of the spatial sampling Δ_s. If we have N_t data points in the temporal direction $n=1,2,\ldots,N_t$ and N_x points in the space direction $i=1,2,\ldots,N_x$, then the total number of matrices available to populate the space is $(N_t-M_t)\times(N_x-M_x)$. It will be usual to have $N_x \gg M_x$ and $N_t \gg M_t$.

To determine the time delay we use average mutual information in the familiar fashion. To determine the space delay D, we do the same, but now we must compute the average mutual information for measurements $a(n,i)$ and measurements $a(n,i+D)$ averaged over all i and n to cover the attractor. Thus, it requires just a slight generalization of existing algorithms for temporal chaos.

To determine the dimension of our "matrix" phase space $M_t M_x$, we need to establish values for the minimum number of coordinates needed to unfold the attractor observed through the projection onto our measured $a(i,n)$. For this we can use the method of false nearest neighbors (Kennel *et al.*, 1992), which works efficiently in the case of temporal chaos alone.

The result of this analysis will be the $M_t \times M_z$ dimensional state space within which we can work to determine the same kind of quantity of interest as in temporal chaos: dimensions, Lyapunov exponents, models, etc.

We denote the members of our state space by $A_{\alpha,\beta}(i,n)$, where the ranges of α and β are $\alpha=1,2,\ldots,M_t$ and $\beta=1,2,\ldots,M_x$ while the "labels" (i,n) run over $i=1,2,\ldots,N_x-M_x$ and $n=1,2,\ldots,N_t-M_t$.

Essentially the same approach can be employed if the object of interest is a "snapshot" of a dynamical process with two spatial dimensions, namely, an image. We need only change the "labels" from x,t to x,y and the variables to $a(x,y)=a(x_0+i\Delta_x, y_0+j\Delta_y)$ with space delays Δ_x,Δ_y. We would establish two embedding dimensions M_x,M_y, and the remainder of the general discussion would go as before. The resulting modeling of images as nonlinear processes in the $M_x M_y$-dimensional space of matrices will then replace the present "state of the art" of linear modeling (Biemond *et al.*, 1990).

To determine quantities such as dimensions of the state-space attractor—that is, the dimension of the collection of matrices located in the $M_t \times M_x$-dimensional space—we must address the question of the *distance* between two rectangular matrices. This question has been studied in some detail by mathematicians. The most interesting work is reported in the monograph by Horn and Johnson (1985), where the idea of *matrix norms* is explored. This is summarized in the book by Golub and Van Loan (1989), where the needed computations are also discussed. The idea is that one can always read a matrix along its rows and then along its columns to consider it a *vector* of length $L=M_t M_x$, then define the usu-

al distances between vectors in that L-dimensional space. This approach has been used by Afraimovich *et al.* (1992) for the analysis of 2D snapshots of Faraday ripples. They calculated the correlation dimension of the patterns at different stages of disorder corresponding to different pumping magnitude. It should be noted, however, that matrices have more structure than just long vectors; in particular, they can be multiplied in a natural way, and thus there are new alternatives for the norms of such objects.

The norm that seems much more relevant for the present purposes is the so-called *spectral norm*. This norm follows from the idea of a singular-value decomposition of a rectangular matrix. Suppose we have a matrix **A** which is $m \times n$; this can be decomposed as

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T , \tag{144}$$

where the $m \times m$ matrix **U** and the $n \times n$ matrix **V** are orthogonal:

$$\mathbf{U}^T \cdot \mathbf{U} = \mathbf{I}, \quad \mathbf{V}^T \cdot \mathbf{V} = \mathbf{I} , \tag{145}$$

and, taking $m \geq n$, $\mathbf{\Sigma}$ is the $n \times n$ diagonal matrix

$$\mathbf{\Sigma} = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n] . \tag{146}$$

The σ_i are known as the singular values of the matrix **A**. They are also the square root of the eigenvalues of the $n \times n$ matrix $\mathbf{A}^T \cdot \mathbf{A}$. We order them as is conventional by $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

The relevance of the singular values to our considerations is that the norm of a matrix defined by

$$\| \mathbf{A} \| = \max_{\| \mathbf{x} \| = 1} \| \mathbf{A} \cdot \mathbf{x} \| \tag{147}$$

is just the largest singular value, $s_1(\mathbf{A})$. Here **x** is an n vector and

$$\| \mathbf{x} \| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2} , \tag{148}$$

the Euclidean distance in our n space. If $n = 1$, the matrix is $\mathbf{A} = (a_1, a_2, \ldots, a_m)$ and

$$s_1(\mathbf{A}) \propto \sqrt{a_1^2 + a_2^2 + \cdots + a_m^2} ,$$

which is a familiar norm.

Once one has a way of finding the appropriate space for the discussion of spatio-temporal data, the various aspects of the analysis we have discussed in previous sections become ready for use in this enlarged context. We consider these comments as suggestive of how the general subject will unfold as spatio-temporal data are studied from the nonlinear dynamics viewpoint. It remains to do the hard work of carrying out such analyses on models and measured data to see what general classification schemes and tools for physics can be extracted.

XI. CONCLUSIONS

A. Summary

Over the past decade we find an enormous literature on the analysis of temporal chaos and correspondingly

significant progress for purposes of studying chaotic laboratory experiments and chaotic field observations of interest to physicists. In this review we have tried to emphasize tools that are of direct utility in understanding the physical mechanisms at work in these measurements. We have placed less emphasis on mathematical results, using them as a guide to what one can learn about fundamental physical processes. Numerical results are critical to the study of nonlinear systems that have chaotic behavior. Indeed, computation plays a larger role in such studies than is traditional in many parts of the physics literature. Progress such as that reported throughout this review rests heavily on the ability to compute accurately and rapidly, and as such would often not have been possible a decade ago. The subject reviewed here is almost "experimental" in that sense through its reliance on computers as an instrument for its study. This bodes well for further analysis of chaos and its physical manifestations, since one can expect even more powerful computing tools to be widely available on a continuing basis.

The point of view taken throughout this review, often implicitly, is that chaotic behavior is a fundamental and interesting property of nonlinear physical systems. This point of view suggests that as a basic property of the nonlinear equations governing real physical systems there is much new to be learned by the study of chaos. While at times we have reported on efforts to "tame" that chaos, as in the section on control of periodic behavior, our main goal has been to bring to physicists a set of tools for extracting important information on the underlying physics of chaotic observations. Absent many of the tools created by those whose contributions we have reviewed here, much of the physics of chaos has been passed by as "noise." We expect the development of these tools to continue and the exploitation of the physics thus exposed to blossom in applications across the spectrum of phenomena where chaos is found. Indeed, this spectrum is so vast that our review has been inevitably unable to cite fully even a fraction of the literature on laboratory and field observations where chaos is known to occur. In any case, such a full recitation might be perceived as a litany of all of classical and semiclassical physics, and thus slightly useless.

The critical aspect of chaotic behavior from the physicist's view is that it is broadband spectrally, yet structured in an appropriate phase space, which itself can be reconstructed from data. This distinguishes it from the conventional view of "noise" and opens up for review and reexamination a wide variety of work which presumes unstructured stochastic behavior to be a driver or major influence on the physical processes being studied. That reexamination has not been undertaken in this review, nor is it a finished project to be found in the literature. Our review has focused on the tools one will use or the ideas underlying the tools, which will evolve for the extraction of new physics from observations of chaotic physics. We have no doubt whatsoever that where physics has been hidden by the assumption of un-

structured stochastic forces, an understanding that there may well be structure there after all will provide exciting new discoveries.

In a broad sense, what we have covered in this review might be cast as "signal processing" in another context. It seems to us substantially appropriate for physicists to have a major role in defining what constitutes signal processing for signals arising from physical systems. We expect that many of the developments over the coming decade in this field will come from the more traditional signal-processing community. This is borne out by a careful look at our references and a glance at the issues now developing in the analysis of temporal chaos. Further, while we have touched upon it only slightly, we also expect that from the same community we shall see an effort in "signal synthesis" in addition to the kind of "signal analysis" that has been our emphasis.

This review is in many ways an update of the earlier review by Eckmann and Ruelle (1985) about eight years ago. Clearly much has been achieved since then, and much rests on ideas carefully exposed in that previous review. It is impossible that we will have covered all the topics seen as important and interesting by all the physicists, mathematicians, chemists, engineers, and others who have contributed in the past eight years to the subject covered here. We offer our apologies to those whose work has only indirectly been cited. We do not deem that as a signal of its lack of importance, but eventually the finiteness of our task prevailed. We have appended a set of additional references in the hopes of identifying some of the omissions in the main text. In any case, we have made no attempt to survey the literature outside physics and what is now called "nonlinear science" for this review. This has excluded interesting work in economics, statistics, chemistry, engineering, biological sciences, and medicine—to name just a few important areas. This is a choice we have made, and we acknowledge the many interesting contributions that we have chosen not to assess or include.

B. Cloudy crystal ball gazing

The study of temporal chaos—the analysis of physical measurements at a point in space—is by now rather thoroughly explored. The various items reviewed try to bring into focus the methods now available in a practical sense for the study of basic properties of nonlinear physical systems in purely temporal measurements. We are certain that new developments in many of the areas we have touched upon and in areas we have not yet even imagined will continue to occur before the next review of this general area. We expect numerous algorithmic improvements to be created in each of the areas we have touched upon. We hope there will be a significant push in the formal and rigorous statistical analysis of many of the algorithms already available. It would be enormously useful, for example, in the analysis of data to have some estimate of the error in all conclusions arising from finite

data sets as well as from numerical, experimental, or instrumental uncertainty. It would be extremely important to extend the full power of the existing topological analysis (Mindlin *et al.*, 1991) in three dimensions to higher dimensions. We anticipate extensive work on the analysis of geometric properties of experimental strange attractors. These may also be more robust against contamination by unwanted signals than the more familiar derived or metric quantities such as dimensions and Lyapunov exponents.

We look forward to the application of chaotic analysis tools, both as covered here and in their further developments, to the detailed study of numerous physical systems. The few examples we have presented here are clearly the "tip of the iceberg." It is in the actual making of these applications where the push for further progress is sure to come most persuasively. The part of the task for the physicist which has been little touched on in this review, since there is little to review on it as a matter of fact, is the building and verification of models that arise from considerations of chaotic physical experiments. We took the liberty to say a few general words about this in the Introduction, but the "proof of the pudding" is in performing the task itself. It is clear, and even mathematically good sense (Rissanen, 1989), that we cannot algorithmically discover fundamental physics. For the physicist the road ahead with the subject of chaotic data analysis will differ from signal processing and return to the analysis of interesting physical systems with the methods reviewed. The point of view in the creation of physical models will change, of course. It will not be necessary, we imagine, always to start with the Navier-Stokes equations to build models of chaos in fluids, though, perhaps, we will learn to project those infinite-dimensional equations down to the few relevant physical degrees of freedom, which will appear in the models.

The discovery of temporal chaos in physical systems is a "finished" topic of research, we believe. Many physical systems have been shown to exhibit chaotic orbits, and we are certain many more will be found. It is no longer enough, however, to have good chaos, one must move on and extract good physics out of it. The methods and techniques in this review are directed towards that in the realm of temporal chaos. An important challenge for the area of research we have reviewed here is to extend both the existing methods and the classification of widely expected and universal behaviors to spatio-temporal chaos.

This is to say that the study of *chaotic fields* remains rather untouched from the point of view taken throughout this review. We certainly do not mean that early and productive steps have not been taken; they have! We simply desire to underline the clear fact that the developments in the analysis of spatio-temporal chaotic data and the connection with model building, signal separation, and the other goals we have outlined are not yet as far along as for temporal chaos. The recent thorough review of Cross and Hohenberg (1993) makes it quite clear that the predominant body of work in this

area has concentrated on the patterns and behaviors coming from identifiable linear instabilities in spatio-temporal dynamical systems. The challenge of moving the study of experimentally observed chaotic fields forward is a large one. Just as much of physics is nonlinear and has postured behind easier problems posed by local linearization for many years, so have we focused on point measurements or "lumped parameter" models when studying a remarkable phenomenon like chaos in physics. The physics of chaotic fields will bring us from this too narrow focus to substantial problems encountered in the laboratory and in geophysical field observations. We expect the next review of this sort to be able to report major accomplishments in the analysis of experimental chaotic fields, and we expect the lessons we have tried to report here to pervade the way we proceed to those accomplishments.

ACKNOWLEDGMENTS

REFERENCES

Abarbanel, H. D. I., R. Brown, and J. B. Kadtke, 1989, Phys. Rev. A **41**, 1782.

Abarbanel, H. D. I., R. Brown, and M. B. Kennel, 1991, J. Nonlinear Sci. **1**, 175.

Abarbanel, H. D. I., R. Brown, and M. B. Kennel, 1992, J. Nonlinear Sci. **2**, 343.

Abarbanel, H. D. I., S. M. Hammel, P.-F. Marteau, and J. J. ("Sid") Sidorowich, 1992, "Scaled Linear Cleaning of Contaminated Chaotic Orbits," preprint, University of California, San Diego, INLS.

Abarbanel, H. D. I., and M. M. Sushchik, 1993, Int. J. Bif. Chaos **3**, 543.

Abraham, N. D., A. M. Albano, B. Das, G. DeGuzman, S. Yong, K. S. Gioggia, G. P. Puccioni, and J. R. Tredicce, 1986, Phys. Lett. A **114**, 217.

Afraimovich, V. S., A. B. Ezersky, M. I. Rabinovich, M. A. Shereshevsky, and A. L. Zheleznyak, 1992, Physica D **15**, 331.

Afraimovich, V. S., N. N. Verichev, and M. I. Rabinovich, 1986, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **29**, 1050.

Andereck, C. D., and F. Hayot, 1992, Eds., *Ordered and Turbulent Patterns in Taylor-Couette Flow* (Plenum, New York).

Arnol'd, V. I., 1978, *Mathematical Methods of Classical Mechanics*, translated by K. Vogtmann and A. Weinstein (Springer, New York).

Auerbach, D., P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, 1987, Phys. Rev. Lett. **58**, 2387.

Badii, R., G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M. A. Rubio, 1988, Phys. Rev. Lett. **60**, 979.

Badii, R., and A. Politi, 1985, J. Stat. Phys. **40**, 725.

Bau, H. H., and J. Singer, 1992, in *Proceedings of the 1st Experimental Chaos Conference*, edited by S. Vohra, M. Spano, M. Shlesinger, L. Pecora, and W. Ditto (World Scientific, Singapore), p. 145.

Beck, C., 1990, Physica D **41**, 67.

Benettin, G., C. Froeschle, and J.-P. Scheidecker, 1979, Phys. Rev. A **19**, 2454.

Benettin, G., L. Galgani, A. Giorgilli, and J.-M. Strelcyn, 1980, Meccanica **15**, 9 (Part I); **15**, 21 (Part II).

Benettin, G., L. Galgani, and J.-M. Strelcyn, 1976, Phys. Rev. A **14**, 2338.

Biemond, J., R. L. Lagendijk, and R. M. Mersereau, 1990, Proc. IEEE **78**, 856.

Birman, J. S., and R. F. Williams, 1983a, Topology **22**, 47.

Birman, J. S., and R. F. Williams, 1983b, Contemp. Math. **20**, 1.

Bradley, E., 1991, in *Algebraic Computation in Control*, Lecture Notes in Control and Information Sciences, edited by G. Jacob and F. Lamnabhi-Lagarrigue (Springer, Berlin), p. 307.

Bradley, E., 1992, in *Proceedings of the 1st Experimental Chaos Conference*, edited by S. Vohra, M. Spano, M. Shlesinger, L. Pecora, and W. Ditto (World Scientific, Singapore), p. 152.

Briggs, K., 1990, Phys. Lett. A **151**, 27.

Broomhead, D. S., and R. Jones, 1989, Proc. R. Soc. London, Series A **423**, 103.

Broomhead, D. S., and G. P. King, 1986, Physica D **20**, 217.

Brown, R., 1992, "Orthonormal Polynomials as Prediction Functions in Arbitrary Phase Space Dimensions," preprint, University of California, San Diego, INLS.

Brown, R., P. Bryant, and H. D. I. Abarbanel, 1991, Phys. Rev. A **43**, 2787.

Bryant, P., R. Brown, and H. D. I. Abarbanel, 1990, Phys. Rev. Lett. **65**, 1523.

Carroll, T. L., and L. M. Pecora, 1991, IEEE Trans. Circuits Syst. **38**, 453.

Carroll, T. L., and L. M. Pecora, 1992, "Cascading and Controlling Synchronized Chaotic Systems" preprint, Naval Research Laboratory, Washington, D.C.

Casdagli, M., 1989, Physica D **35**, 335.

Casdagli, M., D. Des Jardins, S. Eubank, J. D. Farmer, J. Gib-

son, N. Hunter, and J. Theiler, 1991, "Nonlinear Modeling of Chaotic Time Series: Theory and Application," Preprint, Los Alamos National Laboratory No. LA-UR-91-1637.

Cawley, R., and G.-H. Hsu, 1992, "Local-Geometric-Projection Method for Noise Reduction in Chaotic Maps and Flows," Phys. Rev. A **46**, 3057.

Chandrasekhar, S., 1961, *Hydrodynamic and Hydromagnetic Stability* (Clarendon, Oxford).

Chatterjee, S., and A. S. Hadi, 1988, *Sensitivity Analysis in Linear Regression* (Wiley, New York).

Chennaoui, A., K. Pawelzik, W. Liebert, H. G. Schuster, and G. Pfister, 1990, Phys. Rev. A **41**, 4151.

Coffman, K. G., W. D. McCormick, Z. Noszticzius, R. H. Simoyi, and H. L. Swinney, 1987, J. Chem. Phys. **86**, 119.

Cowan, J. D., and D. H. Sharp, 1988, Q. Rev. Biophysics **2**, 365.

Crawford, J. D., 1991, Rev. Mod. Phys. **63**, 991.

Crawford, J. D., and E. Knobloch, 1991, Annu. Rev. Fluid Dynamics **23**, 341.

Cross, M. C., and P. C. Hohenberg, 1993, Rev. Mod. Phys. **65**, 851.

Cremers, A. J., and A. Hübler, 1986, Z. Naturforsch A **42**, 797.

Cvitanović, P., 1988, Phys. Rev. Lett. **61**, 2729.

Devany, R., and Z. Nitecki, 1979, Commun. Math. Phys. **67**, 137.

Ditto, W. L., S. N. Rauseo, and M. L. Spano, 1990, Phys. Rev. Lett. **65**, 3211.

Eckmann, J.-P., S. O. Kamphorst, D. Ruelle, and S. Ciliberto, 1986, Phys. Rev. A **34**, 4971.

Eckmann, J.-P., and D. Ruelle, 1985, Rev. Mod. Phys. **57**, 617.

Eisenhammer, T., A. Hübler, N. Packard, J. A. S. Kelso, 1991, Bio. Cybernetics **65**, 107.

Ellner, S., A. R. Gallant, D. McGaffrey, and D. Nychka, 1991, Phys. Lett. A **153**, 357.

Essex, C., and Nerenberg, M. A. H., 1991, Proc. R. Soc. London, Series A **435**, 287.

Farmer, J. D., E. Ott, and J. A. Yorke, 1983, Physica D **7**, 153.

Farmer, J. D., and J. J. ("Sid") Sidorowich, 1987, Phys. Rev. Lett. **59**, 845.

Farmer, J. D., and J. J. ("Sid") Sidorowich, 1988, in *Evolution, Learning and Cognition,* edited by Y.-C. Lee (World Scientific, Singapore), p. 277.

Farmer, J. D., and J. J. ("Sid") Sidorwich, 1991, Physica D **47**, 373.

Flepp, L., R. Holzner, E. Brun, M. Finardi, and R. Badii, 1996, Phys. Rev. Lett. **67**, 2244.

Ford, J. R., and W. R. Borland, W. R., 1988, *Common Los Alamos Mathematical Software Compendium,* Document CIC No. 148 (Los Alamos National Laboratory, Los Alamos, New Mexico).

Fraser, A. M., 1988, Ph.D. thesis (University of Texas, Austin).

Fraser, A. M., 1989a, IEEE Trans. Inf. Theory **35**, 245.

Fraser, A. M., 1989b, Physica D **34**, 391.

Fraser, A. M., and H. L. Swinney, 1986, Phys. Rev. A **33**, 1134.

Frederickson, P., J. L. Kaplan, E. D. Yorke, and J. A. Yorke, 1983, J. Diff. Eq. **49**, 185.

Gallager, R. G., 1968, *Information Theory and Reliable Communication* (Wiley, New York).

Gilmore, C., 1992, J. Econ. Behav. Organization (in press).

Giona, M., F. Lentini, and V. Cimagalli, 1991, Phys. Rev. A **44**, 3496.

Goldberg, D. E., 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA).

Golub, G. H., and C. F. Van Loan, 1989, *Matrix Computations,* 2nd ed. (Johns Hopkins, Baltimore, MD).

Gollub, J. P., and R. Ramshankar, 1991, in *New Perspectives in Turbulence,* edited by L. Sirovich (Springer, New York), p. 165.

Grassberger, P., and I. Procaccia, 1983a, Phys. Rev. Lett. **50**, 346.

Grassberger, P., and I. Procaccia, 1983b, Physica D **9**, 189.

Grassberger, P., T. Schrieber, and C. Schaffrath, 1991, Int. J. Bif. Chaos **1**, 521.

Greene, J. M., and J.-S. Kim, 1986, Physica D **24**, 213.

Guckenheimer, J., and P. Holmes, 1983, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York).

Hammel, S. M., 1990, Phys. Lett. A **148**, 421.

Hammel, S. M., C. K. R. T. Jones, and J. V. Moloney, 1985, J. Opt. Soc. Am. B **2**, 552.

Hao, Bai-Lin, 1984, *Chaos* (World Scientific, Singapore).

Hao, Bai-Lin, 1990, *Chaos II* (World Scientific, Singapore).

Hénon, M., 1976, Commun. Math. Phys. **50**, 69.

Horn, R. A., and C. R. Johnson, 1985, *Matrix Analysis* (Cambridge University Press, New York).

Hsu, C. S., 1987, *Cell to Cell Mapping,* Applied Mathematical Sciences No. 64 (Springer, New York).

Ikeda, K., 1979, Opt. Commun. **30**, 257.

Isabelle, S. H., A. V. Oppenheim, and G. W. Wornell, 1992, in *1992 International Conference on Acoustics, Speech, and Signal Processing,* IEEE Signal Processing Society Volume 4 (IEEE, New Jersey), p. IV-133.

Kaplan, D. T., and L. Glass, 1992, Phys. Rev. Lett. **68**, 427.

Kaplan, J. L., and J. A. Yorke, 1979, in *Functional Differential Equations and Approximations of Fixed Points,* Lecture Notes in Mathematics No. 730, edited by H.-O. Peitgen and H-O. Walther (Springer, Berlin), p. 228.

Keeler, J. D., and J. D. Farmer, 1986, Physica D **23**, 413.

Kennel, M. B., R. Brown, and H. D. I. Abarbanel, 1992, Phys. Rev. A **45**, 3403.

Kolmogorov, A. N., 1958, Dokl. Akad. Nauk. SSSR **119**, 861 [Math. Rev. **21**, 386 (1960)].

Kostelich, E. J., and J. A. Yorke, 1991, Physica D **41**, 183.

Kowalski, J. M., G. L. Albert, and G. W. Gross, 1990, Phys. Rev. A **42**, 6260.

Kuramoto, Y., 1984, *Chemical Oscillations, Waves and Turbulence,* Springer Series in Synergetics Vol. 19 (Springer, New York).

Landa, P. S., and M. G. Rozenblyum (Rosenblum), 1989, Zh. Tekh. Fiz. **59**, 13 [Sov. Phys. Tech. Phys. **34**, 6].

Lapedes, A., and R. Farber, 1987, "Nonlinear signal processing using neutral networks: Prediction and signal modeling," Preprint, Los Alamos National Laboratory No. LA-UR-87-2662.

Liebert, W., and H. G. Schuster, 1989, Phys. Lett. A **142**, 107.

Lorenz, E. N., 1963, J. Atmos. Sci. **20**, 130.

Lorenz, E. N., 1969, J. Atmos. Sci. **26**, 636.

MacKay, R. S., 1992, "Hyperbolic Structure in Classical Chaos," University of Warwick Scientific Preprint, 4/1993. This paper was presented as a review Lecture at the *International Summer School on Quantum Chaos,* Enrico Fermi Institute, Varenna, Italy, 1992.

Mañé, R., 1981, in *Dynamical Systems and Turbulence, Warwick, 1980,* edited by D. Rand and L.-S. Young, Lecture Notes in Mathematics No. 898 (Springer, Berlin), p. 230.

Manneville, P., 1990, *Dissipative Structures and Weak Turbulence* (Academic, Boston).

Marteau, P.-F., and H. D. I. Abarbanel, 1991, J. Nonlinear Sci. **1**, 313.

McGaffrey, D. F., S. Ellner, A. R. Gallant, and D. W. Nychka,

1992, J. Am. Stat. Assoc. **87**, 682.

Melvin, P., and N. B. Tufillaro, 1991, Phys. Rev. A **44**, R3419.

Mindlin, G. B., X.-J. Hou, H. G. Solari, R. Gilmore, and N. B. Tufillaro, 1990, Phys. Rev. Lett. **64**, 2350.

Mindlin, G. B., H. G. Solari, M. A. Natiello, R. Gilmore, and X.-J. Hou, 1991, J. Nonlinear Sci. **1**, 147.

Mitschke, F., M. Möller, and W. Lange, 1988, Phys. Rev. A **37**, 4518.

Oppenheim, A. V., and R. W. Schafer, 1989, *Discrete-Time Signal Processing* (Prentice Hall, Englewood Cliffs, NJ).

Oppenheim, A. V., G. W. Wornell, S. H. Isabelle, and K. M. Cuomo, 1992, in *International Conference on Acoustics, Speech, and Signal Processing,* IEEE Signal Processing Society Vol. 4 (IEEE, New Jersey), p. IV-117.

Orayevsky, , *et al.,* 1965, *Molecular Oscillators* (Nauka, Moscow).

Osborne, A. R., and A. Provenzale, 1989, Physica D **35**, 357.

Oseledec, V. I., 1968, Trudy. Mosk. Mat. Obsc. **19**, 197 [Moscow Math. Soc. **19**, 197 (1968)].

Ott, E., C. Grebogi, and J. A. Yorke, 1990, Phys. Rev. Lett. **64**, 1196.

Packard, N. H., 1990, Complex Systems **4**, 543.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, 1980, Phys. Rev. Lett. **45**, 712.

Paladin, G., and A. Vulpiani, 1987, Phys. Rep. **156**, 147.

Papoff, F., A. Fioretti, E. Arimondo, G. B. Mindlin, H. Solari, and R. Gilmore, 1992, Phys. Rev. Lett. **68**, 1128.

Parlitz, U., 1992, Int. J. Bif. Chaos **2**, 155.

Pecora, L. M., and T. L. Carroll, 1990, Phys. Rev. Lett. **64**, 821.

Pecora, L. M., and T. L. Carroll, 1991a, Phys. Rev. Lett. **67**, 945.

Pecora, L. M., and T. L. Carroll, 1991b, Phys. Rev. A **44**, 2374.

Pesin, Y. B., 1977, Usp. Mat. Nauk. **32**, 55 [Russ. Math. Sur. **32**, 55 (1977)].

Pikovsky, A. S., 1986, Radio Eng. Electron. Phys. **9**, 81.

Powell, M. J. D., 1981, *Approximation Theory and Methods* (Cambridge University Press, Cambridge).

Powell, M. J. D., 1985, "Radial basis functions for multivariate interpolation: a review" preprint, University of Cambridge.

Powell, M. J. D., 1987, "Radial basis function approximation to polynomials" preprint, Univ. of Cambridge.

Rabinovich, M. I., 1978, Usp. Fiz. Nauk. **125**, 123 [Sov. Phys.-Usp. **21**, 443 (1979)].

Rissanen, J., 1989, *Stochastic Complexity in Statistical Inquiry* (World Scientific, Singapore).

Roux, J.-C., R. H. Simoyi, and H. L. Swinney, 1983, Physica D **8**, 257.

Ruelle, D., "Ergodic Theory of Differentiable Dynamical Systems," 1979, Publications Mathématiques de l'Institut des Hautes Études Scientifiques, No. 50, p. 27.

Ruelle, D., 1990, Proc. R. Soc. London, Series A **427**, 241.

Ruelle, J.-P., 1978, Bol. Soc. Bras. Mat. **9**, 83.

Sakai, H., and H. Tokumaru, 1980, IEEE Trans. AASP **28**, 588.

Salzmann, B., 1962, J. Atmos. Sci. **19**, 329.

Sano, M., and Y. Sawada, 1985, Phys. Rev. Lett. **55**, 1082.

Sauer, T., 1992, Physica D **58**, 193.

Sauer, T., J. A. Yorke, and M. Casdagli, 1991, J. Stat. Phys. **65**, 579.

Schreiber, T., and P. Grassberger, 1991, Phys. Lett. A **160**, 411.

Shaw, R. Z., 1981, Naturforsch A **36**, 80.

Shaw, R., 1984, *The Dripping Faucet as a Model Dynamical System* (Aerial Press, Santa Cruz).

Shimada, I., and T. Nagashima, 1979, Prog. Theor. Phys. **61**, 1605.

Shinbrot, T., W. Ditto, C. Grebogi, E. Ott, M. Spano, and J. A. Yorke, 1992, Phys. Rev. Lett. **68**, 2863.

Shinbrot, T., E. Ott, C. Grebogi, and J. A. Yorke, 1990, Phys. Rev. Lett. **65**, 3215.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London).

Sinai, Y., 1959, Dokl. Akad. Nauk. SSSR **124**, 768 [Math. Rev. **21**, 386 (1960)].

Singer, J., and H. H. Bau, 1991, Phys. Fluids A **3**, 2859.

Singer, J., Y.-Z. Wang, and H. H. Bau, 1991, Phys. Rev. Lett. **66**, 1123.

Smith, L. A., 1988, Phys. Lett. A **133**, 283.

Stoer, J., and R. Burlisch, 1980, *Introduction to Numerical Analysis,* translated by R. Bartels, W. Gautschi, and C. Witzgall (Springer, New York).

Takens, F., 1981, in *Dynamical Systems and Turbulence, Warwick, 1980,* edited by D. Rand and L.-S. Young, Lecture Notes in Mathematics No. 898 (Springer, Berlin), p. 366.

Theiler, J., 1990, J. Opt. Soc. Am. A **7**, 1055.

Theiler, J., 1991, Phys. Lett. A **155**, 480.

Tufillaro, N. B., R. Holzner, L. Flepp, E. Brun, M. Finardi, and R. Badii, 1991, Phys. Rev. A **44**, R4786.

Van Trees, H. L., 1968, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York).

Whitney, H., 1936, Ann. Math. **37**, 645.

Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano, 1985, Physica D **16**, 285.

ADDITIONAL REFERENCES

One of the challenges in a review such as this one is making it finite. In rising to this challenge we have had to make a selection of topics and to identify the appropriate literature that bears quite directly on those topics. We have, of course, done this, and thus risked ignoring vast parts of the vast literature on the analysis of time series from chaotic observations. The items cited in the main body of the text by no means represent the same selection many of our colleagues would have made, but they do represent the literature we consulted and learned from in putting together this review. Many other papers and books have contributed to our understanding of the material we are covering in this review, and we attempt to identify below the main body of these. As ever, we are pleased to hear from authors whose work we have inadvertently failed to cite. We shall be happy to read these papers and we shall happily cite that work as well.

Abarbanel, H. D. I., R. Brown, and M. B. Kennel, 1991, "Lyapunov Exponents in Chaotic Systems: Their Importance and Their Evaluation using Observed Data," Int. J. Mod. Phys. B **5**, 1347.

Abarbanel, H. D. I., S. Koonin, H. Levine, G. J. F. MacDonald, and O. Rothaus, 1990a, "Statistics of Extreme Events with Application to Climate," JASON/MITRE Report JSR-90-305.

Abarbanel, H. D. I., S. Koonin, H. Levine, G. J. F. MacDonald, and O. Rothaus, 1990b, "Issues in Predictability," JASON/MITRE Report JSR-90-320.

Afraimovich, V. S., I. S. Aranson, and M. I. Rabinovich, 1989, "Multidimensional Strange Attractors and Turbulence," Sov. Sci. C Math. Phys. **8**, 3.

Afraimovich, V. S., M. I. Rabinovich, and V. I. Sbitnev, 1985, "Attractor dimension in a chain of coupled oscillators," Sov. Tech. Phys. Lett. **11**, 139.

Albano, A. M., J. Muench, C. Schwartz, A. Mees, and P. Rapp,

1988, "Singular Value Decomposition and the Grassberger-Procaccia Algorithm," Phys. Rev. A **38**, 3017.

Aranson, I. S., M. I. Rabinovich, A. M. Reyman, V. G. Shekhov, 1988, "Measurement of Correlation Dimension in Experimental Devices," Preprint No. 215, Institute of Applied Physics, Gorky.

Baake, E., M. Baake, H. G. Bock, and K. M. Briggs, 1992, "Fitting Ordinary Differential Equations to Chaotic Data," Phys. Rev. A **45**, 5524.

Badii, R., G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M. A. Rubio, 1988, "Dimension Increase in Filtered Chaotic Signals," Phys. Rev. Lett. **60**, 979.

Baker, G. L., and J. P. Gollub, 1990, *Chaotic Dynamics: An Introduction* (Cambridge University Press, New York).

Bayly, B. J., I. Goldhirsch, and S. A. Orzag, 1987, "Independent degrees of freedom of dynamical systems," J. Scientific Computing **2**, 111.

Benzi, R., and G. F. Carnevale, 1989, "A Possible Measure of Local Predictability," J. Atmos. Sci. **46**, 3595.

Bingham, S., and M. Kot, 1989, "Multidimensional Trees, Range Searching, and a Correlation Dimension Algorithm of Reduced Complexity," Phys. Lett. A **140**, 327.

Buzug, Th., T. Reimers, and G. Pfister, 1990, "Optimal Reconstruction of Strange Attractors from Purely Geometrical Arguments," Europhys. Lett. **13**, 605.

Cenys, A., and K. Pyragas, 1988, "Estimation of the Number of Degrees of Freedom from Chaotic Time Series," Phys. Lett. A, **129**, 227.

Cenys, A., G. Lasiene, and K. Pyragas, 1991, "Estimation of Interrelations between Chaotic Observables," Physica D **52**, 332.

Ciliberto, S. and B. Nicolaenko, 1991, "Estimating the Number of Degrees of Freedom in Spatially Extended Systems," Europhys. Lett. **14**, 303.

Crutchfield, J. P., and B. S. MacNamara, 1987, Equations of motion from a data series. Complex Systems **1**, 417.

Ellner, S., 1988, "Estimating Attractor Dimensions from Limited Data: A New Method, with Error Estimates," Phys. Lett. A **133**, 128.

Farmer, J. D., E. Ott, and J. A. Yorke, 1983, "The Dimension of Chaotic Attractors," Physica D **7**, 153.

Farmer, J. D., 1982, "Chaotic Attractors of an Infinite-Dimensional Dynamical System," Physica D **4**, 366.

Farmer, J. D., 1990, "Rosetta stone for connectionism," Physica D **42**, 153.

Fraedrich, K., 1986, "Estimating the Dimensions of Weather and Climate Attractors," J. Atmos. Sci. **43**, 419.

Frederickson, P., J. L. Kaplan, E. D. Yorke, and J. A. Yorke, 1983, "The Liapunov Dimension of Strange Attractors," J. Diff. Eq. **49**, 185.

Friedman, J. H., J. L. Bentley, and R. A. Finkel, 1977, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," ACM Trans. Math. Software **3**, 209.

Fujisaka, H., 1983, "Statistical Dynamics Generated by Fluctuations of Local Lyapunov Exponents," Prog. Theor. Phys. **70**, 1274.

Geist, K., U. Parlitz, and W. Lauterborn, 1990, "Comparison of Different Methods for Computing Lyapunov Exponents," Prog. Theor. Phys. **83**, 875.

Goldhirsch, I., P-L. Sulem, and S. A. Orzag, 1986, "Stability and Lyapunov Stability of Dynamical Systems: a Differential Approach and Numerical Method," Physica D **27**, 311.

Grassberger, P., R. Badii, and A. Politi, 1988, "Scaling Laws for Invariant Measures on Hyperbolic and Nonhyperbolic Attractors," J. Stat. Phys. **51**, 135.

Grebogi, C., E. Ott, and J. A. Yorke, 1988, "Unstable Periodic Orbits and the Dimensions of Multifractal Chaotic Attractors," Phys. Rev. A **37**, 1711.

Hou, X. J., R. Gilmore, G. B. Mindlin, and H. G. Solari, 1990, "An Efficient Algorithm for Fast $O(N \log(N))$ Box Counting," Phys. Lett. A **151**, 43.

Kaplan, J. L., and J. A. Yorke, 1979, "Chaotic Behavior in Multidimensional Difference Equations," Lect. Notes Math. **730**, 204.

Kostelich, E. J., and J. A. Yorke, 1988, "Noise reduction in dynamical systems," Phys. Rev. A **38**, 1649.

Landa, P. S., and M. Rosemblum, 1991, "Time Series Analysis for System Identification and Diagnostics," Physica D **48**, 232.

Liebert, W., K. Pawelzik, and H. G. Schuster, 1991, "Optimal Embeddings of Chaotic Attractors from Topological Considerations," Europhys. Lett. **14**, 521.

Lindenberg, K., and B. J. West, 1986, "The First, the Biggest, and Other Such Considerations," J. Stat. Phys. **42**, 201.

Lorenz, E. N., 1984, "Irregularity: A Fundamental Property of the Atmosphere," Tellus A **36**, 98.

Lukashchuk, S. N., G. E. Falkovich, and A. I. Chernykh, 1989, "Calculation of the dimensions of attractors from experimental data," J. Appl. Mech. Tech. Phys. **30**, 95.

Mackey, M., and L. Glass, 1977, "Oscillation and Chaos in Physiological Control Systems," Science **197**, 287.

Mayer-Kress, G., 1986, Ed., *Dimensions and Entropies in Chaotic Systems* (Springer, Berlin).

Mees, A. I., P. E. Rapp, and L. S. Jennings, 1987, "Singular-value Decomposition and Embedding Dimension," Phys. Rev. A **37**, 340.

Moon, Francis C., 1987, *Chaotic Vibrations* (Wiley, New York).

Myers, C., S. Kay, and M. D. Richard, 1992, "Signal Separation for Nonlinear Dynamical Systems," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE Signal Processing Society, Vol. 4 (IEEE, New Jersey), p. 129.

Nese, Jon M., 1989, "Quantifying Local Predictability in Phase Space," Physica D **35**, 237. Also see his PhD thesis entitled "Predictability of Weather and Climate in a Coupled-Ocean Atmosphere Model: A Dynamical Systems Approach," The Pennsylvania State University, Department of Meteorology, 1989, and references therein.

Osborne, A. R., and A. Provenzale, 1989, "Finite Correlation Dimension for Stochastic Systems with Power-Law Spectra," Physica D **35**, 357.

Parker, T. S., and L. O. Chua, 1990, *Practical Numerical Algorithms for Chaotic Systems* (Springer, New York).

Pawelzik, K., and H. G. Schuster, 1987, "Generalized Dimensions and Entropies from a Measured Time Series," Phys. Rev. A **35**, 481.

Pyragas, K., and A. Cenys, 1987, "Method for the Rapid Determination of the Number of Active Degrees of Freedom of Dynamic Systems with Chaotic Behavior," Litovskii Fizicheskii Sbornik **27**, 437.

Reyl, C., L. Flepp, R. Badii, and E. Brun, 1993, "Control of NMR-Laser Chaos in High-Dimensional Embedding Space," Phys. Rev. E **47**, 267.

Rössler, O. E., 1976, "An Equation for Continuous Chaos," Phys. Lett. A **57**, 397.

Russell, D. A., J. D. Hanson, and E. Ott, 1980, "Dimension of Strange Attractors," Phys. Rev. Lett. **45**, 1175.

Sato, S., M. Sano, and Y. Sawada, 1987, "Practical Methods of Measuring the Generalized Dimension and Largest Lyapunov Exponents in High-Dimensional Chaotic Systems," Prog. Theor. Phys. **77**.

Smith, L. A., 1988, "Intrinsic Limits on Dimension Calculations," Phys. Lett. A **133**, 283.

Stoop, R., and J. Parisi, 1991, "Calculation of Lyapunov Exponents Avoiding Spurious Elements," Physica D **50**, 89.

Theiler, J., 1986, "Spurious Dimension from Correlation Algorithms Applied to Limited Time Series Data," Phys. Rev. A **34**, 2427.

Theiler, J., 1987, "Efficient Algorithm for Estimating the Correlation Dimension from a Set of Discrete Points," Phys. Rev. A **36**, 4456.

Theiler, J., 1988, "Lacunarity in a Best Estimator of Fractal Dimension," Phys. Lett. A **133**, 195.

Theiler, J., 1990, "Statistical Precision of Dimension Estimators," Phys. Rev. A **41**, 3038.

Thompson, J. M. T., and H. B. Stewart, 1986, *Nonlinear Dynamics and Chaos* (Wiley, New York).

Tufillaro, N. B., H. G. Solari, and R. Gilmore, 1990, "Relative Rotation Rates—Fingerprints for Strange Attractors," Phys. Rev. A **41**, 5717.

Wales, D. J., 1991, "Calculating the Rate of Loss of Information from Chaotic Time Series by Forecasting," Nature **350**, 485.

Wright, Jon, 1984, "Method for Calculating a Lyapunov exponent," Phys. Rev. A **29**, 2924.

Young, L-S., 1983, "Entropy, Lyapunov Exponents, and Hausdorff Dimension in Differentiable Dynamical Systems," IEEE Trans. Circuits Systems C AS-30, 599.

Zeng, X., R. Eykholt, and R. A. Pielke, 1991, "Estimating the Lyapunov Exponent Spectrum from Short Time Series of Low Precision," Phys. Rev. Lett. **66**, 3229.