# Bootstrap estimates of chaotic dynamics

Eric J. Kostelich
*Department of Mathematics, Arizona State University, Tempe, Arizona 85287-1804*
(Received 1 August 2000; published 18 June 2001)

Bootstrap sampling is a nonparametric method for estimating the standard error of a statistic. This paper describes the application of bootstrap sampling to estimate the error in local linear approximations of the dynamics on chaotic attractors reconstructed from time series measurements. We present an algorithm for identifying influential points, i.e., observations with an especially large effect on a least-squares fit, and an algorithm to estimate the standard error of regression coefficients obtained from total least squares. We also consider the application of bootstrap methods to assess the uncertainty in Lyapunov exponent computations from chaotic time series.

DOI: 10.1103/PhysRevE.64.016213 PACS number(s): 05.45.Tp

## I. INTRODUCTION

A primary goal of existing algorithms and software to analyze chaotic time series is to allow experimentalists to characterize their data using many of the same tools that theoreticians use to characterize their equations. Under mild hypotheses, it is possible to embed the data to reconstruct an attractor from a time series of measurements [1,2]. The embedding can be used to generate piecewise linear approximations of the dynamics on the attractor. Usually, no analytical form for the function $\mathbf{f}(\mathbf{x})$ that maps a given point $\mathbf{x}$ to its observed image is available, but often it may be approximated as $\mathbf{f}(\mathbf{x})=\mathbf{A}\mathbf{x}+\mathbf{b}$, where the matrix $\mathbf{A}$ and vector $\mathbf{b}$ are obtained by a linear least-squares computation using the points in a suitably small neighborhood of $\mathbf{x}$ [3]. The approximations, in turn, can be used to estimate Lyapunov exponents [4], make short-term predictions [5], or reduce the noise in the observations [6].

It is well known that such computations must be done with care. There are several sources of systematic error, including a poor choice of time delay [7]; noise in the data, which introduces a bias in ordinary least-squares estimators [8]; and a lack of observations [9]. Moreover, embeddings can produce spurious Lyapunov exponents [10]. Nevertheless, with appropriate precautions, one can make piecewise approximations of the dynamics and use them to estimate quantities like Lyapunov exponents. Yet a basic question remains: how can one estimate the uncertainty in the approximations?

This paper describes several applications of a statistical method, called *bootstrap sampling*, to quantify the uncertainty in estimates of dynamical information from data. Section II describes the bootstrap method. Section III outlines a simple, effective, automatic algorithm that uses bootstrap sampling to identify small sets of influential points in least-squares estimates of the local dynamics on a reconstructed attractor. (A point $\mathbf{x}$ is influential if the regression coefficients are especially sensitive to small changes in the coordinates of $\mathbf{x}$.) Section IV shows how bootstrap sampling can be used to estimate the variance in regression coefficients computed from total least-squares methods (also called error-in-variables methods). Section V considers the inherent difficulties in devising confidence intervals for estimates of Lyapunov exponents from embedded chaotic attractors and suggests a heuristic measure of the relative uncertainty using bootstrap estimates of the standard error in one-step expansion rates.

## II. BACKGROUND

*Bootstrap estimates of the standard error*. A typical problem in mathematical statistics is the estimation of some parameter $\theta=\theta(P)$ of a probability distribution $P$. Even if $P$ is known explicitly, formulas for the standard error (standard deviation) of $\theta$ are difficult to derive except in simple cases (e.g., when $\theta$ is the mean of the distribution). Often, $P$ is not known explicitly but is taken as the empirical distribution defined by a given set of observations. The bootstrap is a computational procedure that provides reliable estimates of the standard error (se) of $\theta$, often denoted $\mathrm{se}\,\theta$ or $\sigma(\theta)$, provided that the available data is a reasonable approximation of the underlying probability distribution [11,12].

Let $\mathbf{x}=(x_1,x_2,\ldots,x_n)$ be a collection of $n$ observations that is drawn from an underlying probability distribution $P$ (that may or may not be known explicitly). Let $s(\mathbf{x})$ be a statistic to estimate the parameter $\theta=\theta(P)$. A *bootstrap sample* is a collection $\mathbf{x}^*$ consisting of $n$ data points drawn independently from the original observations at random with replacement and with probability $1/n$. (Each observation is equally likely to be chosen, and a given observation may be chosen once, more than once, or not at all.) The quantity $\theta^*=s(\mathbf{x}^*)$ is a *bootstrap replicate* of $\theta$.

The bootstrap algorithm for estimating the standard error $\sigma(\theta)$ is straightforward. Using a random number generator, one generates $B$ bootstrap samples $\mathbf{x}_i^*$, each of which is used to compute a *bootstrap replicate* $\theta_i^*=s(\mathbf{x}_i^*)$, $i=1,\ldots,B$. The *bootstrap mean* of $\theta$ is the mean of the bootstrap replicates,

$$\mu_B(\theta)=\frac{1}{B}\sum_{i=1}^{B}\theta_i^*. \tag{1}$$

The *bootstrap standard error* of $\theta$ is

 

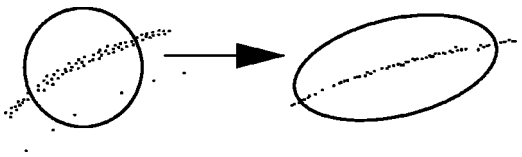

FIG. 1. A schematic illustration of influential points.

$$\sigma_B(\theta)=\left[\frac{1}{B-1}\sum_{i=1}^{B}[\theta_i^*-\mu_B(\theta)]^2\right]^{1/2}. \tag{2}$$

[Equations (1) and (2) are sometimes called the *nonparametric* bootstrap estimates of the mean and standard error, respectively, because they are based on the empirical distribution of the data.] The *ideal* bootstrap estimate of $\sigma(\theta)$ is $\lim_{B\to\infty}\sigma_B(\theta)$. Values of $B$ between 25 and 200 usually yield satisfactory approximations of $\sigma(\theta)$ for most statistics of practical interest [11].

*Eckmann-Ruelle linearization*. In a typical laboratory experiment, a time series of observations $t_i$, $i=1,2,\ldots$ is obtained at equal intervals $\Delta t$. Under mild hypotheses [1,2], an attractor can be reconstructed by generating an appropriate sequence of vectors from the time series. One popular method is time-delay embedding, where one fixes the embedding dimension $m$ and time delay $\tau$ and forms vectors $\mathbf{x}_i=(t_i,t_{i+\tau},t_{i+2\tau},\ldots,t_{i+(m-1)\tau})$, $i=1,2,\ldots$. (Other approaches are also possible [1].) Whatever the method used to generate the vectors $\mathbf{x}_i$, each vector is regarded as a function of the previous one in time order: $\mathbf{x}_{i+1}=\mathbf{f}(\mathbf{x}_i)$.

Although $\mathbf{f}$ is not known in practice, Eckmann and Ruelle [3] suggested a local linear approximation:

$$\mathbf{x}_{i+1}=\mathbf{f}(\mathbf{x}_i)\approx\mathbf{A}_i\mathbf{x}_i+\mathbf{b}_i, \tag{3}$$

where $\mathbf{A}_i$ is an $m\times m$ matrix and $\mathbf{b}_i$ is an $m$ vector. (The subscript emphasizes the dependence of $\mathbf{A}$ and $\mathbf{b}$ on $\mathbf{x}_i$.) Equation (3) is the *Eckmann-Ruelle linearization* of $\mathbf{f}$ at $\mathbf{x}_i$. Both $\mathbf{A}$ and $\mathbf{b}$ can be computed using linear least squares if a sufficiently large number of observations can be found in a suitably small neighborhood of $\mathbf{x}_i$. A collection of Eckmann-Ruelle linearizations can be used to estimate the Lyapunov exponents of the attractor [4]; reduce the noise in the data [6]; make short-term predictions of the underlying chaotic process [5,13]; or apply small controls to keep the dynamics near a given saddle orbit [14].

## III. INFLUENTIAL POINTS IN LINEAR LEAST SQUARES

An important question is how the estimates of $\mathbf{A}$ and $\mathbf{b}$ depend on the observations. Consider a distribution of points like the one illustrated schematically in Fig. 1. The points lie along two strips (manifolds) in the reconstructed attractor. One of them has a much larger natural measure than the other (as suggested by the thicker band). A small ball centered about one of the points encompasses a large number of points along the more probable strip and, in this example, only two points along the less probable strip. The dynamics maps the ball approximately into an ellipse after one iteration.

The corresponding Eckmann-Ruelle linearization is a $2\times 2$ matrix $\mathbf{A}$ that is computed from the data points within the original ball and their corresponding images in the ellipse. The singular values $s_1$ and $s_2$ of $\mathbf{A}$ determine the lengths of the major and minor axes, respectively, of the ellipse [15]. In this example, $s_1$ is determined from a large number of data points, but to within the observational accuracy, the value of $s_2$ is determined almost exclusively by the two points on the less probable strip. (An example of this situation in the Hénon map is presented below.) Hence the estimate of $s_2$ is particularly sensitive to small errors in the measurement of these two points; they are called *influential points* for this reason.

An *outlier* is a point that is relatively distant from the mean of the observations; outliers sometimes result from particularly large errors in measurement. An influential point need not be distant from the mean, but small changes in an influential point have much bigger effects on the regression function relative to small changes in the other observations. However, an outlier need not be an influential point, and an influential point need not be an outlier [8]. Moreover, a small subset of points may form an influential subset, as illustrated in Fig. 1.

Most statistics texts recommend a graphical analysis of the data when fitting a regression function, as the presence of outliers and influential points often is evident from a plot. However, graphical analysis is practical only when there is a small number of independent variables and a small number of regression functions to be fitted. Many heuristic statistical methods, largely based on an analysis of residuals, have been proposed to identify influential points when graphical analysis is impractical [16]. One difficulty with the heuristic methods in the case of chaotic data analysis is that the points with the greatest influence on the Eckmann-Ruelle matrix need not have large residuals.

The bootstrap is a useful way to identify subsets of influential points when computing Eckmann-Ruelle linearizations of the local dynamics. Consider the situation in Fig. 2, which shows a set of points in a neighborhood of the Hénon attractor and their corresponding images. The attractor is a time-delay reconstruction, and in this case,

$$\mathbf{f}\begin{pmatrix}t_{i-1}\\ t_i\end{pmatrix}=\begin{pmatrix}t_i\\ 1.4-t_i^2+0.3t_{i-1}\end{pmatrix}. \tag{4}$$

The neighborhood in Fig. 2(a) shows observations that lie in one of two main bands; each band corresponds to closely spaced pieces of the chaotic attractor. The dynamics squeezes the two bands closer together. The computed estimate

$$\mathbf{A}=\begin{pmatrix}0 & 1\\ 0.3009 & -0.9455\end{pmatrix} \tag{5}$$

is an accurate approximation of ''true'' Jacobian matrix $\mathbf{Df}$ at the reference point $\mathbf{x}_{\mathrm{ref}}=(0.7864,0.4721)$. We generate $B=100$ bootstrap samples and compute a bootstrap replicate $\mathbf{A}^*$ from each. We apply Eq. (2) elementwise to obtain the bootstrap standard error,

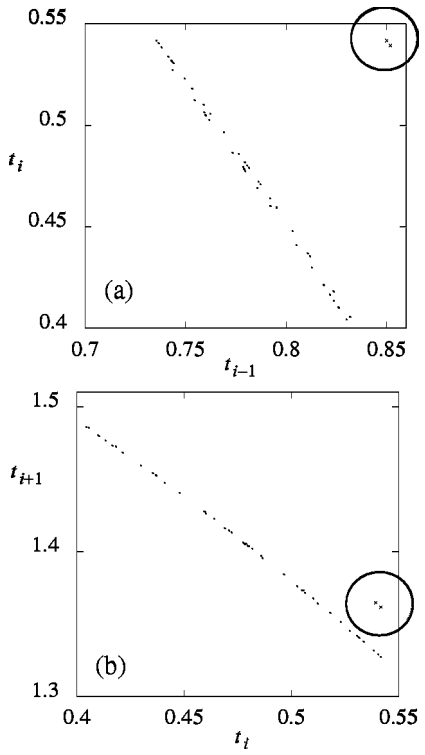


FIG. 2. (a) Observations in a typical neighborhood of the Hénon attractor, and (b) their images. The circles show an influential subset.

$$\sigma_{100}(\mathbf{A}) = \begin{pmatrix} 0 & 0 \\ 0.1379 & 0.0920 \end{pmatrix}.$$

The bootstrap mean and standard errors of the singular values of $\mathbf{A}$ are $s_1 = 1.389 \pm 0.036$ and $s_2 = 0.232 \pm 0.113$. Similarly, the bootstrap standard error of $\det \mathbf{A}$ is about 28% of its true value, 0.3.

The standard errors in $s_2$ and $\det \mathbf{A}$ are more than ten times larger than those in typical neighborhoods on the attractor [17]. They reflect the uncertainty in the local contraction rate due to the small number of points used to estimate it: there is an influential subset consisting of the two points at the upper right of Fig. 2(a). Within the measurement accuracy, all the information about the local contraction rate comes from the two circled points. In bootstrap samples where one or both points are included, the computed value of $\mathbf{A}^*$ is close to that in Eq. (5). However, in bootstrap samples where both points are omitted, typically

$$\mathbf{A}^* \approx \begin{pmatrix} 0 & 1 \\ 0.7432 & -0.6312 \end{pmatrix}.$$

The omission of the points in the influential subset changes the estimate of $\mathbf{A}$ considerably, even though the means of the observations are nearly identical.

The second derivative of the Hénon map is constant, so

$$\frac{\|\mathbf{Df}(\mathbf{x}) - \mathbf{Df}(\mathbf{x}^*)\|}{\|\mathbf{x} - \mathbf{x}^*\|} = 2.$$

In contrast, the observations in this neighborhood yield values of $\|\mathbf{A} - \mathbf{A}^*\| / \|\mu(\mathbf{x}) - \mu(\mathbf{x}^*)\|$ that are as large as 400. The presence of the influential subset implies that the computed approximation of the dynamics does not vary continuously with the observations.

Whenever exceptional estimates of the Eckmann-Ruelle matrix are encountered, one can inspect the corresponding bootstrap samples to identify influential subsets of small numbers of points. Because each point in a bootstrap sample of size $n$ is chosen independently with probability $1/n$, the probability that a particular observation $\mathbf{x}_j$ does *not* appear in a given bootstrap sample is $(1 - 1/n)^n \sim 1/e \approx 0.37$ for modest values of $n$. If two points form an influential subset, then the probability that neither is chosen in a given bootstrap sample is approximately $1/e^2$, or 14%. Hence a modest number of bootstrap replicates is likely to make apparent the influence of the two points on the estimation of $\mathbf{A}$.

This observation underlies following algorithm to identify influential points in an Eckmann-Ruelle linearization.

*Algorithm I.* Let a set of observations $S$ in a small neighborhood on the attractor be given.

(1) Generate $B$ bootstrap samples from $S$.

(2) From each sample, compute a bootstrap replicate $\mathbf{A}^*$ using linear least squares.

(3) Compute the bootstrap mean $\mu_B(\mathbf{A})$ and the bootstrap standard error $\sigma_B(\mathbf{A})$.

(4) For each bootstrap replicate $\mathbf{A}^*$, determine whether $\|\mathbf{A}^* - \mu_B(\mathbf{A})\| \geqslant \|\sigma_B(\mathbf{A})\|$. (The Frobenius norm [19] is suitable for this purpose, and it is less expensive to compute than the spectral norm.)

(5) There are $k \geqslant 0$ bootstrap replicates that satisfy the criterion in step 4. If $k > 0$ then let $E$ be the subset of $S$ whose elements did *not* appear in any of the $k$ corresponding bootstrap samples. If $k = 0$ then let $E$ be the empty set.

Because the points in $E$ do not appear in any bootstrap sample that yields an exceptional map, $E$ may be an influential subset. (If $E$ is empty, then there are no influential points by the criterion in step 4.) The probability that a noninfluential point appears in any given exceptional sample (by the criterion in step 4) is about $1 - e^{-1}$ or 63%; hence if $B$ is too small, then algorithm I may generate a set $E$ that contains many noninfluential observations. One useful heuristic is to choose $B$ so that if an influential subset does exist, then the expected number of exceptional samples identified in step 4 is at least 12. The probability that a noninfluential point appears in all 12 exceptional samples is $(1 - e^{-1})^{12} \approx 0.0041$. If $S$ contains, say, 200 observations, then the expected number of noninfluential points in $E$ is less than 1. This heuristic can reliably identify any single influential point if $B \geqslant 33$; any pair of influential points if $B \geqslant 89$; and any trio of influential points if $B \geqslant 241$.

What does one do with influential points? The answer depends in part on the nature of the underlying data. If one has data from a numerical simulation or other source for which the observational noise is negligible, then influential points might as well be retained in the least-squares computation. If the noise is appreciable, then the presence of influential points means that at least some aspect of the estimated

dynamics relies on a small number of observations, and thus is subject to a relatively large uncertainty. In the Hénon example above, it would be appropriate to discard the influential points if noise were appreciable and compute the best rank-1 approximation to the dynamics; this poses no difficulty for an application that seeks to estimate the largest Lyapunov exponent.

The fractal structure of chaotic attractors and the large variations in the natural measure throughout the attractor mean that influential points arise relatively frequently in practice. In addition, laboratory data often contain isolated ''glitches'' that can be identified using bootstrap sampling. Algorithm I can be applied to any class of models for the dynamics, such as higher-order Taylor polynomials [20] or radial basis functions [21]. Although algorithm I requires significant computational effort, modern computers make its implementation practical; local linear approximations to each point in a data set of $10^5$ points can be computed in 10–60 min for embedding dimensions up to 6 or so.

## IV. BOOTSTRAP ESTIMATES OF THE ERROR IN TOTAL LEAST SQUARES

The presence of measurement noise complicates the statistical estimation of $\mathbf{A}$ and $\mathbf{b}$ even in the absence of influential points. One difficulty is that the estimator for the coefficients of $\mathbf{A}$ is biased, and the size of the bias depends on the noise, not on the number of observations. *Total least squares* provides an alternative approach that reduces the bias [22]; its use was suggested in Ref. [8] as a possible way to enhance the accuracy of Lyapunov exponents computed from Eckmann-Ruelle linearizations. Although the total least-squares algorithm reduces the bias, the application of total least squares to Eckmann-Ruelle linearization appears to give estimators with a significantly larger variance than in ordinary least squares.

In ordinary linear least squares, the observations $\mathbf{x}_i$ and $\mathbf{y}_i$ are assumed to satisfy relations of the form

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b} + \boldsymbol{\epsilon}_i , \tag{6}$$

$i=1,\ldots,n$, where $\mathbf{x}_i$ and $\mathbf{y}_i$ are $m$ vectors of observations, $\mathbf{A}$ is an $m\times m$ matrix, and the $\boldsymbol{\epsilon}$'s are $m$ vectors of independent, random error terms of mean $\mathbf{0}$ and common variance. All of the error in the model (6) occurs in the measurement of $\mathbf{y}_i$.

Without loss of generality, we may subtract off the mean of the observations $\mathbf{x}_i$ and $\mathbf{y}_i$ and re-cast the least squares problem in a matrix form as follows. Let $\mathbf{Y}_0$ be the $n\times m$ matrix whose $i$th row is $\mathbf{y}_i-\mu(\mathbf{y})$, and let $\mathbf{X}_0$ be the $n\times m$ matrix whose $i$th row is $\mathbf{x}_i-\mu(\mathbf{x})$. Equation (6) is equivalent to the overdetermined system

$$\mathbf{Y}_0 = \mathbf{X}_0\mathbf{A}^{\mathrm{T}}. \tag{7}$$

(We assume $n>m$.) The ordinary least-squares algorithm attempts to find a matrix $\hat{\mathbf{Y}}_0$ that minimizes $\|\mathbf{Y}_0-\hat{\mathbf{Y}}_0\|_F$ subject to the constraint that each column of $\hat{\mathbf{Y}}_0$ is in the column space of $\mathbf{X}_0$ [19]. Once such a $\hat{\mathbf{Y}}_0$ is found, then any matrix $\hat{\mathbf{A}}^{\mathrm{T}}$ that satisfies $\hat{\mathbf{Y}}_0=\mathbf{X}_0\hat{\mathbf{A}}^{\mathrm{T}}$ is called an *ordinary least-squares* solution of Eq. (7) [23]. The Eckmann-Ruelle linearization (3) is obtained by setting $\mathbf{A}=(\hat{\mathbf{A}}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{b}=\mu(\mathbf{y})-\mathbf{A}\mu(\mathbf{x})$. One can regard the ordinary least-squares approach as finding the smallest adjustment to the observation matrix $\mathbf{Y}$ so that Eq. (7) has a solution.

It can be shown that the matrix

$$\hat{\mathbf{A}}^{\mathrm{T}} = (\mathbf{X}_0{}^{\mathrm{T}}\mathbf{X}_0)^{-1}\mathbf{X}_0{}^{\mathrm{T}}\mathbf{Y}_0 \tag{8}$$

provides a least-squares solution of Eq. (7) [18]. The model (6) implies that $\mathbf{Y}_0$, and hence $\hat{\mathbf{A}}^{\mathrm{T}}$, is a random matrix. The estimator (8) is said to be an *unbiased* estimator of $\mathbf{A}^{\mathrm{T}}$: the expectation of $\hat{\mathbf{A}}^{\mathrm{T}}$ is the transpose of the underlying ''true'' matrix $\mathbf{A}$ in the model (6).

Now consider the time-delay reconstruction of an attractor from a chaotic time series whose values are contaminated with measurement noise. (Assume that the errors are independent and identically distributed with mean $\mathbf{0}$ and a common covariance matrix that is diagonal.) In this case, there is error in all the observations, so the underlying model in the Eckmann-Ruelle linearization (3) is

$$\mathbf{y}_i = \mathbf{A}(\mathbf{x}_i+\boldsymbol{\delta}_i)+\mathbf{b}+\boldsymbol{\epsilon}_i , \tag{9}$$

where the $\boldsymbol{\delta}$'s and $\boldsymbol{\epsilon}$'s are vectors of independent and identically distributed measurement errors. In this case, the estimator (8) is a *biased* estimator of $\mathbf{A}^{\mathrm{T}}$. The size of the bias depends on the covariance matrix of the error terms and is independent of the number of observations [23–25].

The total least-squares algorithm adjusts both sets of observations and finds parameters $\mathbf{A}$ and $\mathbf{b}$ so that the relation $\mathbf{y}=\mathbf{A}\mathbf{x}+\mathbf{b}$ holds exactly for each adjusted observation. Let $[\mathbf{X}_0;\mathbf{Y}_0]$ denote the augmented $n\times 2m$ observation matrix; the first $m$ elements of the $i$th row are $\mathbf{x}_i-\mu(\mathbf{x})$ and the second $m$ elements are $\mathbf{y}_i-\mu(\mathbf{y})$. The total least-squares problem seeks to find a new augmented matrix $[\hat{\mathbf{X}}_0;\hat{\mathbf{Y}}_0]$ such that the quantity

$$\|[\mathbf{X}_0;\mathbf{Y}_0]-[\hat{\mathbf{X}}_0;\hat{\mathbf{Y}}_0]\|_F$$

is minimized, subject to the constraint that the column space of $\hat{\mathbf{Y}}_0$ is contained in the column space of $\hat{\mathbf{X}}_0$. If such a minimizing matrix can be found, then any matrix $\hat{\mathbf{A}}^{\mathrm{T}}$ that solves $\hat{\mathbf{Y}}_0=\hat{\mathbf{X}}_0\mathbf{A}^{\mathrm{T}}$ is a total least-squares solution of Eq. (7) [23]. We obtain the Eckmann-Ruelle linearization (3) by setting $\mathbf{A}=(\hat{\mathbf{A}}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{b}=\mu(\hat{\mathbf{y}})-\mathbf{A}\mu(\hat{\mathbf{x}})$, where $\mu(\hat{\mathbf{x}})$ and $\mu(\hat{\mathbf{y}})$ are the means of the adjusted $\mathbf{x}$ and $\mathbf{y}$ observations, respectively. A solution of the total least-squares problem can be formulated as an iterative nonlinear minimization problem [22] or, in the case of Eq. (9), in terms of the singular value decomposition of $[\mathbf{X}_0;\mathbf{Y}_0]$ [23].

The total least-squares solution has some desirable asymptotic properties. It can be shown that the total least-squares estimator of the matrix $\mathbf{A}$ is less biased than the ordinary least-squares estimator and that the bias tends to zero as the number of observations $n$ tends to infinity. (See Ref. [23] for details.)

However, the behavior of the total least-squares estimator is harder to characterize for intermediate values of $n$. The total least-squares algorithm estimates new observations $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ as well as a matrix of coefficients, $\hat{\mathbf{A}}$, such that $\hat{\mathbf{Y}} = \hat{\mathbf{X}}\hat{\mathbf{A}}^{\mathrm{T}}$. Thus total least squares has more degrees of freedom than ordinary least-squares regression. As a result, one may expect that total least-squares estimators of the regression parameters have larger variance than in the ordinary least-squares case.

There is no convenient expression for the standard error of the parameter matrix $\hat{\mathbf{A}}$ in total least squares, but we can estimate it using bootstrap sampling. As an illustration, we consider the application of the total least-squares algorithm to the collection of data shown in Fig. 2. One hundred bootstrap samples are drawn and the total least-squares algorithm is applied to each. The bootstrap mean of the matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0.3084 & -0.9405 \end{pmatrix},$$

which is close to the ordinary least-squares estimate in Eq. (5). However, the bootstrap standard error is

$$\sigma_{200}(\mathbf{A}) = \begin{pmatrix} 0 & 0 \\ 1.115 & 0.7765 \end{pmatrix},$$

which is more than ten times larger than that of the ordinary least-squares estimator using the same data.

Numerical experiments using bootstrap sampling suggest that total least-squares estimators have large variance when they are applied to Eckmann-Ruelle linearizations. When the number of observations is limited ($n$ is between 25 and 200 for typical laboratory data sets), the total least-squares algorithm appears to be much more sensitive to outliers and to influential points than ordinary least squares. If the measurement noise is significant, then it is probably better to preprocess the data using a filtered embedding [13] or a nonlinear noise reduction method [6] than to employ total least squares on unfiltered data. Alternatively, if independent information about the measurement noise is available, then one can attempt to apply a bias correction to the ordinary least-squares estimator [24].

## V. THE UNCERTAINTY IN LYAPUNOV EXPONENT CALCULATIONS

How can one assess the uncertainty in a Lyapunov exponent calculation from time series data? Every algorithm for computing Lyapunov exponents relies on an estimate of the linearized dynamics in a neighborhood of every point on the reconstructed attractor. As discussed above, there are many sources of error in such estimates. We argue that it is not particularly useful to treat these errors as random variables. Instead, we suggest that the uncertainties in the local linearizations be treated as *intrinsic* errors, and we describe a heuristic procedure to quantify this intrinsic uncertainty. Numerical experiments suggest that this heuristic assessment of the uncertainty agrees reasonably well with the range of Lyapunov exponent estimates obtained from different embeddings of time series data.

The distinction between intrinsic and random errors is important when one wants to bound the uncertainty in a calculation. Consider a set of measurements $\{\theta_i\}_{i=1}^n$ that approximate some unknown quantity $\Theta$. If the measurement errors $\epsilon_i$ are *random*, then we model the observations as $\theta_i = \Theta + \epsilon_i$, where the $\epsilon$'s are random variables chosen from some probability distribution. If, for instance, the $\epsilon$'s are assumed to be independent and identically distributed with mean 0 and variance $\varepsilon^2$, then a reasonable estimate of $\Theta$ is the mean of the observations, $\overline{\theta} = \Sigma_{i=1}^n \theta_i / n$, whose standard deviation is

$$\sigma(\overline{\theta}) = \frac{\varepsilon}{\sqrt{n}}. \tag{10}$$

Thus the uncertainty in $\overline{\theta}$ decreases as more observations become available.

A contrasting situation occurs for *intrinsic* errors. Suppose that $\{\theta_i\}_{i=1}^n$ is a set of real numbers of comparable magnitudes, stored as floating-point values in a computer. Consider the uncertainty in their computed mean. Typically, intermediate results are rounded toward zero about as often as they are rounded away from zero, but it does not follow that the rounding errors approximately cancel out and therefore can be ignored. The floating-point representation of each $\theta_i$ has an intrinsic uncertainty $\varepsilon_i$, due to the finite precision. If each $\theta_i$ is of comparable magnitude, then each $\varepsilon_i$ is roughly equal to a common value $\varepsilon$. The uncertainty in the computed mean is approximately

$$\sum_{i=1}^{n} \varepsilon_i / n = \sum_{i=1}^{n} \varepsilon / n = \varepsilon. \tag{11}$$

In contrast to the random error model leading to Eq. (10), the availability of more observations does not decrease the uncertainty in the floating-point estimate of the mean.

Benettin's algorithm, described below, estimates the Lyapunov exponents on an attractor as the geometric mean of local expansion rates computed along each point of a given orbit. We suggest that bootstrap sampling can be used to estimate the uncertainty in the local expansion rates, and propose a measure, analogous to that in Eq. (11), for quantifying the uncertainty in the estimated Lyapunov exponents that arises from sampling variability on the attractor. Although the procedure lacks a rigorous justification, preliminary numerical results are promising.

We recall some basic definitions. Let $\mathbf{f}$ be a diffeomorphism of an $m$-dimensional manifold $M$ to itself. Let $\mathbf{Df}^k(\mathbf{x}_0)$ denote the derivative of the $k$-times iterated map, $\mathbf{f}^k$, evaluated at $\mathbf{x}_0$; by the chain rule, $\mathbf{Df}^k(\mathbf{x}_0) = \mathbf{Df}(\mathbf{x}_{k-1}) \cdots \mathbf{Df}(\mathbf{x}_0)$, where $\mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i)$. For each point $\mathbf{x} \in M$ and vector $\mathbf{v} \in T_{\mathbf{x}}M$ (the space of tangent vectors to $\mathbf{x}$), define

$$\lambda(\mathbf{x}, \mathbf{v}) = \lim_{k \to \infty} \frac{1}{k} \log \|\mathbf{Df}^k(\mathbf{x})\mathbf{v}\| \tag{12}$$

whenever the limit exists. It can be shown that there are at most $m$ distinct values of $\lambda(\mathbf{x},\mathbf{v})$, called the *Lyapunov exponents* of $\mathbf{f}$ at $\mathbf{x}$. If $\mathbf{f}$ is ergodic for some invariant measure $\mu$, then the Lyapunov exponents are the same for $\mu$-almost every $\mathbf{x}$. The Oseledec multiplicative ergodic theorem states conditions under which the Lyapunov exponents are guaranteed to exist and gives a splitting of $T_{\mathbf{x}}M$ into orthogonal subspaces $E_1(\mathbf{x}),\ldots,E_m(\mathbf{x})$ such that $\lambda(\mathbf{x},\mathbf{v})=\lambda_j$ for $\mathbf{v}\in E_j(\mathbf{x})$. (See Ref. [26] for details.)

Roundoff error makes a direct evaluation of $\mathbf{Df}^k$ impractical for large values of $k$. Benettin *et al.* [27] have developed an iterative numerical procedure, summarized in algorithm B, that estimates the Lyapunov exponents for a mapping of the form $\mathbf{x}_{k+1}=\mathbf{f}(\mathbf{x}_k)$ by computing one-step expansion rates $\alpha$ along an appropriate set of orthonormal subspaces. The algorithm is repeated for $N$ steps, where $N$ is as large as practical; the initial condition is $\mathbf{x}_0$. To simplify the notation, the left arrow ($\leftarrow$) indicates that the quantity on the left is overwritten by the quantity on the right. The $\mathbf{v}$'s and $\mathbf{w}$'s are temporary vectors that are redefined on every iteration.

*Algorithm II.*

Step 0. Set $k\leftarrow 0$ and let $\mathbf{u}_1,\ldots,\mathbf{u}_m$ be an arbitrary set of orthonormal $m$ vectors, called the *Lyapunov basis*.

Step 1. Set $\mathbf{w}_j\leftarrow\mathbf{Df}(\mathbf{x}_k)\mathbf{u}_j$, $j=1,\ldots,m$.
Step 2. Let $\alpha_1(k+1)=\|\mathbf{w}_1\|$. Set $\mathbf{u}_1\leftarrow\mathbf{w}_1/\alpha_1(k+1)$.
Step 3. For $j=2,\ldots,m$:
(a) Set $\mathbf{v}_j\leftarrow\mathbf{w}_j-\sum_{i=1}^{j-1}\langle\mathbf{w}_j,\mathbf{w}_i\rangle\mathbf{w}_i$.
(b) Let $\alpha_j(k+1)=\|\mathbf{v}_j\|$.
(c) Set $\mathbf{u}_j\leftarrow\mathbf{v}_j/\alpha_j(k+1)$.
(The angle brackets denote the standard inner product. The Gram-Schmidt orthonormalization should be replaced by a QR decomposition if $m$ is larger than 5 or so.)
Step 4. Set $k\leftarrow k+1$. If $k<N$ then go to step 1.

The $j$th Lyapunov exponent is estimated as

$$\lambda_j=\frac{1}{N}\sum_{i=1}^{N}\log\alpha_j(i). \tag{13}$$

The $\alpha$'s are the *one-step* expansion rates; i.e., $\alpha_i(k)$ is the local expansion rate at $\mathbf{x}_k$ along the subspace spanned by the $i$th Lyapunov vector. Some of the initial one-step expansion rates may be omitted to allow transients to decay; in that case, the sum in Eq. (13) runs from an appropriate $N_0$ to $N$, and the factor $1/N$ is replaced by $1/(N-N_0+1)$. Sometimes the renormalization of the $\mathbf{u}$'s is done after every $p>1$ iterations, in which case $\mathbf{Df}$ is replaced by the derivative of the $p$-times iterated map $\mathbf{Df}^p$, and the factor $1/N$ in Eq. (13) is replaced by $1/(Np)$. When Benettin's algorithm is applied to an embedding, the Eckmann-Ruelle matrix $\mathbf{A}_k$ computed at $\mathbf{x}_k$ is used as an approximation of $\mathbf{Df}(\mathbf{x}_k)$ [or of $\mathbf{Df}^p(\mathbf{x}_k)$, as appropriate].

A key question is how to quantify the uncertainty in the estimates of the Lyapunov exponents using Eq. (13) when the tangent maps of the underlying dynamical system are estimated with Eckmann-Ruelle linearizations. Golia and Sandri [28] have suggested a resampling technique to generate new time series from an observed one in such a way that the new time series can be considered products of the same underlying dynamical system. They perform a Lyapunov exponent calculation on each resampled time series, then use the variance of the computed exponents as an estimate of the uncertainty. However, the resampling method is valid only if the generated time series can be shown to shadow the original one sufficiently closely [28], which may be difficult to do if the underlying dynamical system is not known explicitly or if observational noise is appreciable. Gençay [29] has proposed a bootstrap method that relies, in part, on products of random permutations of the Jacobian matrices for approximating the limit (12). Ziehmann *et al.* [30] have argued that Gençay's approach is inappropriate if the matrix products in Eq. (12) do not commute.

The premise behind efforts to construct a confidence interval for Lyapunov exponents is that the errors in estimating the local dynamics arise primarily from random quantities. While noise in the observations is a source of randomness, significant uncertainties in the estimated dynamics can arise even when the observational error is negligible (as may occur in time series from numerical simulations).

For example, spurious Lyapunov exponents can arise as an artifact of the embedding. Consider a time-delay embedding of a time series from the Hénon map in a five-dimensional phase space. The embedding dimension guarantees a diffeomorphism between the original attractor and the reconstructed one (because 5 is more than twice the box dimension of the Hénon attractor [1]). At each point, the Eckmann-Ruelle linearization produces a $5\times 5$ matrix. Algorithm B yields five estimated Lyapunov exponents, three of which are spurious. The difficulty arises because the Eckmann-Ruelle matrix in a 5-dimensional embedding is not the tangent map of the underlying dynamical system, even though it consistent with the dynamics. The spurious exponents arise as part of a deterministic process [10]. (The Appendix describes a heuristic method that avoids, or at least minimizes, the computation of spurious exponents, but the point here is that no probability model accounts for the appearance of spurious Lyapunov exponents in time series embeddings.)

Systematic errors in local estimates of the dynamics also arise from the natural measure of the attractor. Some regions of the attractor may be so sparsely populated that it is impossible to obtain enough observations in a given neighborhood to compute an Eckmann-Ruelle linearization. If the neighborhood is enlarged to include more points, then nonlinearities may become significant and Eq. (3) may be a poor approximation to the local dynamics. One can fit a quadratic model instead, but in an $m$-dimensional embedding, a quadratic model requires the estimation of up to $(m^3+m^2)/2$ more parameters than for the linear model (3), and their variance is larger than in a linear model. Alternatively, there may be many observations, but due to the strongly volume-contracting nature of the dynamics, all the manifolds in a neighborhood lie on a low dimensional hyperplane to within the observational accuracy. If there are few, if any, observations in transverse directions from which to estimate the local contraction rate, then the Eckmann-Ruelle linearization

TABLE I. Lyapunov exponents computed from time series of the Hénon map.

| Noise | Estimated exponent | Range | $u(\lambda)$ |
|---|---|---|---|
| $\eta=0$ | $\lambda_1=0.632\pm0.014$ | [0.594,0.662] | 0.008 |
| | $\lambda_2=-2.338\pm0.050$ | [−2.504,−2.228] | 0.302 |
| $\eta=0.005$ | $\lambda_1=0.636\pm0.013$ | [0.596,0.666] | 0.016 |
| | $\lambda_2=-2.619\pm0.089$ | [−2.876,−2.444] | 0.490 |

may be unreliable. As suggested in Fig. 1, these difficulties tend to occur in a neighborhood of every point in certain regions of the attractor.

These considerations suggest that uncertainties in local expansion rates should be regarded as intrinsic errors. For example, consider a map of the form $x_{n+1}=ax_n \bmod 1$. The Lyapunov exponent for almost every initial condition is $\lambda=\log a$. In any numerical simulation, roundoff error yields a common uncertainty $\varepsilon$ in the representation of $\log a$, and Eq. (11) applies. Suppose instead that $a$ is estimated from data; how might the uncertainty in $\lambda$ be quantified?

Bootstrap sampling is one possibility for quantifying the uncertainty in local expansion rates, provided that there are enough observations in a neighborhood to give a reasonable approximation to both the natural measure and to the linearized dynamics. On each iteration, the Eckmann-Ruelle matrix $\mathbf{A}_k$, computed from all the observations in a suitable neighborhood centered at $\mathbf{x}_k$, is substituted for $\mathbf{Df}(\mathbf{x}_k)$ in algorithm B. The matrix $\mathbf{A}_k$ is used to update the Lyapunov basis at every step. In addition, at each point $\mathbf{x}_k$, $B$ bootstrap replicates $\mathbf{A}_k^*$ of $\mathbf{A}_k$ are generated; steps 2 and 3 of the algorithm are repeated, substituting $\mathbf{A}_k^*$ for $\mathbf{A}_k$, to generate bootstrap replicates $\alpha_j^*(k)$ of $\alpha_j(k)$ for each $j$. The bootstrap mean and standard error of each $\alpha_j(k)$ are computed using Eqs. (1) and (2). (The bootstrap procedure can be used regardless of whether the $\mathbf{A}_k$'s are estimated from ordinary least squares, total least squares, or some other method.)

If we regard $\sigma_B(\alpha_j(k))$ as an estimate of the intrinsic local error in the expansion rate at the $j$th point on the trajectory, then the analysis leading to Eq. (11) suggests that a bootstrap measure of the uncertainty in the $j$th Lyapunov exponent associated with the trajectory is

$$u(\lambda_j)=\frac{1}{N}\sum_{k=1}^{N}\sigma_B(\alpha_j(k)). \tag{14}$$

We emphasize that Eq. (14) is *not* a confidence interval in the conventional statistical sense. However, numerical experiments suggest that it gives a reasonable assessment of the range of values of $\lambda_j$ that are computed in various cases.

Table I shows the Lyapunov exponents computed from Hénon time series as follows. One hundred initial conditions are chosen at random within the basin of attraction. For each initial condition, the Hénon map (4) is iterated to produce a time series of 2000 values (the first few iterates are discarded to remove transients). Each time series is embedded in two dimensions using a time-delay embedding. An Eckmann-Ruelle linearization is computed at each point, and Benettin's algorithm is used to estimate the Lyapunov exponents from each time series (i.e., each exponent is based on a reconstructed attractor consisting of 2000 points). The process is repeated for each of the 100 time series, giving 100 different estimates of the Lyapunov exponents. Their mean and standard error are reported in the form $\mu\pm$se in the first two rows in the second column of the table. (This procedure is analogous to that proposed by Golia and Sandri [28], except that the time series have been obtained by iterating the Hénon map from different initial conditions.) The procedure is repeated, except that uniformly distributed random noise is added to each time series to simulate observational noise with mean zero and variance equal to 0.5% of the variance of the original time series. The mean and standard error of the Lyapunov exponent estimates is reported in the third and fourth rows of the second column in Table I. (The third column of Table I gives the range of all 100 estimates.) Finally, for each time series, we generate bootstrap estimates of the uncertainty in the one-step expansion rates as described above. For each time series, we compute $u(\lambda_1)$ and $u(\lambda_2)$, as defined by Eq. (14). The fourth column in Table I shows their average values.

The ''true'' values of the exponents, based on Benettin's algorithm applied to $10^7$ iterations of Eq. (4), are $\lambda_1=0.6047$ bits/iteration and $\lambda_2=-2.342$ bits/iteration. The positive Lyapunov exponent tends to be overestimated from the embedded time series, and the negative exponent is too large (in absolute value) when estimated from the noisy time series; the discrepancies may be due to systematic errors as described above. However, the uncertainties given by Eq. (14) for $\lambda_1$ and $\lambda_2$ are representative of the range of estimates of the exponents obtained from the different time series, particularly in the presence of observational noise [31].

Table II shows estimates of the largest three Lyapunov exponents obtained from different time-delay embeddings of a laboratory time series from a Belousov-Zhabotinskii chemical reaction [32]. The units are bits per 125 time steps (the shortest period of any embedded periodic orbit within the attractor is approximately 125 times the sampling rate [33]). Benettin's algorithm has been applied using Eckmann-Ruelle linearizations from embeddings in three, four, and five dimensions and with time delays of 60, 90, and 120 time steps. (The latter time delays give comparable values of the mutual information [7], and a graphical analysis of the data suggests that a three-dimensional embedding space suffices to reconstruct the attractor.)

The parenthesized values are the uncertainties in each calculation as given by Eq. (14). The uncertainties in most cases are consistent with the range of estimates obtained from the

TABLE II. Estimates of the three largest Lyapunov exponents from a time series of the Belousov-Zhabotinskii chemical reaction. Numbers in parentheses are estimates of the uncertainty computed from Eq. (14).

| | $\lambda$ | $\tau=60$ | $\tau=90$ | $\tau=120$ |
|---|---|---|---|---|
| $m=3$ | $\lambda_1$ | 1.19 (0.35) | 1.41 (0.40) | 1.19 (0.55) |
| | $\lambda_2$ | −0.29 (0.53) | −0.29 (0.60) | −0.03 (0.73) |
| | $\lambda_3$ | −6.24 (1.80) | −3.53 (1.61) | −3.68 (1.84) |
| $m=4$ | $\lambda_1$ | 1.04 (0.24) | 0.93 (0.39) | 0.84 (0.46) |
| | $\lambda_2$ | −0.33 (0.30) | −0.06 (0.44) | −0.01 (0.54) |
| | $\lambda_3$ | −2.86 (0.76) | −2.84 (1.00) | −1.63 (0.88) |
| $m=5$ | $\lambda_1$ | 0.73 (0.19) | 0.98 (0.48) | 0.91 (0.43) |
| | $\lambda_2$ | 0.08 (0.25) | 0.14 (0.49) | 0.16 (0.50) |
| | $\lambda_3$ | −2.34 (0.48) | −1.36 (0.75) | −1.23 (0.71) |

different embeddings [34]. The uncertainties in $\lambda_1$ and $\lambda_3$ are about 25–50% of the estimated value of the corresponding exponent. This result suggests that Lyapunov exponent computations from embedded flows are subject to a larger inherent uncertainty than those from discrete maps.

Although the numerical results above are suggestive, there are several caveats.

(1) The bootstrap estimates of the standard errors of the local expansion rates are meaningful only if the underlying model [e.g., Eq. (3)] is a good approximation of the dynamics at each point. The bootstrap procedure does not account for systematic errors due to nonlinearities or to artifacts of the embedding.

(2) The bootstrap estimates may be unreliable if there are insufficient data to approximate the natural measure of the attractor in a given neighborhood. In general, the bootstrap method fails if the statistic in question depends sensitively on the tails of the underlying distribution (see e.g., Ref. [11], Chap. 7).

(3) The quantity $u(\lambda)$ is not a confidence interval in the conventional sense. Therefore the usual statistical heuristics do not apply. For instance, one cannot conclude that if the estimated value of the largest Lyapunov exponent satisfies $\lambda_1 > 3u(\lambda_1)$, then the likelihood is greater than 99% that $\lambda_1$ is positive.

Finally, and most importantly, the arguments leading to the definition of $u(\lambda)$ are not rigorous. While the numerical results in Tables I and II are promising, they are not comprehensive. More work is needed to determine whether the use of $u(\lambda)$ (or some similar quantity) can be justified theoretically.

## VI. CONCLUSIONS

Estimates of the local dynamics on a reconstructed attractor may be especially sensitive to the measured values of only a few points, even if many observations are used to fit the model. Bootstrap sampling provides a convenient way to identify such influential points. Bootstrap sampling also provides a way to characterize the uncertainty in estimates of one-step expansion rates in Benettin's algorithm for computing Lyapunov exponents. There are many sources of systematic error that make it problematic to derive a confidence interval for the estimated values of Lyapunov exponents. However, if one regards the uncertainties in one-step expansion rates as intrinsic errors (much as one treats roundoff error in interval analysis), then their average value may give a useful characterization of the reliability of Lyapunov exponents computed from a particular time series.

## ACKNOWLEDGMENTS

The author thanks Ying-Cheng Lai and Sharon Lohr for helpful discussions. This work was supported by the National Science Foundation under grant number ECS-9807529.

## APPENDIX

The computation of spurious Lyapunov exponents can be minimized if care is taken to prevent overfitting of the Eckmann-Ruelle linearization. The embedding dimension must be sufficiently large to guarantee the existence of a diffeomorphism between the reconstructed attractor and the underlying dynamics [1]. However, most of the observations in a given neighborhood on the reconstructed attractor may lie in a lower dimensional subspace. In such cases, the best rank-$r$ solution of the least-squares problem (7) should be obtained for an appropriate $r<m$.

For this purpose, we form the matrix of observations $\mathbf{X}_0$ as outlined in Sec. IV and compute its singular value decomposition [35] to obtain the singular values $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_m \geqslant 0$. As $s_i$ is the square root of the $i$th largest eigenvalue of $\mathbf{X}_0{}^{\mathrm{T}}\mathbf{X}_0$, it provides a measure of the total variance of the observations along the corresponding eigendirection (i.e., the $i$th right singular vector of $\mathbf{X}_0$).

The goal is to determine the value of $r$ such that $s_{r+1}, \ldots, s_m$ are ''negligible.'' Given $r$, we change coordinates and project the observations onto the subspace spanned by the first $r$ right singular vectors of $\mathbf{X}_0$. We fit the model (3) using least squares to obtain an Eckmann-Ruelle matrix of rank $r$, then change coordinates back to $\mathbf{R}^m$.

One criterion is to regard a singular value as negligible if it is smaller than the measurement error in each component of $\mathbf{X}_0$ [36]. Another criterion is to define $V=s_1^2+\cdots+s_m^2$, which is a measure of the total variance of the observations; then select a fraction $p$, $0<p\leqslant 1$, of the variance to be explained by the least-squares procedure for each Eckmann-Ruelle linearization. The choice of $p$ should be based on an estimate of the noise level in the observations and/or an estimate of the size of the nonlinear terms given the radius of a typical neighborhood over which the linearization is computed. For example, if the time series consists of measurements that are accurate to seven significant bits, then it is reasonable to take, say, $p=0.99$. We let $r$ be the smallest integer such that $(s_1^2+\cdots+s_r^2)/V\geqslant p$. This is the criterion that is used in all of the Eckmann-Ruelle linearizations reported here.


[1] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[2] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, Berlin, 1981), p. 366.

[3] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[4] See, for example, M. Sano and Y. Sawada, Phys. Rev. Lett. **55**, 1082 (1985).

[5] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).

[6] See, for instance, E. J. Kostelich and T. Schrieber, Phys. Rev. E **48**, 1752 (1993), and references therein.

[7] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[8] E. J. Kostelich, Physica D **58**, 138 (1992).

[9] J.-P. Eckmann and D. Ruelle, Physica D **56**, 185 (1992).

[10] T. D. Sauer and J. A. Yorke, Phys. Rev. Lett. **83**, 1331 (1999); T. D. Sauer, J. A. Tempkin, and J. A. Yorke, *ibid.* **81**, 4341 (1998).

[11] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).

[12] R. Beran and G. Ducharme, *Asymptotic Theory for Bootstrap Methods in Statistics* (University of Montreal, Centre de Reserches Mathematiques, Montreal, 1991).

[13] T. Sauer, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Addison-Wesley Publishing Company, Reading, MA, 1994), p. 175.

[14] E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).

[15] See, for example, K. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems* (Springer-Verlag, New York, 1997).

[16] S. Chatterjee and A. S. Hadi, *Sensitivity Analysis in Linear Regression* (John Wiley & Sons, New York, 1988).

[17] Bootstrap sampling is not needed to estimate the standard error of the matrix **A**, because the standard error can be written explicitly in terms of the observation matrix [18]. However, bootstrap sampling is needed to identify small influential subsets of observations.

[18] J. Neter and W. Wasserman, *Applied Linear Statistical Models* (Richard D. Irwin, Inc., Homewood, IL, 1974), Chap. 7.

[19] The Frobenius norm of $\mathbf{X}$ is $\|\mathbf{X}\|_F=(\Sigma_{i,j}x_{ij}^2)^{1/2}$.

[20] See, for example, H. D. I. Abarbanel, R. Brown, J. L. Sidorowich, and L. Sh. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).

[21] See, for example, M. J. D. Powell, in *Algorithms for Approximation*, edited by J. C. Mason and M. G. Cox (Clarendon, Oxford, 1987), p. 143.

[22] W. H. Jeffreys, Astron. J. **85**, 177 (1980).

[23] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis* (Society for Industrial and Applied Mathematics, Philadelphia, 1991).

[24] See, for example, G. A. F. Seber, *Linear Regression Analysis* (John Wiley & Sons, New York, 1977); H. Schneeweiss, Metrika **23**, 101 (1976); W. A. Fuller, *Measurement Error Models* (John Wiley & Sons, New York, 1987).

[25] Ideally, given a noisy time series $\{t_1,t_2,\ldots,t_N\}$, one would like to compute corrections $\{\eta_1,\eta_2,\ldots\eta_N\}$ such that on the attractor reconstructed from the adjusted time series $\{t_1+\eta_1,\ldots,t_N+\eta_N\}$, a local model like Eq. (3) is satisfied in every neighborhood. In principle, then, there should be a functional relationship between the vectors $\boldsymbol{\delta}_i$ and $\boldsymbol{\epsilon}_i$ in Eq. (9). However, the problem of computing a single set of time series corrections is not well posed without further conditions, and there are significant numerical difficulties as well. One approach is to try to find corrections to segments of the time series that minimize an appropriate sum of squares; see, for instance, E. J. Kostelich and J. A. Yorke, Physica D **41**, 183 (1990).

[26] See, for example, C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, Boca Raton, FL, 1995), Chap. 8, and references therein.

[27] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Meccanica **15**, 21 (1980).

[28] S. Golia and M. Sandri, Phys. Rev. Lett. **78**, 4197 (1997).

[29] R. Gençay, Physica D **89**, 261 (1996).

[30] C. Ziehmann, L. A. Smith, and J. Kurths, Physica D **126**, 49 (1999).

[31] The standard deviation of the estimates for $\lambda_2$ from the different time series is considerably smaller than the value obtained from Eq. (14). This result does not imply that the resampling procedure of Golia and Sandri is invalid, because the noisy trajectories may not be shadowed by a real trajectory of the Henon map.

[32] K. Coffman, W. D. McCormick, Z. Noszticzius, R. H. Simoyi, and H. L. Swinney, J. Chem. Phys. **86**, 119 (1987).

[33] D. P. Lathrop and E. J. Kostelich, Phys. Rev. A **40**, 4028 (1989).

[34] The symbolic dynamical analysis in Ref. [33] gives an estimate of $\lambda_1$ of approximately 0.5 bits per 125 time steps. Hence $\lambda_1$ may be consistently overestimated. The estimated value of $\lambda_3$ is anomalously negative for the three-dimensional embedding with $\tau=60$. In this case, however, it was possible to

compute a rank-3 linear estimate of the local dynamics from the available data only about half of the time, so it is likely that there is a significant systematic error in this estimate of $\lambda_3$.

[35] The singular value decomposition of the $n \times m$ matrix $\mathbf{X}$ is $\mathbf{X} = \mathbf{USV}^{\mathrm{T}}$, where $\mathbf{S} = \mathrm{diag}(s_1, s_2, \ldots, s_m)$ contains the *singular values* of $\mathbf{X}$, which are the square roots of the (necessarily real) eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$). By convention, $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_m \geqslant 0$. The matrix $\mathbf{V}$ is an $m \times m$ orthonormal matrix whose columns are the orthonormalized eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$; they are called the *right singular vectors* of $\mathbf{X}$. The columns of the $n \times m$ matrix $\mathbf{U}$ are the $m$ orthonormalized eigenvectors of $\mathbf{XX}^{\mathrm{T}}$; they are called the *left singular vectors* of $\mathbf{X}$. For more details, consult, for example, G. W. Stewart, *Introduction to Matrix Computations* (Academic Press, New York, 1973).

[36] See for instance J. J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart, *The Linpack Users' Guide* (Society for Industrial and Applied Mathematics, Philadelphia, 1979), Chap. 11.