Noise reduction in chaotic time-series data: A survey of common methods

Eric J. Kostelich
Department of Mathematics, Box 871804, Arizona State University, Tempe, Arizona 85287

Thomas Schreiber*
Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark
(Received 23 April 1993)

This paper surveys some of the methods that have been suggested for reducing noise in time-series data whose underlying dynamical behavior can be characterized as low-dimensional chaos. Although the procedures differ in details, all of them must solve three basic problems: how to reconstruct an attractor from the data, how to approximate the dynamics in various regions on the attractor, and how to adjust the observations to satisfy better the approximations to the dynamics. All current noise-reduction methods have similar limitations, but the basic problems are reasonably well understood. The methods are an important tool in the experimentalist's repertoire for data analysis. In our view, they should be used more widely, particularly in studies of attractor dimension, Lyapunov exponents, prediction, and control.

PACS number(s): 05.45.+b

I. INTRODUCTION

Noise limits one's ability to extract quantitative information from time-varying signals. In many cases, one is interested in a time series of data produced by a system whose underlying behavior can be characterized as low-dimensional chaos. Two important measures associated with the dynamics are the dimension of the attractor (which describes approximately the number of degrees of freedom) and the Lyapunov exponents (which quantify the sensitivity of the process to initial conditions).

There is much literature on physical systems that exhibit chaotic behavior and on methods for estimating the dimension of an attractor reconstructed from time-series data (see, for instance, the papers in [1,2]). Algorithms for estimating Lyapunov exponents have been described by Wolf *et al.* [3] and by Eckmann *et al.* [4], among others.

Both the dimension and the Lyapunov exponents describe a kind of scaling behavior in the limit as the distances between points on the attractor approach zero [5]. Each of them is sensitive to the presence of small amounts of noise. Simple numerical experiments show that a noise level of 1% of the time-series extent makes it impossible to measure the correlation dimension of the attractor using distances less than 3% of the attractor extent [6]. Noise in laboratory data may completely obscure the underlying fractal structure unless the data are preprocessed to reduce the noise [7].

There are several aspects to the noise-reduction problem. One question concerns the circumstances under which low-pass filtering and similar methods are justified for the analysis of chaotic laboratory data. Another question involves ways in which the dynamics can be used to identify and correct errors in the observations arising from noise, at least in cases where the behavior of the process can be characterized as low-dimensional chaos. Different approaches have been suggested to the latter question. Examples include Kostelich and Yorke [6,8], Schreiber and Grassberger [9], Sauer [10], Cawley and Hsu [11], and Farmer and Sidorowich [12].

We do not want to attempt a quantitative comparison of these noise-reduction methods. Instead, it is our opinion that the methods have important similarities—indeed, one can view them as variations on a common theme. All of the methods must address three basic problems: (1) they must have a strategy for reconstructing the underlying attractor from the observed time series; (2) they must estimate the local dynamical behavior by choosing a class of models and fitting the parameters statistically; and (3) they must adjust the observations to make them more consistent with the models. For this reason, all of the methods have similar advantages and shortcomings. Nevertheless, the various methods perform well in many cases. In our opinion, experimentalists and others should take advantage of these ideas in their analysis of data.

In the next section, we summarize briefly one popular method for reconstructing an attractor from a time series of data and show how noise complicates the problem of computing the dimension of the attractor. We discuss improved embedding theorems in Sec. III and show how the effects of noise can be reduced with better embedding methods. In Sec. IV, we discuss how to approximate the dynamics using the data. We show how the observations can be adjusted to reduce noise in Sec. V and consider error measures in Sec. VI.

II. THE EFFECT OF NOISE: AN EXAMPLE USING TIME-DELAY EMBEDDINGS

The first step in any analysis of chaotic data is to reconstruct an attractor from the data. This is the *em-*

bedding problem, which can be summarized briefly as follows. Suppose one has an apparatus whose behavior can be described in principle by a set of differential equations. There might be a large number of equations, and one might not know exactly what they are. (For example, the Belousov-Zhabotinskii chemical reaction exhibits complex dynamics and more than 30 intermediate chemical species have been identified in it. However, many of the associated rate constants are not known [13,14].) After a time, the system reaches an asymptotic state in which the behavior is governed by the attractor A for the underlying set of equations. The experimental data are functions of points on A. The embedding problem is the following: given a time series of measurements $\{s_i\}$, how can a set be reconstructed that is equivalent in an appropriate mathematical sense to A? In particular, there should be a one-to-one correspondence between points on A and points on the reconstructed attractor, and information about the derivatives of the flow should be preserved.

The *Takens time-delay-embedding method* [15] is probably the most common attractor reconstruction method in the literature. (We discuss other strategies in Sec. III.) Given the time series $\{s_i\}$, the reconstructed attractor consists of the m vectors

$$\mathbf{x}_i = (s_i, s_{i+\tau}, s_{i+2\tau}, \ldots, s_{i+(m-1)\tau}) ,$$

where τ is the *time delay* and m is the *embedding dimension*. Takens showed that if the embedding dimension is sufficiently large, then the set $\{\mathbf{x}_i\}$ has the same dimension and Lyapunov exponents as the original attractor for "most" choices of the time delay [15]. (We give a more precise statement below.)

One of the oldest applications of time delay embeddings is to the estimation of the dimension of the attractor. There are many notions of dimension; see [7] and the references therein for more details. In some sense, the dimension describes the number of degrees of freedom associated with the dynamics.

One popular definition is the *correlation dimension*, introduced by Grassberger and Procaccia [16]. Suppose the reconstructed attractor consists of N points. Let

$$C(\epsilon) = \frac{2P(\epsilon, N)}{N(N-1)} , \tag{1}$$

where $P(\epsilon, N)$ is the number of pairs of points on the attractor whose distance apart is less than ϵ. Then $C(\epsilon)$ is a number between 0 and 1. In the limit as $\epsilon \to 0$ and $N \to \infty$, and in the absence of noise, we have $C(\epsilon) \sim \epsilon^d$, where d is the *correlation dimension* of the attractor.

In practice, the finite number of data points and noise in the data place a lower bound on the values of ϵ that can be used to estimate d. Typically, one computes $C(\epsilon)$ for each ϵ in a decreasing sequence. The dimension d is estimated by using linear regression over a range of $(\ln\epsilon, \ln C(\epsilon))$ pairs. One can compare the scaling relations over various ranges by performing the linear regression over different subsets of the ϵ's and looking for consistent values of d.

Ideally, the values for d should be relatively constant for a large interval of ϵ values. For example, if $C(\epsilon)$ has been computed for a decreasing sequence of 30 values of ϵ, one might perform the linear regression over the largest five ϵ's to get an estimate $\hat{d}(\epsilon_1, \ldots, \epsilon_5)$ of the dimension. One could repeat the calculation using the second largest five ϵ's to get $\hat{d}(\epsilon_2, \ldots, \epsilon_6)$, and so on. In this way, one would get estimates of the derivative $d[\ln C(\epsilon)]/d(\ln\epsilon)$.

In a good experimental data set, one might find reasonably constant values of $\hat{d}$ for ϵ ranging from 10% down to 1% or less of the attractor extent. However, a "plateau" in the values of d often is not apparent. The values of $\hat{d}$ may yield a plot that is qualitatively similar to Fig. 1.

In their analysis of weak turbulence in a Couette-Taylor fluid flow experiment, Brandstater and Swinney [17] tried to estimate the correlation dimension over various ranges of ϵ and divided their plot into four regions labeled A–D, as shown in Fig. 1. In region A the lack of data points is the dominant feature. Therefore, the values of $\hat{d}$ are subject to large fluctuations and cannot be used to estimate the dimension reliably. The behavior in region B is attributed to the noise, which "smears out" the fractal structure below a certain level of resolution. One wants to see the behavior in region C for as large a range of ϵ values as possible. The behavior in region D reflects the lack of scale invariance since ϵ is of the order of the size of the entire attractor.

The data in Fig. 1 consist of a time series of 40 000 values from a laser experiment by Flepp and co-workers [18]. The curves illustrate attempts to estimate the correlation dimension of the data before and after a nonlinear noise-reduction method was applied. The region labeled C on each curve corresponds to an approximate power-law relationship between the correlation $C(\epsilon)$ defined in Eq. (1) and the ball size ϵ. Before noise reduction (represented by diamonds), a scaling region is difficult to discern, because the noise obscures the fine scale structure up to 1/16 of the attractor extent. A wide range of estimates of the attractor dimension are possible with the noisy data, depending on the range of ϵ used. In this example, it is difficult simply to determine whether the process is low dimensional using only the noisy data. Hence, even small levels of noise significantly complicate esti-

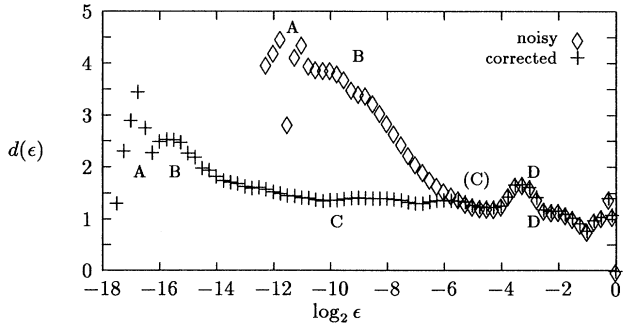

FIG. 1. A typical plot of correlation dimension as a function of the range of distances ϵ used to estimate the scaling exponent. Noise obscures the fine scale structure up to 1/16 of the attractor extent, making an accurate estimate of the attractor dimension impossible.

mates of dimension, a quantity that in principle should be straightforward to measure.

A similar situation arises in the estimation of Lyapunov exponents. In his numerical study of the Lorenz equations, Abarbanel [19] found that the negative Lyapunov exponent cannot be determined when the noise level is as small as 10^{-4}, and none of the exponents can be determined with satisfactory precision in the presence of 1% noise. This is true even if the minimum required embedding dimension is known. (Higher embedding dimensions lead to the additional problem of spurious negative exponents.)

Although dimension estimation provides an illustrative example, there are many other applications in which noise is a serious concern. The ability to calculate accurate approximations to the dynamics allows one to make short-term predictions of the process [20–23] or to determine small changes to an accessible system parameter in order to stabilize the process around a saddle periodic orbit. (The so-called *control of chaos* is a subject of considerable recent interest [24].) Obviously, noise limits one's ability to estimate the dynamics, and a good noise-reduction procedure is essential in such applications.

III. NOISE REDUCTION AND EMBEDDING TECHNIQUES

A. Generalizations of the Takens embedding theorem

The Takens time-delay-embedding method discussed in the preceding section is one of several approaches to the problem of attractor reconstruction. A fundamental objective is to find a reconstruction of the attractor that minimizes the effects of noise in the input time signal.

Casdagli *et al.* [25] consider notions of distortion and noise amplification in attractor reconstruction. Because of the noise, each point in an m-dimensional time-delay reconstruction of the attractor can be regarded as a random m vector. The mean of the underlying density is a nearby "true" vector if the noise has mean zero and is uncorrelated with the signal. Their objective is to try to choose the time delay in such a way that the variance of the random observations around the true state does not have unacceptably large components along regions of interest on the attractor.

Sauer, Yorke, and Casdagli [26] have proved important generalizations of the Takens time-delay-embedding theorem which are especially relevant to the analysis of experimental data. Their work justifies the use of filtering in the reconstruction of attractors to reduce the effect of noise. In the rest of this section, we summarize their main results and related work.

An embedding is a reconstruction for which there is a one-to-one mapping to the original attractor that preserves information about the derivatives of the original flow. This implies that an embedding produces a reconstructed set with the same dimension and Lyapunov exponents as the original attractor.

The Takens embedding theorem asserts that a one-to-one mapping exists for a generic set of time delays, provided that the embedding is done in enough dimensions. How many are necessary?

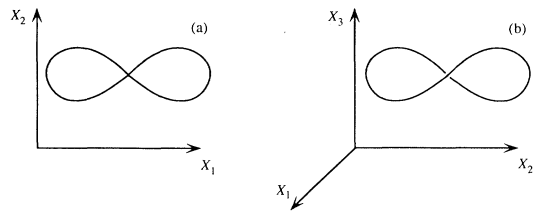

FIG. 2. (a) No small perturbation of the reconstruction function eliminates the self-intersection in the limit cycle. (b) Although the limit cycle intersects itself, arbitrarily small perturbations will remove the self-intersection.

Suppose that the input time series is periodic (a sine wave, for instance). Dynamically this corresponds to a limit cycle, which can be embedded in two dimensions. However, in a two-dimensional Takens time-delay reconstruction, the measurement function may produce a self-intersecting limit cycle, as shown in Fig. 2(a). No small perturbation of the measurement function removes the self-intersection. It is still possible for the curve to intersect itself in three dimensions, but there is a set of arbitrarily small perturbations to the measurement function that produce a one-to-one mapping between the underlying attractor and its reconstruction.

This example provides a heuristic motivation for the rigorous result proved by Sauer, Yorke, and Casdagli [26]: if the box-counting dimension [27] of the attractor is d, then m dimensions suffice to produce an embedding, where $m > 2d$. For a limit cycle ($d = 1$), three dimensions are sufficient. In fact, almost every reconstruction in three dimensions avoids a self-intersection of the sort depicted in Fig. 2(b). This notion of probability one can be made precise in a way that will not be discussed here (see [26] for details). Moreover, typical choices of time delay are likely to produce an embedding. (In the example of a limit cycle, one has an embedding unless the time delay is an integer multiple of the period of the signal.)

An embedding preserves characteristic quantities such as the dimension. This does not mean that reliable estimates of these numbers can be extracted from a given data set because different reconstructions are not equally sensitive to noise and sparse data.

B. Noise reduction with filtered embeddings

The Takens time-delay-embedding method, although popular, is not the only way to reconstruct an attractor. Suitable linear combinations of input time series values can reduce the noise in the reconstructed attractor. Let

$$\mathbf{x}_i = B \begin{pmatrix} s_i \\ s_{i+1} \\ \vdots \\ s_{i+w-1} \end{pmatrix} \tag{2}$$

where B is an $m \times w$ matrix. This sequence of vectors

produces an embedding if the rank of B is sufficiently large and if B does not collapse periodic points of the underlying attractor of integral period less than or equal to w [26]. For example, B can be chosen to yield a Fourier embedding (*e.g.*, $w = 64$ and $m = 16$, so that each point on the reconstructed attractor is a suitably truncated fast Fourier transform of a window of 64 time-series values). Sauer [10] has shown that this "low-pass" embedding can be used to preprocess data with 100% additive Gaussian noise.

Broomhead and King [28] have suggested the use of singular value analysis to reconstruct the attractor. (This is an alternative choice for B.) One can choose a window of the original time series and project it onto the subspace spanned by the corresponding singular vectors. Heuristically, each window of the time series is projected onto the subspace that contains the largest fraction of the total variance of the data. The remaining orthogonal directions presumably contain most of the noise. (See [28] for details.) Albano *et al.* [29] used a similar procedure to reduce the noise and estimate the Grassberger-Procaccia dimension of the attractor reconstructed from time-series data. Other applications of singular value decomposition to noise reduction are described in later sections.

Another alternative is to use a Takens time-delay reconstruction after applying a suitable filter to the entire time series. For example, one can compute the power spectrum of the input signal, make some assumptions about the frequency distribution of the noise, remove or suppress those frequency components, and invert the transform. This is the idea behind traditional bandpass filters (see [30] for a concise description of the Wiener filter).

Traditional filters work well in situations where most of the noise is restricted to certain frequencies. However, the time series of many chaotic signals have broadband components [31,17]. Thus bandpass filters will cause distortion because some of the suppressed frequency components are part of the dynamics. If it is not done carefully, bandpass filtering can fail to give an embedding. If *infinite input response* filters are used, the dimension of the reconstructed set may deviate considerably from that of the original attractor. See [32] for details.

Despite these caveats, we recommend the use of these preprocessing methods because they can reduce the noise significantly. The numerical algorithms are widely available and easily implemented, even on small computers. It is often desirable to try different attractor reconstructions from the same data because the methods distort regions of interest on the attractor in different ways. Depending on the application, one reconstruction may be better than another.

C. Multiple-probe measurements

Multiple redundant measurements can be exploited to reduce noise. For example, several measurements can be taken during the fundamental period of a driven system; several devices can be operated simultaneously (e.g., velocity data can be taken at several nearby locations in a flow) [33] or the data can be taken at time intervals that are short compared to the correlation time of the signal (oversampling).

Suppose there are L probes and at time i we record $s_i^{(1)}, \ldots, s_i^{(L)}$. Each of the L time series contains the same dynamical information and comparable amounts of noise (we assume the measurement errors are independent). The series with the largest variance has the smallest *relative* noise level (highest signal-to-noise ratio). Following the above considerations about embeddings, almost every nonsingular linear combination $V_i = \sum_{k=1}^{L} a_k s_i^{(k)}$ contains the same information about the dynamics. (One need not restrict attention to linear combinations of the time series for attractor reconstruction, but they are convenient.)

A useful strategy is to pick the linear combination which has the highest signal-to-noise ratio. It can be found by maximizing the variance of the V_i subject to the constraint $\sum_{k=1}^{L} a_k^2 = 1$. The requisite vector is the eigenvector corresponding to the largest eigenvalue of the $L \times L$ covariance matrix Γ whose (k,l)th entry is $\Gamma_{kl} = \langle s^{(k)} s^{(l)} \rangle - \langle s^{(k)} \rangle \langle s^{(l)} \rangle$, where the angular brackets denote the average value over all time steps i. This procedure is similar to the singular value decomposition and is also called principal component analysis [34].

IV. MODELING DYNAMICS FROM DATA

The embedding methods described in the preceding section are useful preprocessing tools. However, they are linear methods applied to portions of the time series that are restricted in either the time or frequency domain. They do not exploit the underlying dynamical behavior to identify and correct errors in the observations.

The objective of every noise-reduction method is to find a simpler dynamical system that is consistent with the data. There are three aspects to the problem: (1) the nature of the noise, (2) the class of models to be fitted using a statistical method, and (3) the adjustment of the observations to be more consistent with the model. We discuss these in turn in the following subsections.

A. Measurement error and dynamical noise

Measurement noise refers to the corruption of observations by errors which are independent of the dynamics. The dynamics satisfy $\mathbf{x}_i = \mathbf{f}(\mathbf{x}_{i-1})$, but we measure $\mathbf{x}_i + \eta_i$. [More typically, the measurements consist of scalars $s_i = h(\mathbf{x}_i) + \delta_i$, where h is a smooth function that maps points on the attractor to real numbers and the δ_i are independent and identically distributed random variables.] *Dynamical noise*, in contrast, is a feedback process wherein the system is perturbed by a small random amount at each time step:

$$\mathbf{x}_i = \mathbf{f}(\mathbf{x}_{i-1} + \eta_{i-1}). \tag{3}$$

Dynamical and measurement noise are two notions of the error that may not be distinguishable *a posteriori*

based on the data only. Both descriptions can be consistent to some extent with the same signal. The *shadowing* problem addresses the question of whether there is a trajectory near the observed one, possibly for a different nearby initial condition or for a slightly different map.

Suppose the system contains dynamical noise and that there is a map $\tilde{\mathbf{f}}$ near $\mathbf{f}$ in some metric for which $\mathbf{x}_i = \tilde{\mathbf{f}}(\mathbf{x}_{i-1})$, $i = 1, \ldots, n$. Then the observations can be considered as the exact output of the unknown but slightly different map $\tilde{\mathbf{f}}$, or they can be considered as the noisy output from the known map $\mathbf{f}$. Often one of the two points of view permits a simpler description than the other. For instance, if we detect correlations between δ_i and $\mathbf{x}_{i+\tau}$ for some τ, then we can reject the hypothesis that the noise arises from independent measurement errors. Provided that the series does not contain *exactly* the same phase-space point twice, we can construct a smooth function that interpolates between all measured points; relative to this function, the data form a noise-free orbit.

Strictly speaking, it is not necessary to think of this problem only in terms of separating a deterministic signal from some random fluctuations. It is possible that the "noise" might arise from a high-dimensional, deterministic dynamical system. The noise-reduction problem therefore is a question of how to separate the low-dimensional dynamics from a complex signal. This requires one to choose a class of models for the dynamics and to fit a model to the data in the regions of interest on the attractor. The following sections describe some approaches to this problem.

B. Time ordering and estimation of the dynamics

The overall objective of noise-reduction methods is to find a simple dynamical system that is consistent with the data. Therefore, one must find an approximation to the dynamics from a class of models and adjust the observations to satisfy the dynamical approximations better. Although these two steps are not independent of each other, we will discuss them separately, pointing out along the way where they overlap.

A basic problem is how to exploit the time ordering of the data to produce a more self-consistent trajectory and reduce the noise. Suppose that m dimensions suffice to reconstruct the attractor. In principle, one can write

$$s_{m+1} = f(s_1, \ldots, s_m) + \eta_{m+1} \tag{4}$$

where η_{m+1} denotes a noise term and the s_i are the time series values (possibly preprocessed using the methods outlined above). The subscripts denote the time ordering of the data; i.e., s_{i+1} is the observation immediately following s_i. We call Eq. (4) a *forward-in-time* representation of the dynamics. The prediction problem consists of finding an approximation $\hat{f}$ of f to get an estimate $\hat{s}_{m+1}$ of s_{m+1}, viz. $\hat{s}_{m+1} = \hat{f}(s_1, \ldots, s_m)$.

Typically, an ensemble of nearby trajectories is used to find $\hat{f}$ from a class of models. In some sense, $\hat{s}_{m+1}$ can be regarded as a "maximum likelihood" estimate of s_{m+1}.

This produces a naive (but unsatisfactory) scheme for noise reduction: start with the first m time-series values. Output a prediction $\hat{s}_{m+1}$ for s_{m+1}. Next, use the original observations $s_2, \ldots, s_{m+1}$ to determine $\hat{s}_{m+2}$, and so on. This scheme adjusts all but the first m time-series values.

Can this process be iterated to reduce the noise? That is, if we take the output from the first run and use it as the input to a second run, will the resulting time series be less noisy? The answer is *no* if the underlying time series is chaotic, because errors in $s_1, \ldots, s_m$ typically are amplified due to the sensitivity to initial conditions. For the same reason, the output time series drifts away from the original. (The same is true if we write s_i as a function of the succeeding m observations and work backwards in time.)

One approach which largely avoids this problem was suggested by Schreiber and Grassberger [9]. We review it here to outline the main ideas; a refinement of their procedure is described in a later section. The objective of the procedure is to use past and future values to adjust one or more observations in the middle. Let

$$0 = F(s_1, \ldots, s_m, s_{m+1}) + \eta_{m+1} \tag{5}$$

express the functional dependence of past and future values up to a noise term. If m dimensions suffice to embed the attractor, then any one of the observations in Eq. (5) is an implicit function of the others. We call Eq. (5) an *implicit* representation of the dynamics. Schreiber and Grassberger [9] use a linear approximation for F to find the least-squares estimate

$$\hat{s}_{m/2} = \sum_{\substack{k=1 \\ (k \neq m/2)}}^{m+1} a_k s_k + b \tag{6}$$

for the value in the middle of the sequence. This method uses information from both the future and the past to adjust the observations.

The map is determined in the following way. Given $s_1, \ldots, s_{m+1}$, one locates several closely matching sequences of $m+1$ observations. The middle value in each sequence is expressed as a linear combination of the others as in Eq. (6). The coefficients (except for $a_{m/2}$) are determined from a least-squares fit. Notice that the coefficient $a_{m/2}$ must be excluded to prevent a trivial fit (wherein $a_k = 0$ for $k \neq m/2$ and $a_{m/2} = 1$).

In a similar way, one can use $s_2, \ldots, s_{m+2}$ to determine a new set of coefficients to get an estimate $\hat{s}_{m/2+1}$ for the next output point, and so on to the end of the time series. This procedure adjusts all but the first and last $m/2$ values. The output is a less noisy time series if the linearization in Eq. (6) is an accurate approximation of the dynamics.

Because past and future values are used at each step, the procedure can be iterated without the rapid drift due to the sensitive dependence on initial conditions. (One must be careful to avoid especially large corrections, which can arise because the linearization is a poor approximation of the dynamics or because of occasional "glitches" in the data.) Schreiber and Grassberger's method is straightforward to program, runs quickly on a desktop workstation, and can reduce the noise by a factor of 10 in some cases [9].

Schreiber [35] proposed a very simple variant of the above method which gives stable results especially for very short and noisy data: the linear approximation (6) is replaced by a constant, which can be determined with much less data. The resulting algorithm can be coded in a few lines and is quite efficient.

C. Eckmann-Ruelle linearization

One can work directly with each m-dimensional point on the reconstructed attractor to determine a local linear approximation of the dynamics. As above, the subscript denotes the time ordering of the data: $\mathbf{x}_{i+1}$ is the point immediately following $\mathbf{x}_i$ in time. The dynamics are given by $\mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i)$ for an unknown function $\mathbf{f}$.

One assumes that $\mathbf{f}$ is at least piecewise differentiable and considers a low-order polynomial expansion of $\mathbf{f}$. Eckmann and Ruelle [36] suggested that the local linear approximation $\mathbf{f}(\mathbf{x}_{\rm ref}) = A\mathbf{x}_{\rm ref} + \mathbf{b}$ can be computed from the data using least squares. Here A is an approximation of the Jacobian matrix of partial derivatives of $\mathbf{f}$ evaluated at $\mathbf{x}_{\rm ref}$.

Suppose that $\mathbf{x}_{\rm ref}$ has k neighbors within a suitably small neighborhood U. Linear regression is used to find the matrix A and vector $\mathbf{b}$ that minimize the sum of squares $\sum_j \|\mathbf{x}_{j+1} - (A\mathbf{x}_j + \mathbf{b})\|^2$, where the sum runs over all indices j such that $\mathbf{x}_j \in U$. A different A and $\mathbf{b}$ are computed for each neighborhood on the reconstructed attractor.

The noise reduction scheme described in Kostelich and Yorke [6] uses Eckmann-Ruelle linearization to approximate the dynamics at each point on the reconstructed attractor. To avoid the drift due to the sensitivity on initial conditions discussed above, it must be followed by a separate trajectory adjustment step in which a nearby time series is computed that is more consistent with the linearization of the dynamics at each point. (This is outlined in Sec. V.)

There are several difficulties with this kind of function approximation that have been considered in detail in [37]. Two of the more important considerations are the following: (1) The accuracy of the approximation depends on how well a linear map describes the dynamics. Some regions of the attractor may contain few observations. A larger ball size increases the number of available points but makes nonlinearities more prominent. This situation becomes more common as the dimension of the attractor increases. (2) The linear least-squares procedure produces biased estimates of the matrix A and vector $\mathbf{b}$ because errors in measurement are present in all the observations; i.e., there is error in both the original points $\{\mathbf{x}_j\}$ and their successors $\{\mathbf{x}_{j+1}\}$. Ordinary least squares produces unbiased estimates only when the errors are confined to the latter set.

It is difficult to make general statements about the effect of these problems in practice. Much depends on the dimension of the attractor, the number of data points, the noise level, etc. Nevertheless, Eckmann-Ruelle linearization has useful applications both in noise reduction and Lyapunov exponent estimation.

D. Local projective maps

An alternative approach uses geometrical considerations to reduce the noise. Since the reconstructed attractor is a subset of a smooth manifold in an m-dimensional phase space, one can estimate the local tangent plane at each point with singular value decomposition. Let $\{\mathbf{x}_i\}_{i=1}^k$ be the set of all points in a small neighborhood of some reference point with mean $\bar{\mathbf{x}}$, and let X be the $k \times m$ matrix whose ith row is $\mathbf{x}_i - \bar{\mathbf{x}}$. Then

$$X = U^T \Sigma V \tag{7}$$

where the columns of U and V form an orthonormal basis for the columns and rows of X, respectively [38]. The entries σ_i of the diagonal matrix Σ are the *singular values* of X: they are the non-negative square roots of the eigenvalues of the covariance matrix $X^T X$. (The decomposition can be arranged so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$.) The sum $\sigma_1^2 + \cdots + \sigma_m^2$ equals the total variance in the observations $\mathbf{x}_i$ [38]. The columns of the orthonormal matrix V are the corresponding singular vectors.

The noise spreads out the observations in all directions, but most components lie along a lower-dimensional hyperplane through $\bar{\mathbf{x}}$. The components in the remaining orthogonal directions presumably are mostly noise. Thus one can ask how many dimensions account for a certain fraction of the total variance. For example, if p is the integer such that $\sigma_1^2 + \cdots + \sigma_p^2$ is about 95% of the total variance, then the span of the first p singular vectors is a good approximation to the tangent plane of the attractor in a neighborhood of $\bar{\mathbf{x}}$. (Obviously, the observations must lie in a small ball for this to be true. Whether one considers 95% of the variance or some other fraction depends upon the estimated underlying noise level. See [39] for details.)

Cawley and Hsu [11] and Sauer [10] suggested that noise in the observations can be reduced by projecting the observations onto the subspace spanned by a suitable collection of singular vectors at each point on the attractor. (Alternatively, instead of projecting each observation directly onto the subspace, one can move the observation only part of the way. This minimizes possible corruption of the data due to a poor approximation to the tangent plane.)

Figure 3 is a schematic diagram of the process. The method is purely geometric; the time ordering of the data is not used to compute the map.

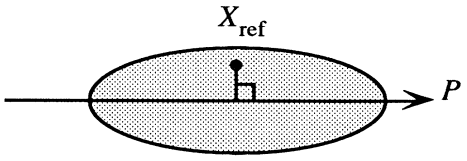

FIG. 3. Schematic diagram of the noise-reduction process in one dimension using singular value decomposition. The line P is the span of the first singular vector. The ellipse indicates that noise spreads out the observations in a neighborhood of the reference point. The reference point (and possibly some nearby points) are projected orthogonally onto P in order to reduce the noise.

When the attractor is reconstructed using a Takens time-delay embedding, the projections typically produce m different corrections for each time series value. Thus it is necessary to incorporate a trajectory adjustment step to compromise between the projections.

E. Local projections with constraints

Grassberger *et al.* [40] have considered a more general procedure in which the projective maps and the corrections to the observations are computed in one step as part of a least-squares minimization problem. Like the projective method above, this method circumvents the problem of biased estimates of derivatives arising from errors in all of the observations and obviates a separate trajectory adjustment step. In addition it exploits the time ordering of the data in a clever way. Figure 4 was obtained using the method described in this section; see [41] for the parameters used.

As above, one assumes that when the embedding dimension m is higher than strictly necessary, the observations can be projected onto a d-dimensional manifold, which is assumed to be approximately linear in a suitably small neighborhood of the reference point. Thus the components of each observation that lie in a q-dimensional subspace perpendicular to the tangent plane of the manifold are zeroed out by the projection, where $q = m - d$.

However, Grassberger *et al.* [40] suggested that the projection should be done in such a way that changes are made only to a central block of coordinates in each observation. (In three dimensions, only the middle value of each observation would be adjusted. If $m = 7$, then one might try to project the middle three observations while leaving the first two and last two essentially unchanged.) Large corrections to coordinates at the end of each observation will propagate due to the sensitive dependence on initial conditions. Similarly, large corrections in the first few coordinates will grow in backwards time. Hence this method uses information from the future and past dynamics to find the best correction to the middle coordinate values. When only the middle value of each observation is adjusted, the method becomes very similar to the one by Schreiber and Grassberger discussed in Sec. IV B.

Let $\mathbf{x}_{\mathrm{ref}} = (s_i, \ldots, s_m)$ be an $(m+1)$ vector whose coordinates are consecutive values from the input time series. (We assume that the attractor can be embedded in m or fewer dimensions.) As above, we let U denote a suitably small neighborhood of $\mathbf{x}_{\mathrm{ref}}$. The components of each $\mathbf{x}_j \in U$ correspond to portions of nearby trajectories on the reconstructed attractor, each of which lies on a manifold of dimension $m - q$. An implicit functional relationship between the components of each $\mathbf{x}_j$ is defined by $\mathbf{F}(\mathbf{x}_j) = \mathbf{0}$, where $\mathbf{F}$ is a vector function with q components. (This can be regarded as a set of q constraints.)

The relation $\mathbf{F}$ is assumed to be approximately linear in U. For each observation $\mathbf{x}_j \in U$, one seeks a vector θ_j, an $(m+1) \times q$ matrix A and a q-vector $\mathbf{b}$ such that (1) $A^T(\mathbf{x}_j + \theta_j) + \mathbf{b} = \mathbf{0}$; (2) $A^T P A = I$; and (3) $\sum \theta_j^T P^{-1} \theta_j$ is a minimum, where the sum is over all the points in U. The corrections θ_j, the matrix A, and the vector $\mathbf{b}$ can be computed in a single minimization problem. The matrix P is a fixed $(m+1) \times (m+1)$ diagonal weight matrix. The entries of P are chosen to penalize large corrections to the first and last few coordinates of each point in the neighborhood. The columns of A are the eigenvectors corresponding to the q smallest eigenvalues of the weighted covariance matrix of the observations. Additional details are given in [40].

Intuitively, the columns of A form an orthogonal basis for a subspace that is perpendicular to the tangent plane of the manifold containing the observations. One seeks the smallest corrections that project each point onto the tangent plane, but distances are measured with a metric determined by the weight matrix P [40]. (The columns of A are not an orthonormal basis for the perpendicular subspace unless P is the identity matrix, in which case the method becomes identical to the one by Cawley and Hsu discussed above.)

Finally, we note that a similar approach, in which a local linear approximation and the adjustments to the observations are computed in one step, was outlined in [37] for cases where the observations are near saddle periodic orbits.

F. Global function fits and other methods

Global models try to find one function $\hat{f}$ which gives the best fit to f, where the sum of squares is taken over all the data. A classical method is to express $\hat{f}$ as a linear combination of a set of k basis functions:

$$\hat{f} = \sum_{j=1}^{k} \alpha_j \psi_j. \tag{8}$$

The coefficients α_j are then chosen to minimize the rms approximation error, which is a linear optimization problem that can be solved by standard techniques. A popular choice of functions ψ_j (at least in the context of chaotic dynamics) are radial basis functions [42,43]: $\psi_j(\mathbf{x}) = \phi(\|\mathbf{c}_j - \mathbf{x}\|)$. All basis functions thus have the same functional form and are distinguished only by the different center points $\mathbf{c}_j$. A variety of choices is possible for $\phi(r)$: examples include Gaussian functions, exponential functions, low-order polynomials, and rational functions. The coefficients usually are obtained by minimizing

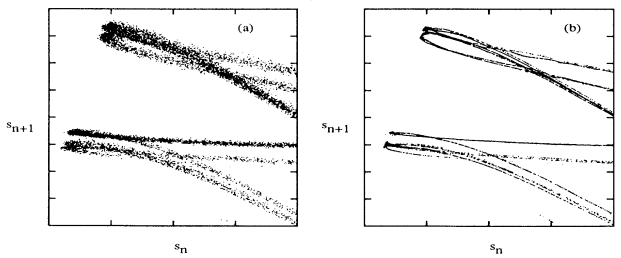

FIG. 4. Enlargements of phase portraits of the Zürich NMR laser data (same data as in Fig. 1): (a) unprocessed data, (b) the same region after noise reduction with constrained projections (Sec. IV E). This is a case where the phase portrait shows effective noise reduction.

$$E^2 = \frac{1}{N}\sum_i \left[s_{i+m+1} - \hat{f}(s_i, \ldots, s_{i+m})\right]^2$$

where the sum runs over all the time series values in the data set. The numerical method of choice for this least-squares problem is singular value decomposition; see Press *et al.* [30].

The success of the method depends on the appropriate choice of the form of the basis functions (like the width of the Gaussian functions whenever they are used). In addition, one needs a strategy for choosing the centers $\mathbf{c}_j$ and a criterion for deciding how many basis functions are required. More basis functions leads to closer approximations of the input time series. Of course the original map f is not known, so as more centers are chosen we model more details of the noisy data. When the number of basis functions equals the number of input points, we have an interpolation between all the observations: the interpolation is exact on the observations, the noise is interpreted as part of the dynamics, and the function $\hat{f}$ cannot be used for trajectory adjustment. More material on radial basis-function approximation can be found in [21,42,44].

Another method of global nonlinear function approximation is the modulation by neural networks. In experienced hands they have proven capable of learning nonlinear functions useful for time-series analysis. Weigend, Huberman, and Rumelhart [45] successfully applied this technique to predict sunspot numbers. Neural networks were used successfully in a recent Santa Fe Institute time series prediction contest [23], but the performance of the different networks varied widely. Further references about function approximation by neural networks can be found in [46].

Genetic algorithms under certain circumstances can learn to make predictions [47], but so far seem of limited use for noise reduction. They do not make predictions for every reference point, but only for certain regions in phase space.

Although global function fits have some appeal, the choice of basis functions induces some bias. For instance, the accuracy of the dynamical approximations at each point on the attractor depends in a nontrivial way on both the shape of the basis functions, the distribution of the centers, and the curvature of the trajectories. In the case of neural nets the nature of this bias is unknown. The main advantage of global models is that they can provide stable fits even for small amounts of data.

Finally we remark that straightforward global approximations work mainly for forward-in-time fits as in Eq. (4). In many cases the dynamical equations cannot be solved globally to give a unique value for coordinates other than s_{m+1}, even when the map is invertible. For example, the Hénon map is $s_3 = 1 - as_2^2 + bs_1$. If we try to solve for s_2 we obtain $s_2 = \pm\sqrt{1 + bs_1 - s_3/a}$, which usually has two branches. On a noisy orbit the square root can even become imaginary.

V. TRAJECTORY ADJUSTMENT

Every noise-reduction method requires some strategy to adjust the observations to respect the dynamics more closely. Two requirements must be fulfilled by the new orbit: it should be more consistent with the estimated dynamics than the original signal and it should remain close to the original signal. (In cases where a Takens time-delay embedding is used to reconstruct the attractor, the observations should be adjusted so that the output is a scalar time series.)

Usually the new orbit is obtained as the result of some optimization procedure. It is not possible to satisfy all the criteria simultaneously, so one must make compromises in any particular algorithm. For example, a given point on the attractor may belong to several neighborhoods. Hence, different maps could be used to adjust the point, each of which moves the point in a different way. An open question concerns the best way to reconcile all the possible adjustments.

A. One-step methods

The trajectory adjustment step is trivial if the dynamics are represented according to the Schreiber-Grassberger scheme in Eq. (6). The central coordinate $s_{m/2}$ is well controlled in both the stable and the unstable directions on the attractor, because past and future values are used to adjust it. Thus the estimate $\hat{s}_{m/2}$ can be taken as a replacement for the observed value. The value of the coefficient $a_{m/2}$ must be fixed between 0 and 1 to avoid a trivial least-squares problem. Hence a smaller value can be used to lend more weight to the dynamics, and a value closer to 1 can be used to obtain smaller corrections. The method automatically yields a scalar signal when the attractor is reconstructed with a Takens time-delay embedding.

Only slightly more work is needed in the constrained projective procedure described in Sec. IV E. In this case, proposed corrections are obtained for *all* components of the observations instead of a single central coordinate. In a Takens time-delay embedding, the same scalar measurement appears as a coordinate in several vectors, so the different proposed corrections can be averaged together. (See [40] for more details.) The same considerations apply to the geometric method of Cawley and Hsu [11] and Sauer [10].

B. Two-step methods

Forward-in-time reconstructions require an explicit nontrivial adjustment step, particularly when the attractor is reconstructed using Takens time-delay embeddings. Some methods require the new orbit to fulfill the dynamics exactly. This is reasonable when the dynamics are known with high accuracy. Algorithms of Hammel [48] and Farmer and Sidorowich [12] are examples in this category. When the map is known, it is possible to refine trajectories to arbitrarily high precision.

This is not the case in laboratory situations because the dynamics must be estimated from noisy data. Instead, one must minimize the deviation from the individual adjustments subject to the constraint that the

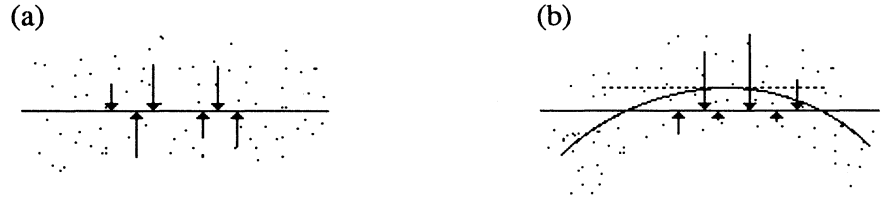

FIG. 5. Consider a noisy one-dimensional set. In the absence of curvature (a), the mean of the corrections is zero. If the underlying manifold is curved (b), projection onto the fitted straight line yields a nonzero mean correction, which can be subtracted to come closer to the desired tangent.

distance from the original orbit must not be too large. Kostelich and Yorke [6] sought the trajectory $\{\hat{x}_{i+k}\}_{k=0}^{p}$ that minimizes the sum of squares

$$\sum_{k=0}^{p} \|\hat{\mathbf{x}}_{i+k} - \mathbf{f}(\hat{\mathbf{x}}_{i+k-1})\|^2 + w\|\hat{\mathbf{x}}_{i+k} - \mathbf{x}_{i+k}\|^2 + \|\hat{\mathbf{x}}_{i+k+1} - \mathbf{f}(\hat{\mathbf{x}}_{i+k})\|^2 \quad (9)$$

where $\mathbf{f}$ denotes the estimated dynamics at each point. Equation (9) expresses the idea that the distances should be small between each fitted point and its preimage, each fitted point and the original observation, and the fitted point and its image. Since their method uses a Takens time-delay reconstruction of the attractor, the minimization in Eq. (9) is done so that the output consists of a sequence of scalar values. The distances between each point and its image are weighted twice as heavily as the distance to the original observation when the parameter w is set to 1. Larger values of w may be appropriate in cases where the input contains large, isolated errors in the data so that the trajectory is not moved excessively to accommodate an occasional erroneous observation.

The minimization problem is more complicated if the model $\mathbf{f}$ is not a piecewise linear function. Strategies for this case have been discussed by Davies [49]. If the nonlinear minimization problem is solved by gradient descent, the weight w can be absorbed into the stepsize for the descent.

We remark that no matter how the sum of squares in Eq. (9) is minimized, the Jacobian of $\mathbf{f}$ must be available. (This is trivial for piecewise linear approximations of the dynamics.) In contrast, the one step methods described above do not need the Jacobian function. This can make one-step methods numerically more robust. For instance, errors in the coefficients computed using the method of Schreiber and Grassberger are not important as long as they predict $s_{m/2}$ reasonably well. The method of Kostelich and Yorke depends more sensitively on the linear maps because they are also used for trajectory adjustment.

C. Recommended improvements

In this section we mention some modifications which were omitted from the description of the algorithms for clarity. Their implementation is not required in order to reduce the noise in most cases, but they can enhance the performance of the methods considerably.

All of the noise-reduction methods occasionally may make anomalously large corrections to an observation. This happens mainly at points where the stable and unstable manifolds are almost tangential. They can also be caused by singular least-squares fits or by a large, isolated error in the data. Since it is not possible to determine the underlying problem with an automatic computer program, most implementations restrict the maximum size of any correction. Kostelich and Yorke [6] did not alter such points at all, but simply flagged them in the output. Alternatively, such points can be moved only some fraction of the computed distance; these observations can be corrected gradually in subsequent iterations over the data set.

Locally linear and projective models introduce errors due to the presence of small nonlinearities, such as the curvature of the attractor. This can be detected because the curvature error is systematic and makes the mean of the corrections nonzero. Forcing the corrections to all points in a small neighborhood to have zero mean partially compensates for this. Sauer [10] proposed this modification, which becomes particularly important when one attempts to remove very small noise levels and curvature errors dominate. Figure 5 illustrates how curvature causes a nonzero mean correction. If this trend is subtracted, the linear approximation approaches the desired tangential form (dashed line).

VI. HOW MUCH NOISE IS TAKEN OUT?

This is a fundamental question. The answer depends partly on the assumptions one makes about the nature of the noise in the input signal. Two error measures have been used for the development of algorithms. When the noise-free signal is known, the error in the input and output is given by the rms distance between the data s_i and the original noise free signal s_i^0:

$$E_0 = \left(\frac{1}{N} \sum_{i=1}^{N} (s_i - s_i^0)^2 \right)^{1/2} .$$

A similar quantity $\hat{E}_0$ can be computed for the cleaned data $\hat{s}_i$. If $\hat{E}_0 < E_0$, then the noise has been reduced. A slightly weaker error measure can be computed when the exact dynamical evolution equation $\mathbf{x}_i^0 = \mathbf{f}(\mathbf{x}_{i-1}^0)$ is known. It measures the deviation from deterministic behavior according to the equation

$$E_{\mathrm{dyn}} = \left(\frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{f}(\mathbf{x}_{i-1})\|^2 \right)^{1/2} .$$

As before, the dynamical error can be compared before and after the noise-reduction procedure has been applied.

These are natural measures when the original trajectory or the "true" dynamics are known. However, other points of view are possible. For example, the output from a noise-reduction scheme might be considered as a cleaner set of data for a slightly different value of the parameter. It is also possible for a noise-reduction procedure to translate the data slightly, resulting in large values for E_{dyn}. Nevertheless, the data might still be regarded as a less noisy realization of the dynamics in slightly different coordinates.

The main disadvantage of these quantities is that in a typical experimental situation neither the noise-free data nor the exact dynamics is known. Thus we must employ criteria which are accessible given only the data. Figure 4 shows an example where the processed data "looks less noisy," but we need a more objective, quantitative criterion.

One reasonable assumption is that the data are composed of a deterministic part plus random noise. Thus we can measure the success of noise-reduction algorithms by testing whether the signal has become more deterministic and the subtracted noise is random. Sections VI A and VI B outline methods to test this assumption.

A. Correlation sums

A widely used tool for detecting and quantifying low-dimensional behavior is the correlation sum C defined in Eq. (1). This quantity is used to estimate the correlation dimension d of an attractor. Here we are interested in the scale-dependent effective dimension $d(\epsilon) = d[\ln C(\epsilon)]/d(\ln \epsilon)$. For an infinite amount of noiseless data, $d = \lim_{\epsilon \to 0} d(\epsilon)$.

For laboratory data, this scaling of interpoint distances exists only over a finite range of length scales. As pointed out in Sec. II, the scaling breaks down for larger distances (region D in Fig. 1). At small scales two effects can be seen. When points become sparse, the effective dimension of the attractor fluctuates wildly and no scaling is found (region A). If enough pairs are still found with a distance smaller than the noise level, one measures an effective dimension $d(\epsilon)$ that is close to the dimension of the whole phase space (usually given by the embedding dimension, region B).

If the data set is too short, then region B may not be observable because points become sparse at or above the noise level. This implies that it is not possible to gather enough information on small scales to determine whether a point is displaced by noise and to correct the error. In this case, nonlinear noise reduction tends to break down.

If there are enough points, one finds a crossover region between the noise dominated length scales (region B) where $d(\epsilon)$ equals the embedding dimension and the "correct" scaling (region C) where $d(\epsilon) \approx d$. The length scale at which this crossover occurs is proportional to the noise level. (For Gaussian noise, this observation can be quantified [50].) If one succeeds in reducing the noise in a chaotic time series, then this region is pushed to a smaller scale, as seen in Fig. 1. In the best case, the region is reduced to a scale where spareseness is the dominating effect.

B. Prediction error

If a noise-reduction procedure successfully removes most of the noise, then the processed data set should have better short-term predictability then the raw data set. In other words, the data should be more self-consistent. This can be quantified by computing the *out-of-sample* prediction error with some nonlinear predictor before and after noise reduction. In the simplest case, the data are split into two parts, one of which must be long enough to fit a nonlinear mapping $\hat{\mathbf{f}}(\mathbf{x})$. (One can use the same maps as in the noise reduction itself, but this does not have to be the case.) The prediction error

$$E_p = \left(\frac{1}{N_p} \sum_{i=N-N_p+1}^{N} \|\mathbf{x}_i - \hat{\mathbf{f}}(\mathbf{x}_{i-1})\|^2 \right)^{1/2}$$

is computed over the remaining N_p data points. It is essential to take the out-of-sample error because one can always find a mapping which yields zero in-sample prediction error simply by interpolating between the data points.

If the data set is too short to be split into two sufficiently long parts, then *take-one-out* statistics can be used to obtain an out-of-sample error [7,51]. For each reference point $\mathbf{x}_i$, a predictor $\hat{\mathbf{f}}_i$ is fitted using all the data in the neighborhood except $\mathbf{x}_i$. The resulting map $\hat{\mathbf{f}}_i$ is used only to predict $\mathbf{x}_i$. In this way, one obtains

$$E_p = \left(\frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{f}}_i(\mathbf{x}_{i-1})\|^2 \right)^{1/2}$$

as an estimate of the out-of-sample prediction error. This error measure is computed before and after noise reduction. (More material on error measures for nonlinear time-series prediction can be found in [22,23].)

Three effects contribute to the one-step prediction error in low-dimensional chaotic systems.

(1) The observed value $\mathbf{x}_i$ is contaminated by noise, inducing an error proportional to the noise level.

(2) The data values used as arguments of the prediction function are contaminated with noise, leading to an error approximately proportional to the noise level. The proportionality constant is determined by the Jacobian determinant of $\hat{\mathbf{f}}$.

(3) The fitted prediction function is only an approximation to the true dynamics. The accuracy of the fit can be a complicated function of the noise level.

For dissipative maps, effect (1) usually dominates (2). The size of effect (3) is mostly unknown. However, we expect the deviation of the predictor from the true dynam-

ics to increase monotonically with the noise level. In the presence of all three effects, E_p will increase faster than linearly with the noise level. If we compare E_p before and $\hat{E}_p$ after noise reduction we only obtain an upper bound on the amount of noise reduction. However, if $\hat{E}_p < E_p$, then the data have been rendered more self-consistent.

C. Power spectra

Chaotic data are characterized by broadband power spectra [31,17]. Thus it is generally not possible to reduce noise in such data using methods based on power spectra. However, there are cases where signal and noise can be readily distinguished in some part of the spectrum, for example, at high frequencies or around a dominant frequency. Nonlinear noise-reduction methods should suppress noise throughout the spectrum, and their effects also should be visible in these places.

Let us take densely sampled flow data as an example. High-frequency noise can be removed with a simple low-pass filter. Nonlinear noise-reduction methods tend to suppress the high frequencies as well. However, some of the high frequencies are part of the dynamics. Nonlinear noise-reduction methods also can remove some of the lower-frequency noise that is not part of the dynamics. Such distinctions cannot be made with a low-pass or Wiener filter [6].

We do not recommend the power spectrum as a measure of the amount of noise reduction in chaotic time-series data, because it is essentially a linear method. As we discuss below, however, the power spectrum is a useful characterization of the corrections applied to the input data. Most nonlinear methods have limiting cases where they behave like linear filters (for example, when large neighborhoods are used to fit in local linear models or when inappropriate basis functions are used to fit global models). Successful nonlinear noise reduction should make sure we do not pick one of these cases.

D. Consistency tests

So far we have only described tests which can be applied to the cleaned signal in order to check whether it is likely to be low-dimensional deterministic chaos. If the input data consist of uncorrelated noise added to a deterministic signal, then the corrections applied to the data should resemble a random process. One can exploit all that is known about the sources of noise in the data. In most cases, it has only short correlation times. The spectrum will not always be flat, but if it has unexpected features, then one must consider whether the method used to process the data is appropriate.

For example, it is a good idea to check the distribution of the corrections to the data. Measurement errors are expected to have a normal distribution, whereas discretization errors are uniformly distributed in $[0, 2^{-r-1}]$ if the data are stored as r-bit integers.

In addition, one should look for cross correlations between the signal and the corrections. They should be small if the noise-reduction procedure has successfully removed random measurement errors. Significant cross correlations can arise for different reasons, including some threshold in the measurement device, fluctuations in the scale of measurement ("multiplicative noise"), varying noise levels in different regions of the phase space, and corruption of the data due to an inappropriate noise-reduction method.

VII. CONCLUSIONS

The basic problems involved in the analysis of chaotic time-series data are reasonably well understood. Several methods for reducing noise are available, provided that the underlying dynamics are low dimensional. The underlying attractor can be reconstructed from the data in several possible ways, some of which can act as linear filters to lower the noise level.

More sophisticated noise-reduction methods are available which exploit the local dynamical behavior to identify and correct errors arising from noise. A variety of methods exists for estimating the dynamics. One-step methods, such as those suggested by Grassberger and coworkers [7,40], adjust the observations as part of the process of determining a local linear model for the dynamics. Projective schemes, such as that implemented by Cawley and Hsu [11], move the observations onto a subspace that approximates the tangent plane to the manifold containing the attractor at each point. Kostelich and Yorke [6,8] compute an estimate of the Jacobian matrix of partial derivatives of the map at each point on the attractor.

Trajectory adjustment refers to the process by which the input data are changed to be more consistent with the estimated dynamics. Although the true dynamics are not known in most experimental situations, there are several ways to measure the self-consistency of the data and to check whether the noise level has been reduced.

Even low levels of noise can complicate the estimation of basic quantities such as attractor dimension and Lyapunov exponents. For this reason, the use of noise-reduction methods is strongly recommended in any analysis of chaotic time-series data.

ACKNOWLEDGMENTS

The authors thank L. Flepp, J. Simonet, and E. Brun for the data used in Figs. 1 and 4. E.K. is supported by the National Science Foundation Applied and Computational Mathematics Program under Grant No. DMS-9017174. T.S. receives a European Communities grant within the framework of the SCIENCE program, Contract No. B/SCI*-900557.

* Permanent address: Physics Department, University of Wuppertal, D-5600 Wuppertal 1, Gauss-Strasse 20, Germany.

[1] B.-L. Hao, *Chaos II* (World Scientific, Singapore, 1990).
[2] *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress (Springer-Verlag, Berlin, 1986).
[3] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[4] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986).
[5] A readable discussion of the definitions can be found in D. Gulick, *Encounters with Chaos* (McGraw-Hill, New York, 1992).
[6] E. J. Kostelich and J. A. Yorke, Physica D **41**, 183 (1990).
[7] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos **1**, 521 (1991).
[8] E. J. Kostelich and J. A. Yorke, Phys. Rev. A **38**, 1649 (1988).
[9] T. Schreiber and P. Grassberger, Phys. Lett. A **160**, 411 (1991).
[10] T. Sauer, Physica D **58**, 193 (1992).
[11] R. Cawley and G.-H. Hsu, Phys. Rev. A **46** 3057 (1992); R. Cawley and G.-H. Hsu, Phys. Lett. A **166** 188 (1992).
[12] J. D. Farmer and J. Sidorowich, in *Evolution, Learning and Cognition*, edited by Y. C. Lee (World Scientific, Singapore, 1988); Physica D **47** 373 (1991).
[13] J. C. Roux, R. H. Simoyi, and H. L. Swinney, Physica D **8**, 257 (1983).
[14] K. Coffman, W. D. McCormick, Z. Noszticzius, R. H. Simoyi, and H. L. Swinney, J. Chem. Phys. **86**, 119 (1987).
[15] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, New York, 1981), p. 366.
[16] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).
[17] A. Brandstater and H. L. Swinney, Phys. Rev. A **35**, 2207 (1987).
[18] L. Flepp, R. Holzner, E. Brun, M. Finardi, and R. Badii, Phys. Rev. Lett. **67**, 2244 (1991); M. Finardi, L. Flepp, J. Parisi, R. Holzner, R. Badii, and E. Brun, *ibid.* **68**, 2989 (1992).
[19] H. D. I. Abarbanel, in [22].
[20] J. D. Farmer and J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).
[21] M. Casdagli, Physica D **35**, 335 (1989).
[22] *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank, Santa Fe Institute Studies in the Science of Complexity, Proc. Vol. XII (Addison-Wesley, Reading, MA, 1992).
[23] *Predicting the Future and Understanding the Past: A Comparison of Approaches*, Proceedings of the NATO ARW on Time Series Analysis and Forecasting held in Santa Fe, NM, 1992, edited by A.S. Weigend and N.A. Gershenfeld, (Addison-Wesley, Reading, MA, 1993).
[24] See for example F. Romeiras, C. Grebogi, E. Ott, and W. P. Dayawansa, Physica D **58**, 165 (1992); E. Ott, C. Grebogi, and J.A. Yorke, Phys. Rev. Lett. **64** 1196 (1990); G. Nitsche and U. Dressler, Physica D **58** 153 (1992); W. L. Ditto, S. N. Rauseo, and M. L. Spano, Phys. Rev. Lett. **65**, 3211 (1990).
[25] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, Physica D **51**, 52 (1991).
[26] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[27] The box-counting dimension can be defined in a manner analogous to the correlation dimension. Let $N(\epsilon)$ be the minimum number of cubes of side length ϵ needed to cover the attractor. Then $N(\epsilon) \sim \epsilon^d$ as $\epsilon \to 0$, where d is the box-counting dimension. For more details, see J. D. Farmer, E. Ott, and J. A. Yorke, Physica D **7**, 153 (1983).
[28] D. Broomhead and G. P. King, Physica D **20**, 217 (1986).
[29] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp, Phys. Rev. A **38**, 3017 (1988).
[30] W. H. Press, B. P. Flannery, S. A. Teukolski, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1988).
[31] P. R. Fenstermacher, H. L. Swinney, and J. P. Gollub, J. Fluid Mech. **94**, 103 (1979).
[32] R. Badii, G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M. A. Rubio, Phys. Rev. Lett. **60**, 979 (1988); R. Badii *et al.*, *ibid.* **60**, 979 (1988); M. Muldoon (unpublished); Milan Pauluš and Ivan Dvořák, Physica D **55**, 221 (1992).
[33] P. L. Read (unpublished).
[34] R.W. Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography* (Elsevier, Amsterdam, 1988).
[35] T. Schreiber, Phys. Rev. E **47**, 2401 (1993); A. Pikovsky, Sov. J. Commun. Technol. Electron. **31**, 81 (1986).
[36] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[37] E. J. Kostelich, Physica D **58** 138 (1992).
[38] See, for example, G. W. Stewart, *Introduction to Matrix Computations* (Academic, New York, 1973).
[39] D. Broomhead, R. Jones, and G. P. King, J. Phys. A **20**, L563 (1987).
[40] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T Schreiber, Chaos (to be published).
[41] H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L.G. Flepp, J. Simonet, R. Badii, and E. Brun, Phys. Rev. E **48**, xxxx (1993).
[42] D. Broomhead and D. Lowe, Complex Syst. **2**, 321 (1988).
[43] J. Holzfuss, and J. B. Kadtke, Int. J. Bifurcation Chaos (to be published).
[44] See L. A. Smith, Physica D **58** 50 (1992), and references therein.
[45] A. S. Weigend, B. A. Huberman, and D. E. Rumelhart, Int. J. Neural Syst. **1** 193 (1990); A. S. Weigend, B. A. Huberman, and D. E. Rumelhart, in [22].
[46] S. A. Solla, in [22].
[47] J. L. Breeden and N. Packard, Technical Report No. CCSR-92-11, 1992 (unpublished).
[48] S. M. Hammel, Phys. Lett. A **148**, 421 (1990); M. E. Davies, *ibid.* **169** 251 (1992).
[49] M. E. Davies, Int. J. Bifurcation Chaos (to be published); M. E. Davies (unpublished).
[50] T. Schreiber, Phys. Rev. E (to be published).
[51] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, 1982).