Symbol sequence statistics in noisy chaotic signal reconstruction

X. Z. Tang* and E. R. Tracy
Physics Department, College of William and Mary, Williamsburg, Virginia 23185

A. D. Boozer
Physics Department, University of Virginia, Charlottesville, Virginia 22901

A. deBrauw†
Physics Department, Carleton College, Northfield, Minnesota 55057-4025

R. Brown
Institute for Nonlinear Science, University of California, San Diego, California 92093-0402
(Received 3 October 1994)

A method is discussed for reconstructing chaotic systems from noisy signals using a symbolic approach. The state space of the dynamical system is partitioned into subregions and a symbol is assigned to each subregion. Consequently, an orbit in a continuous state space is converted into a long symbol string. The probabilities of occurrence for different symbol sequences constitute the symbol sequence statistics. The symbol sequence statistics are easily measured from the signal output and are used as the target for reconstruction (i.e., for assessing the goodness of fit of proposed models). Reliable reconstructions were achieved given a noisy chaotic signal, provided the general class of the model of the underlying dynamics is known. Both observational and dynamical noise were considered, and they were not limited to small amplitudes. Substantial noise produces a strong bias in the symbol sequence statistics, but such bias can be tracked and effectively eliminated by including the noise characteristics in the model. This is demonstrated by the robust reconstruction of the Hénon and Ikeda maps even when the signal to noise ratio is ≈ 1. Applications of this method include extracting control parameters for nonlinear dynamical systems and nonlinear model evaluation from experimental data.

PACS number(s): 05.45.+b, 05.40.+j, 02.50.Ph

I. INTRODUCTION

In this paper we consider the inverse problem of reconstructing a model of a dynamical system from measured time-series data. Such a model might be used for prediction, or control, or many other potential applications. Constructing models from time-series data has a long history (see, e.g., Ref. [1]). However, substantial gaps in modeling capability remain.

Here we treat the following situation: the system we wish to model is low dimensional and deterministic. However, no physical system can ever be fully isolated from its environment. Therefore, real systems will always be subject to driving by various forms of dynamical noise. In addition, the observational data we are given may be polluted by substantial amounts of measurement noise. Decomposing a given signal as "chaos plus noise" is not a well-defined procedure without further constraints on the decomposition process. In most physics applications, these constraints are provided by the fact that one often has some knowledge of the class of models to be considered, both for the deterministic part and the noise (e.g., the noise is white, $1/f$, shot noise, etc.). For the decomposition of the signal into chaos plus noise to be useful the deterministic part of the model should have relatively few degrees of freedom and be as simple as possible. This can be made more precise by using, for example, a minimum description length criteria [2] or some other objective measure to avoid overfitting. In addition, the noise part of the model should be as small in amplitude as possible.

The long-term goal of the present work is to develop techniques for the analysis and modeling of chaotic systems from short and/or noisy data sets. Our work is model based and not black box based. One could be led to the choice of models either by application of some classification scheme (e.g., Refs. [3] or [4]) applied to a "mystery" signal, or by appropriate experimental design, or by taking it as a working hypothesis to be tested. We also assume that the desired model is continuous in the state space variables. In principle, the techniques described here can be used in *ab initio* reconstruction, but we do not anticipate that this will be their main area of application.

In a series of papers (see, e.g., Ref. [4] and references therein) Crutchfield and co-workers have emphasized the utility of a symbolic approach to the characterization and modeling of nonlinear systems. They show, for example, how to construct an optimal finite-state model using only

*Present address: Dept. of Applied Physics, Columbia University, 500 West 120th Street, New York, NY 10027.

†Present address: Peace Corps, P.O. Box 208, Lilongwe, Malawi.

the symbol statistics. Rechester and White [5,6] have used a symbolic approach to develop a symbolic kinetic equation as a means to estimate the underlying invariant density. In the realm of signal diagnostics (as opposed to modeling) Schwarz *et al.* [7] have used symbol statistics to analyze solar flare data. Symbolic techniques have also been used to analyze physiological signals (see, e.g., the review by Elbert *et al.* [8] and references therein). Also, a "symbol plane" approach has been developed by Cvitanović, Gunaratne, and Procaccia [9] to extract topological information from time-series data. See also [10] for an application of symbol plane analysis to experimental signals.

The interesting aspects of the present work center on the use of symbol sequence statistics as the target for the reconstruction of continuous models. In a previous Letter [11], we considered the effects of additive or observational noise, especially the high noise situation, while keeping the dynamical noise trivially small (numerical roundoff). Here we will treat both observational and dynamical noise, and neither will be limited to small amplitudes.

In the present paper we are interested in the basic nature of reconstruction using symbol sequence statistics. Therefore, we will focus on the following two questions: (1) Is reconstruction via the symbol sequence statistics robust (i.e., is it stable against fluctuations in sampling)? (2) Is the procedure convergent (i.e., is it possible to eliminate bias in the estimation of the model)?

As one would expect, we find that symbol sequence statistics are quite robust under moderate amounts of observational noise ($\leq 10\%$) or dynamical noise ($\leq 1\%$). Consequently, parameter estimation produces reliable results. Substantial observational and/or dynamical noise induces a strong bias in the symbol sequence statistics. In spite of this, the resultant bias in parameter estimation is much less dramatic and saturates as the noise amplitude increases. As we will show, by including the noise characteristics as part of the model the bias in parameter estimation can be eliminated. Therefore, symbol sequence statistics are much more robust measures for characterizing low-dimensional deterministic chaotic systems than other quantities, such as the Lyapunov exponents, the invariant density, etc.

The outline of the paper is as follows. In the next section (Sec. II) we give a brief introduction to the topic of symbol sequence statistics—how to measure them and how to use them as a target for modeling. We discuss the topic of symbolic dynamics in order to compare and contrast it with the symbol sequence statistics. Next, we introduce the error landscape and discuss its generic properties by examining the error landscapes in detail for the quadratic and Hénon maps. There are two major effects which limit the accuracy of reconstruction using the symbol sequence statistics: finite sample fluctuations and noise. In Sec. III we first discuss the effects of finiteness of the data stream and noise on one's ability to estimate the symbol sequence statistics. Once this issue has been addressed, we then consider the effects on the error landscape. Sections II and III are the foundations upon which Sec. IV is built. In Sec. IV we discuss, finally, the reconstruction of chaotic systems using the symbol sequence statistics including finite sample and noise effects. In Sec. V we end with a summary and conclusions.

II. SYMBOL SEQUENCE STATISTICS AND THE ERROR LANDSCAPE

In this section we give an introduction to the symbol sequence statistics, and compare it with the more familiar topic of symbolic dynamics. The key concept is the introduction of a discrete partition in the continuous phase space. This converts the analog signal into a symbol sequence, from which the symbol sequence statistics can be estimated. The observed symbol sequence statistics can be used as a target for measuring the goodness of fit of proposed models. This motivates the introduction of the error function $\mathcal{E}$. A plot of this function vs the model parameters constitutes the error landscape, and reconstruction of the dynamical system amounts to finding the global minimum of this landscape. We examine the error landscape for the clean quadratic and Hénon maps, and demonstrate that these landscapes are fractal.

A. Symbolic dynamics and symbol sequence statistics

Although we will not need much of the apparatus from symbolic dynamics, our approach is inspired by the power and simplicity of the ideas to be found there. Hence, for completeness, we discuss the relationship between these two techniques. Symbolic dynamics is the study of the symbolic representation of a continuous dynamical system, $f:M\mapsto M$. This is obtained by introducing a partition, $\mathbf{p}=\{P_1,P_2,\ldots,P_q\}$, which divides the state space M into q disjoint sets, each of which is labeled with a symbol $s_i\in\{1,2,\ldots,q\}\equiv S$. Consequently, the time evolution $(\mathbf{x}_0,\mathbf{x}_1,\ldots)$ of the dynamical system, $f:M\mapsto M$, is translated into a sequence of symbols labeling the partition elements visited by an orbit $s=\{s_0,s_1,\ldots\}$. After this translation the continuum dynamics f is replaced by a shift operation.

For a sufficiently long orbit, a statistical study can be implemented to estimate the probability of occurrence of different symbol sequences of various lengths. Here one must assume that the statistics is time stationary. Equivalently, we assume that all transients have died out and the system has reached its asymptotic state on an attractor in the underlying state space.

We introduce the following notation: the observed time series is denoted by $x_n=x(t_n)$ for $n=1,2,\ldots,N$. (Here we will treat only the case of scalar signals $x_n\in\mathbb{R}^1$. Generalizations are straightforward.) This signal consists of a contribution from a dynamical signal, $y_1(t_n)$, and a contribution from measurement noise $\eta(t_n)$. Thus, $x(t_n)=y_1(t_n)+\eta(t_n)$. The dynamical signal is assumed to come from a low-dimensional dynamical system that is coupled to its environment. The effect of this coupling will be modeled as dynamical noise denoted by $\mathbf{u}(t)$. Thus, the dynamical signal comes from a discrete-time map,

$$\mathbf{y}(t_{n+1})=\mathbf{F}[\mathbf{y}(t_n)]+\mathbf{u}(t_n)\ , \tag{1}$$

or a flow,

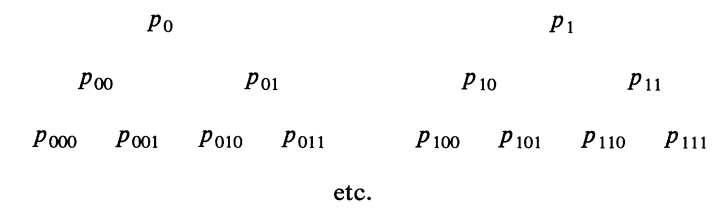
$$\frac{d\mathbf{y}}{dt}=\mathbf{F}(\mathbf{y})+\mathbf{u}(t)\ , \tag{2}$$

where $\mathbf{y}=(y_1,y_2,\ldots,y_d)$ and $\mathbf{F}$ is the low-dimensional dynamical system. Here $\mathbf{y}$, $\mathbf{F}$, and $\mathbf{u}$ are all $\in\mathbb{R}^d$ where d is the number of degrees of freedom associated with the low-dimensional dynamical system. Although $\mathbf{u}\in\mathbb{R}^d$, however, we do not assume that $\mathbf{u}$ can be generated by d (or any finite number of) degrees of freedom (e.g., $\mathbf{u}$ may represent coupling to a heat bath or thermal reservoir with many degrees of freedom which we choose not to model in detail, but only statistically as "noise"). In addition, in the present work we will consider only maps. Many flows of interest (e.g., flows bounded within a finite region of the phase space and recurrent) can often be reduced to a mapping.

To convert the analog signal stream $x(t_n)$ into a symbol sequence we pass it through a threshold function which takes

$$\{x_n\}=(x_1,x_2,\ldots,x_N)\mapsto\{s_n\}=(s_1,s_2,\ldots,s_N)\ ,$$

where $s_n\in(0,1)$. In the following work, a simple two-symbol partition was used: if $x_n<x^*$ then $s_n\equiv 0$ and if $x_n>x^*$ then $s_n\equiv 1$ with x^* chosen to be the center of mass of the data distribution. For a data sample of length N the probabilities of various symbol sequences can now be estimated. The symbol sequence statistics can be conveniently summarized by the symbol tree:



Here p_{001} is the probability of observing the sequence 001, etc.

The symbol tree is a compact summary of (coarse-grained) information about multiple-time-step correlations in the signal. Each level in the tree corresponds to a particular length of symbol sequence. Our goal is to construct a state space evolution rule [$\mathbf{F}(\mathbf{y})$] which generates the same symbol tree as the data down to some level L, where L denotes the length of the sequence under consideration (typically $L=5$ was used).

In symbolic dynamics the notion of a generating partition is of central importance. A partition $\mathbf{p}$ is generating if the partition, together with all its images and preimages, divide the state space into arbitrarily fine regions in a one-to-one manner [12]. That is to say, as the length of sequences becomes arbitrarily long a given sequence labels an individual point on the attractor [13]. Hence the symbol sequence probability $p_{s_1s_2\ldots s_L}$ is a coarse-grained description of the underlying invariant distribution. As one goes to deeper tree levels (longer sequences), this description becomes finer.

The symbolic dynamics of systems with generating partitions are particularly useful in characterizing ergodic dynamical systems [14,15]. By ergodic one means that time averages and state space averages are equivalent for any continuous function in the domain of the state space under consideration [16].

A hierarchy of entropies can be defined based upon the symbol tree. At the top of this hierarchy is the topological entropy $h^0(\mathbf{p})$ under partition $\mathbf{p}$:

$$h^0(\mathbf{p})\equiv\lim_{L\to\infty}\frac{\ln N(L)}{L}\ .$$

$h^0(\mathbf{p})$ measures the asymptotic rate at which $N(L)$, the number of different admissible sequences of length L, increases as one descends to lower tree levels ($L\to\infty$). [$N(L)$ should not be confused with the length of the measured data stream, denoted simply as N. The distinction should be clear from the context.]

If one takes into account the probability distribution of different symbol sequences, a Shannon-entropy-like quantity $H(\mathbf{p},L)$ can be defined as

$$H(\mathbf{p},L)\equiv-\sum p_{s_1s_2\ldots s_L}\ln(p_{s_1s_2\ldots s_L})\ , \tag{3}$$

where the sum is over all possible symbol strings appearing at level L. $H(\mathbf{p},L)$ increases as one descends to lower tree levels and the asymptotic rate is given by the metric entropy under partition $\mathbf{p}$,

$$h^1(\mathbf{p})\equiv\lim_{L\to\infty}\frac{H(\mathbf{p},L)}{L}\ .$$

Higher order entropies $h^q(\mathbf{p})$ can be defined as [17]

$$h^q(\mathbf{p})\equiv\lim_{L\to\infty}\frac{H^q(\mathbf{p},L)}{L}\ ,$$

where $H^q(\mathbf{p},L)$ is the order-q Renyi entropy [18] at tree level L and is given by

$$H^q(\mathbf{p},L)\equiv(1-q)^{-1}\ln\sum[p_{s_1s_2\ldots s_L}]^q\ .$$

Considered as functions of the partition $\mathbf{p}$, these entropies are maximal for generating partitions, and such maxima are defined as the ith-order entropy h^i of the dynamical system $f:M\mapsto M$,

$$h^i\equiv\sup_{\mathbf{p}}h^i(\mathbf{p})\ .$$

The h^i's are invariant under change of coordinate repre-

sentation, hence they provide an experimentally accessible way to characterize a dynamical system. The metric entropy of the dynamical system, h^1 (the so-called Kolmogorov-Sinai invariant) is of particular interest since it is related to the positive Lyapunov exponents, arguably the most important measure of chaos. It is known that the metric entropy of a differentiable dynamical system which preserves an ergodic measure with compact support can be no larger than the sum of its positive Lyapunov exponents, λ_i [19]:

$$h^1 \leq \sum_{\lambda_i > 0} \lambda_i \ .$$

Equality (the so-called Pesin identity) holds if the ergodic measure is invariant and has smooth density [20]. Thus, a positive metric entropy for a deterministic system can be taken as an indicator of the presence of chaos.

The mathematical rigor of symbolic dynamics and the experimental convenience of coarse graining would seem to guarantee this approach a much larger role than it presently has in the experimental study of dynamical systems. However, the power of the symbolic approach is overshadowed by the difficulty of constructing a generating partition for any given system [12]. This has proven to be nontrivial even in the well-studied Hénon system [5,12]. Important progress has been made for Axiom-A systems where a prescription for constructing generating partitions has been given and the Pesin identity verified [21].

It is important to realize that, in addition to the above considerations, in the presence of noise the very concept of a generating partition becomes ill-defined since increasingly long symbol sequences do not partition the state space on ever finer scales, as was pointed out in Ref. [13]. Hence, the interpretation of the symbol statistics and the associated entropy hierarchy must be modified. The symbol sequence statistics of noisy dynamical systems are still a measure of multiple-time-step correlations in a noisy signal. This is true even though they might be biased away from their noise-free values, and biased in different ways for different types of noise.

It is, of course, possible that there are optimal partitions through which such noise-induced bias can be minimized. Such optimal partitions might be associated with maximum entropy partitions (at least at low noise [13]) but a general method for finding such a partition in the noisy case is not known. This is true partly because the desired properties of such a partition have yet to be precisely articulated.

B. The error landscape

In general, the reconstruction of chaotic systems from time-series data is done by varying the model dynamics such that some error function is minimized. The error function we have chosen is the "distance" between the observed symbol tree and the tree generated by the model. The higher levels of the symbol tree can be computed if the lower levels are known. Therefore, for some fixed lowest level L, matching trees reduces to comparing only this lowest level. For example, at level $L=5$, there are $2^5=32$ possible five-symbol substrings. Hence, the tree branch space at this level is 32-dimensional and the symbol sequence statistics (to level 5) are completely characterized by the 32-dimensional tree branch vector $\mathbf{T}$ with entries $(T_0 \equiv p_{00000},\ T_1 \equiv p_{00001}, \ldots, T_{31} \equiv p_{11111})$. We define the distance between two different tree branch vectors, $\mathbf{T}$ and $\mathbf{T}'$, as the Euclidean norm of their difference: $d(\mathbf{T},\mathbf{T}') \equiv \|\mathbf{T}-\mathbf{T}'\|$.

For the purpose of introducing the error landscape we will consider the artificial case with no measurement or dynamical noise. Thus, the observations represent pure deterministic dynamics. We denote the tree branch vector measured from the data by $\mathbf{T}_0$. Choosing a set of parameters $\vec{\lambda}$ for our dynamical model $\mathbf{F}(\mathbf{y};\vec{\lambda})$, we iterate Eq. (1) and calculate $\mathbf{T}(\lambda)$, the tree branch vector for the model dynamics. The error function $\mathcal{E}(\vec{\lambda})$ is a function on parameter space defined as

$$\mathcal{E}(\vec{\lambda}) \equiv \|\mathbf{T}(\vec{\lambda})-\mathbf{T}_0\| \ .$$

A plot of $\mathcal{E}(\vec{\lambda})$ constitutes the error landscape. [In [11] this function was denoted $h(\vec{\lambda})$.] Reconstructing the dynamical system is equivalent to finding a global minimum in this landscape.

Our first example examines the error landscape associated with the quadratic map, $x_{n+1}=\lambda x_n(1-x_n)$. We chose a target value for λ, denoted by λ_0. (We chose $\lambda_0=3.75$, which is known to generate chaotic behavior.) An initial condition was chosen and iterated using the map to produce a time series. The time series was then used to produce a tree branch vector, $\mathbf{T}_0=\mathbf{T}(\lambda_0)$. This procedure was then repeated for many different values of λ and the error functions $\mathcal{E}(\lambda;\lambda_0) \equiv \|\mathbf{T}(\lambda)-\mathbf{T}(\lambda_0)\|$ were computed. The results are shown in Fig. 1.

Note the "chaotic monument valley" studded with high peaks shown in Fig. 1(a). These peaks correspond to stable high-order periodic windows. Blowing up the region around λ_0 reveals an extremely complex landscape, Fig. 1(b). Since periodic windows are dense for the quadratic map this rugged landscape is likely to be a fractal. In fact, Ershov's recent theoretical work on the parameter dependence of the invariant density for the tent map [22] suggests that the error landscape we obtain should be continuous, but not differentiable. We expect these qualitative characteristics to be generic even in higher-dimensional parameter spaces. Explicitly, we expect the error landscape associated with the symbol sequence statistics to be fractal, with peaks corresponding to high-order stable periodic motion.

Due to the complexity of the error landscape it is clear that most conventional search routines will have difficulty in finding the global minimum. In our study we employed the adaptive simulated annealing (ASA) code developed by Ingber [23]. This code has proved to be capable of finding the global minima.

Our second example examines the error landscape generated by the Hénon map [24],

$$x_{n+1}=a-x_n^2+by_n \ ,$$

$$y_{n+1}=x_n \ ,$$

on the parameter space of $\vec{\lambda} \equiv (a,b)$. The parameter

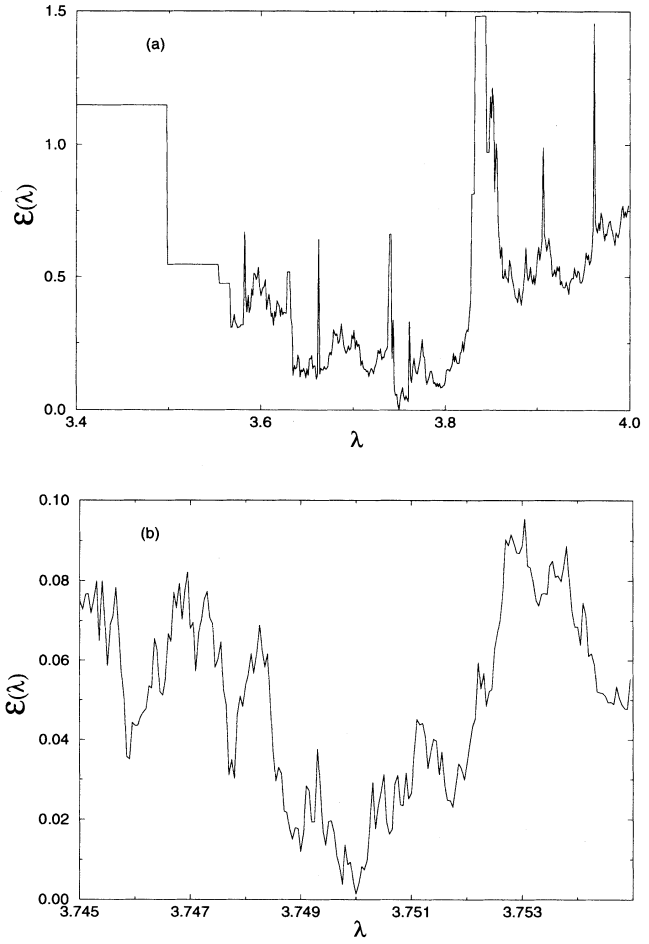


FIG. 1. The error landscape of the quadratic map ($\lambda_0=3.75$). The fifth tree level is used. (a) overview; (b) blowup view.

space portrait of the Hénon map, a so-called "isodiagram," has already been given by Gallas [25]. Basically, it consists of vast regions of low-order periodic motion or unstable dynamics ($|\mathbf{x}|\to\infty$), and a bounded "chaotic puddle." Inside the chaotic puddle there are many islands corresponding to stable periodic motion. The most remarkable ones Gallas called "shrimps." The dominant shrimps lie along a line approximated by

$$b=-0.583a+1.025\ . \tag{4}$$

In addition, the thin shrimp "legs" extended parallel to this line. Since the system can undergo a period-doubling cascade as the parameters are varied these structures are expected to repeat themselves on all scales.

The symbol sequence statistics for periodic motion (low entropy) are very different from the symbol sequence statistics for chaotic motion (high entropy). This distinction implies that, for a given dynamical system, there should be some similarity between the "isodiagram" and the error landscape, $\mathcal{E}(a,b)$. We have generated the error landscape for the Hénon map for target parameter values (a_0,b_0), see Fig. 2(a). The error landscape we obtained recovers all the features quoted by Gallas [25], and there is a global minimum at $\vec{\lambda}_0=(a_0,b_0)$. Such agreement implies the ability to examine bifurcations using such error landscapes.

For our last example we considered the error landscape generated by the Ikeda map [30],

$$\mathbf{z}_{n+1}=a-b\mathbf{z}_n\exp\left[i\kappa-\frac{i\eta}{1+|\mathbf{z}_n|^2}\right],$$

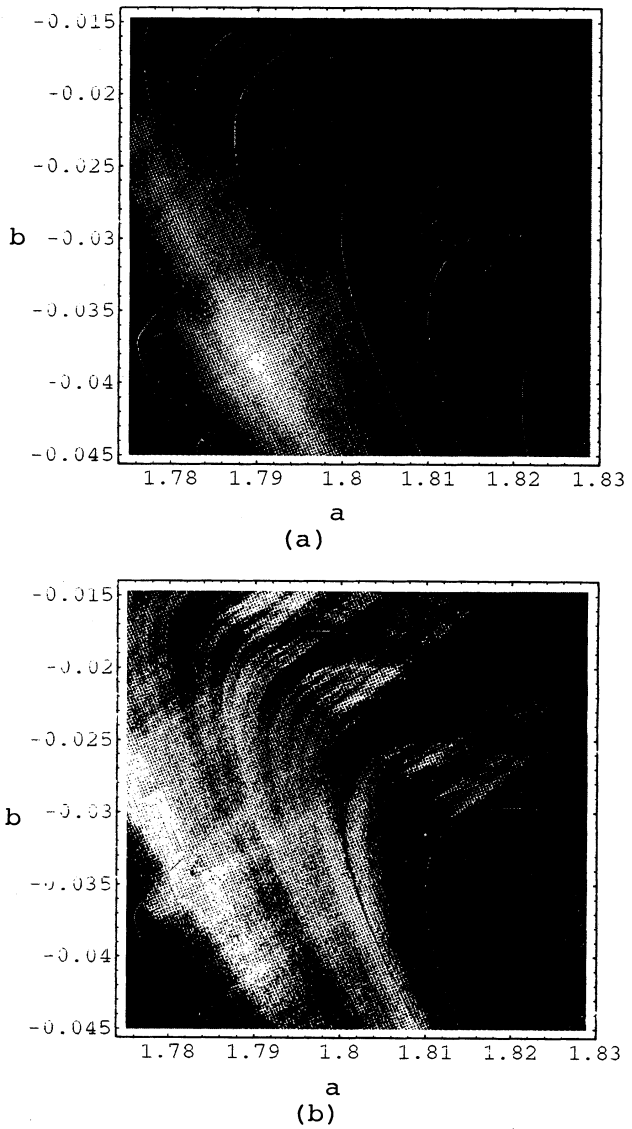


FIG. 2. The error landscape of the Hénon map ($a_0=1.79$, $b_0=-0.0385$) is shown for the parameter range $a\in(1.775,1.83)$ and $b\in(-0.045,-0.015)$. (a) Without noise, the 15th tree level is used (the minimum is at the center of the yellow area at the lower left while the maxima are red); (b) with observational noise of amplitude 0.5, the fifth tree level is used. The minima [now near the point around (1.784, −0.0365)] is in orange while red is still assigned for the maxima (shrimps).

with $z=x+iy$, as an independent test with similar results, i.e., the error landscape is fractal with peaks corresponding to high-order stable periodic windows.

III. EFFECTS OF FINITENESS OF THE DATA STREAM AND NOISE

Section II considered the artificial case of zero noise as a means of introducing the symbol sequence statistics and the error landscape. We also ignored the issue of the effects of estimating the symbol sequence statistics using only a finite amount of data. The present section forms a bridge between the ideal situation discussed in Sec. II and the applications to chaotic reconstruction in Sec. IV. Before using the symbol sequence statistics to do chaotic reconstruction, one must deal with the fact that we can only estimate the symbol sequence statistics from measured data; hence one must deal with the fluctuations in estimates due to the finiteness of the data stream, and the effects of noise. These technical issues must be dealt with before we can claim to understand how to apply our technique to experimental data. The contents of Sec. III are as follows: first we introduce some basic measures which can be used to quantify the effects we wish to study; second we discuss the effects of finite sample size; third we discuss the effects of measurement noise; fourth we discuss the effects of dynamical noise. In the final part of this section we consider the resultant impact of these effects on the error landscape.

A. Basic concepts

For the sake of clarity, we shall concentrate on an individual tree level, L. Two quantities, the Shannon entropy (more precisely, a Shannon-like entropy) and the bias function, are useful in addressing the effects of noise. The Shannon entropy for the Lth tree level is defined as

$$H_L \equiv -\frac{1}{\ln(2^L)} \sum p_{s_1 s_2 \cdots s_L} \ln(p_{s_1 s_2 \cdots s_L}) , \tag{5}$$

where the summation is over all substrings of length L. [This is proportional to $H(p,L)$ of Eq. (3) defined earlier.] H_L is a direct analog to the metric entropy since it is proportional to the metric entropy h^1 for large tree level ($\lim_{L\to\infty} H_L = h^1 \ln 2$).

As a measure of information gained or lost to the noise we define B_L the bias function of the Lth tree level at noise level σ to be

$$B_L(\sigma) \equiv \frac{H_L(\sigma)-H_L(0)}{H_L(\infty)-H_L(0)} . \tag{6}$$

$H_L(\infty)$ denotes the pure noise case. For white noise, of course, this is the maximum entropy state and $H_L(\infty)=1$. Here the term "white" signifies that the noise is delta-correlated in time, i.e., that $\langle \eta_i \eta_j \rangle = \langle \eta_i^2 \rangle \delta_{ij}$ for measurement noise, and the elements of the correlation tensor for dynamical noise satisfy $\langle u_m(t_i) u_n(t_j) \rangle = \langle u_m^2 \rangle \delta_{ij} \delta_{mn}$. In our numerical experiments we have used uniformly distributed and Gaussian distributed white noise, both of which had zero mean. In the case of Gaussian noise, the noise level σ signifies the rms of the noise ($\sigma = \langle \eta^2 \rangle^{1/2}$, etc.). For uniformly distributed noise (the type used most often in this paper), it is more convenient to set σ equal to the width. Where necessary, we will distinguish between the amplitudes of dynamical and observational noise. Unless this distinction is made explicitly, comments regarding the effects of noise should be understood to apply to both types.

The symbol sequence statistics are influenced by three factors: the finiteness of the data sample, dynamical noise, and measurement noise. To examine these effects consider a large ensemble of symbol trees. We will assume that each tree in the ensemble is constructed from a symbol sequence of length N and there are M trees in the ensemble. Furthermore, we assume that the symbol sequences used to construct the trees are polluted with noise level σ. Therefore, each tree in the ensemble is only a biased estimate of the "true" symbol tree one would obtain from an infinitely long noise-free symbol sequence.

Let $\mathbf{T}_L^{(i)}(N,\sigma) \in \mathbb{R}^{2^L}, i=1,2,\ldots,M$ denote a vector constructed from the Lth level of the ith symbol tree in the ensemble. This is a 2^L-dimensional vector whose jth component $[\mathbf{T}_L^{(i)}(N,\sigma)]_j = p_{s_1 s_2 \cdots s_L}$ is the probability of occurrence of the sequence $(s_1, s_2, \ldots, s_L)$, where j is the decimal equivalent of the binary symbol string $s_1 s_2 \ldots s_L$. (Thus, if $L=5$ then $0 \equiv 00000$, $1 \equiv 00001$, $2 \equiv 00010$, etc.)

The center of mass of the ensemble, denoted as $\langle \mathbf{T}_L(N,\sigma) \rangle$, is given by

$$\langle \mathbf{T}_L(N,\sigma) \rangle = \frac{1}{M} \sum_{i=1}^{M} \mathbf{T}_L^{(i)}(N,\sigma) .$$

This should be contrasted with the Lth level of the "true" tree

$$\mathbf{T}_L \equiv \lim_{N\to\infty} \mathbf{T}_L(N,0) .$$

(The true tree can only be estimated, of course, by computing with a very long *clean* data stream.) Bias effects will be measured by examining the noise dependence of the ensemble relative to the fixed reference point given by the true tree. Toward this end, we compute the distance of the ith sample tree from the true tree,

$$D_L^{(i)}(N,\sigma) = \| \mathbf{T}_L^{(i)}(N,\sigma) - \mathbf{T}_L \| ,$$

and the distance of the ith sample tree from the center of mass of the ensemble,

$$d_L^{(i)}(N,\sigma) = \| \mathbf{T}_L^{(i)}(N,\sigma) - \langle \mathbf{T}_L(N,\sigma) \rangle \|$$

(N.B. at zero noise, D_L and d_L are identical).

In order to characterize the distribution of $\mathbf{T}_L^{(i)}(N,\sigma), i=1,2,\ldots,M$, we compute both the means and the standard deviations of $D_L^{(i)}(N,\sigma)$ and $d_L^{(i)}(N,\sigma)$:

$$\langle D_L(N,\sigma) \rangle = \frac{1}{M} \sum_{i=1}^{M} D_L^{(i)}(N,\sigma) , \tag{7a}$$

$$\langle d_L(N,\sigma) \rangle = \frac{1}{M} \sum_{i=1}^{M} d_L^{(i)}(N,\sigma) , \tag{7b}$$

$$\sigma_L^{(D)}(N,\sigma)=\left[\frac{1}{M-1}\sum_{i=1}^{M}[D_L^{(i)}(N,\sigma)-\langle D_L(N,\sigma)\rangle]^2\right]^{1/2}, \tag{7c}$$

$$\sigma_L^{(d)}(N,\sigma)=\left[\frac{1}{M-1}\sum_{i=1}^{M}[d_L^{(i)}(N,\sigma)-\langle d_L(N,\sigma)\rangle]^2\right]^{1/2}. \tag{7d}$$

Another important measure is the distance between the center of mass and the true tree:

$$\mathcal{D}_L(N,\sigma)=\|\langle \mathbf{T}_L(N,\sigma)\rangle-\mathbf{T}_L\| . \tag{8}$$

To clarify the meaning of these quantities, consider the first tree level, $\mathbf{T}_1^{(i)}(N,\sigma)=(p_0,p_1)$, as an example. Although $\mathbf{T}_1^{(i)}(N,\sigma)$ is nominally two dimensional its entries are probabilities. Hence, $p_0+p_1=1$, and the two-dimensional vector $\mathbf{T}_1^{(i)}(N,\sigma)$ must fall on the line segment connecting (1,0) and (0,1). The first level of the true tree $\mathbf{T}_1$ lies somewhere on this line as well. We choose $\mathbf{T}_1$ as the origin and study the distribution of the $\mathbf{T}_1^{(i)}(N,\sigma)$'s along this line segment. We denote this distribution by $\mathcal{P}_{\mathbf{T}_1}(x)$ where x is the distance between $\mathbf{T}_1$ and $\mathbf{T}_1^{(i)}(N,\sigma)$ along the line segment.

In the zero noise limit, $\mathbf{T}_1^{(i)}(N,\sigma)$ is distributed symmetrically about $\mathbf{T}_1$, with $\langle \mathbf{T}_1(N,0)\rangle=\mathbf{T}_1$ in the $N\to\infty$ limit. Also, as mentioned previously, in the absence of noise the distributions of D and d are identical [$\mathcal{P}(D)=\mathcal{P}(d)$], as shown in Fig. 3. Also shown in Fig. 3 are the effects of adding a substantial amount of noise: the center of mass of the distribution of the $\mathbf{T}_1(N,\sigma)$'s becomes biased away from the true tree. However, the distribution about the biased center of mass, $\langle \mathbf{T}_1(N,\sigma)\rangle$, remains roughly the same, i.e., $\mathcal{P}(d(N,\sigma))\approx\mathcal{P}(d(N,0))$. Analogous results are shown for the fifth level of the tree in Fig. 4. This will be discussed in more detail in the following sections.

B. Effects of finite sample size on the symbol sequence statistics

As we pointed out in the introductory comments to this section, any symbol tree we compute is only an estimate of what might be called the *ideal* tree, i.e., the tree one would obtain using an infinitely long sample. In practice, the ideal tree vector is approximated by computing the tree using the entire data set:

$$\mathbf{T}_L(\sigma)\equiv\lim_{N\to\infty}\mathbf{T}_L(N,\sigma)$$
$$\approx\frac{1}{M}\sum_{i=1}^{M}\mathbf{T}_L^{(i)}(N,\sigma)=\langle \mathbf{T}_L(N,\sigma)\rangle .$$

(N.B. at zero noise the ideal tree and the true tree are identical.) In practice the true tree is unknown, and we only have experimental access to an estimate of the ideal (noisy) tree. In general the elements of the ensemble will be scattered in a cloud about the ideal vector, Fig. 4. The radius of this cloud can be defined as

$$r(N,\sigma)\equiv\left[\frac{1}{M}\sum_{i=1}^{M}\|\mathbf{T}_L^{(i)}(N,\sigma)-\langle \mathbf{T}_L(N,\sigma)\rangle\|^2\right]^{1/2} . \tag{9}$$

[i.e., r is the rms value of the set $d_L^{(i)}(N,\sigma)$.] Our numerical experiments indicate that $r(N,\sigma)$ scales as $N^{-1/2}$ and is only weakly dependent on the noise level, Fig. 5.

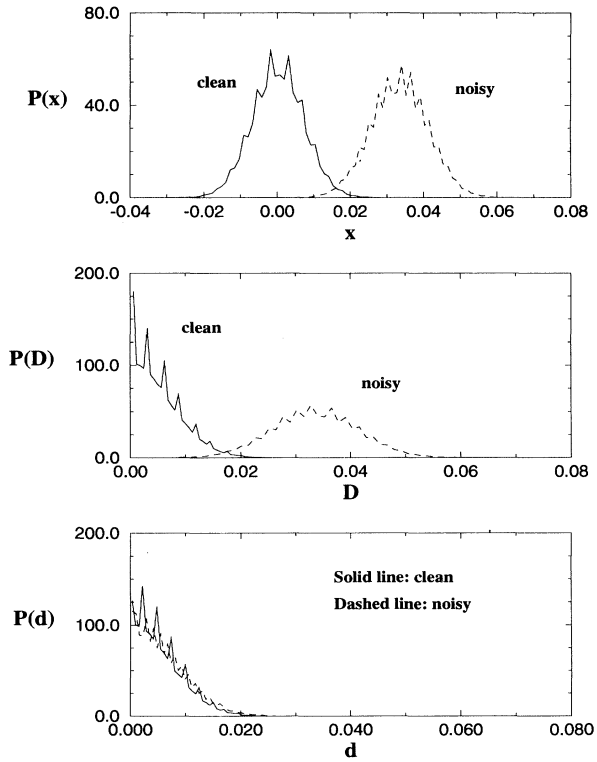


FIG. 3. The distribution of $\mathbf{T}_1(N=5000,\ \sigma=0)$ and $\mathbf{T}_1(N=5000,\ \sigma=1.0)$ in the tree branch space. There are a total of 10^5 samples generated by the Hénon map with $a=1.790$ and $b=-0.0260$. Upper: $\mathcal{P}(x)$; middle: $\mathcal{P}(D)$; lower: $\mathcal{P}(d)$.

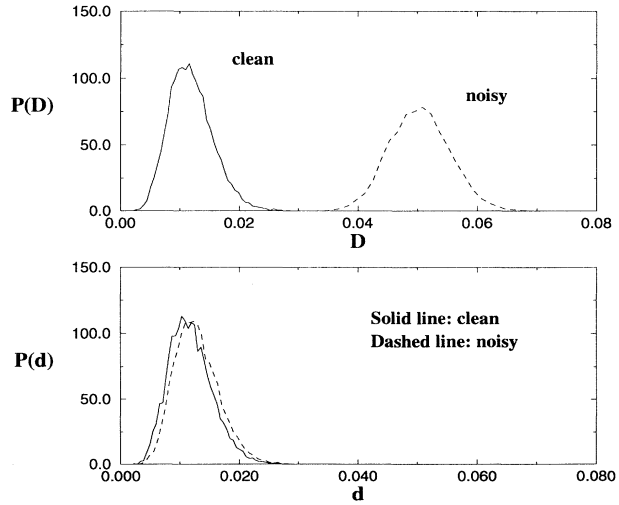


FIG. 4. The distribution of $\mathbf{T}_5(N=5000,\ \sigma=0)$ and $\mathbf{T}_5(N=5000,\ \sigma=0.7)$ in the (32-dimensional) tree branch space. There are a total of 10^4 samples generated by the Hénon map with $a=1.790$ and $b=-0.0260$.

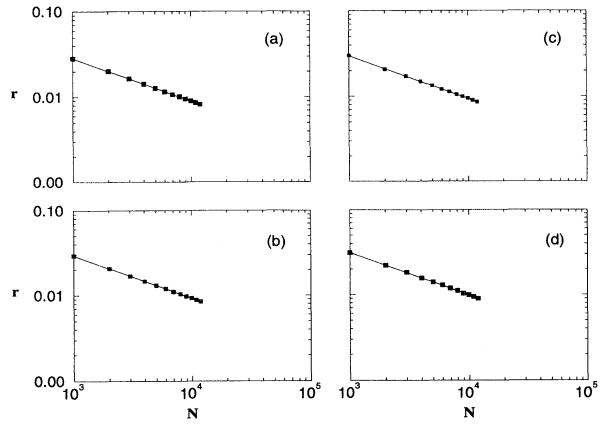


FIG. 5. The fluctuation of the symbol tree due to the finiteness of the data stream scales as $r \propto N^{-1/2}$ with r the radius of the cloud and N the length of the data stream. The solid lines are best-fit curves given by $r = r_0 N^{-1/2}$. (a) clean Hénon map, $r_0 = 0.8996$; (b) noisy Hénon map with observational noise of amplitude 0.5 and dynamical nosie of amplitude 0.02 (0.3) in $x(y)$ component, $r_0 = 0.9226$; (c) clean Ikeda map, $r_0 = 0.9434$; (d) noisy Ikeda map with observational noise of amplitude 0.1 and dynamical noise of amplitude 0.2 in both x and y components, $r_0 = 0.9882$.

C. Effect of observational noise on the symbol sequence statistics

The effect of observational noise upon the symbol tree is relatively easy to understand. We begin by observing that noise can induce a bias in the tree entries. To illustrate this consider the following example: a deterministic dynamical system with a stable periodic orbit completely contained within a single partition. Without loss of generality assume this partition is labeled by zero. A time series representing this orbit would produce a symbol string consisting of nothing but zeros. Hence, $[\mathbf{T}_L^{(i)}(N,0)]_0 = 1$ and all other components of the vector $\mathbf{T}_L^{(i)}(N,0)$ are zero for $i = 1,2,\ldots,M$.

In the presence of sufficiently large noise (observational or dynamical) the probability of a 1 in the symbol sequence is nonzero. Hence, the vector is "biased" away from its "true" value. If the noise is uncorrelated in time, and we take the $\sigma \to \infty$ limit, then all symbol sequences become equally probable. In this case all of the Lth level symbol tree vectors in the ensemble approach $\mathbf{T}_L^{(\max)} = 2^{-L}(1,1,1,\ldots,1)$, the maximum entropy vector. Thus, the ideal tree $\mathbf{T}_L(\sigma) \approx \langle \mathbf{T}_L(N,\sigma) \rangle$ follows some "bias curve" from its initial position, $(1,0,0,\ldots,0)$, to the maximal entropy position $2^{-L}(1,1,1,\ldots,1)$, as σ increases from 0 to ∞. The shape of this curve and its high-noise end point depend on the characteristics of the noise, e.g., colored noise will generate a different bias curve than white noise.

More generally, the motion of a chaotic dissipative dynamical system typically asymptotes to a strange attractor. For these types of attractors the invariant density (physical measure) has a highly nontrivial distribution consisting of many sharp peaks [26]. The peaks of such distributions are closely related to low-order unstable periodic orbits that reside in the neighborhood of the attractor [27].

The observed density is the density one obtains from a time series. It will be a convolution of the "true" density with observational noise. The resulting observed density will resemble the true invariant density except the irregularities will have been "smoothed" by the convolution with the noise distribution. Substantial observational noise is expected to broaden the peaks and lead to more uniform symbol sequence statistics. As these statistics become more uniform there will be a resulting increase in the Shannon entropy.

In Fig. 6 we plot the Shannon entropy as a function of observational noise level. The figure indicates that as the noise level increases the Shannon entropy, Eq. (5), increases. Similarly, the figure indicates that the bias function, Eq. (6) grows as the noise level increases. The effect of sample size on symbol sequence statistics is shown in Fig. 7.

D. Effects of dynamical noise on the symbol sequence statistics

Here we wish to examine the effect of nontrivial levels of dynamical noise. By nontrivial we mean many orders of magnitude larger than numerical roundoff. Unlike observational noise, dynamical noise affects the physical trajectory of the system. As the amplitude of dynamical noise increases the invariant distribution on the attractor changes. The most notable change in the topology of the attractor happens at the "tips" or "sharp folds" of the attractor where the stable and unstable directions are degenerate. In Fig. 8 we show clean and noisy attractors associated with the Hénon and Ikeda maps. One can see that dynamical noise causes the tips of the attractor to stretch further before folding back on themselves. This is especially true as higher levels of dynamical noise are introduced.

It is important to realize that sufficiently large dynamical noise can push the trajectory out of the basin of attraction. Therefore, in general there is an upper bound (sometimes quite low) on the dynamical noise amplitude that can be tolerated before stable dynamics is lost. This is particularly true when the system is near a "crisis," i.e., when parts of the attractor are close to the basin boundary. This upper limit is independent of the reconstruction method being used.

These subtleties aside, we observe through numerical experiments that the effect of dynamical noise upon symbol sequence statistics, and the resulting symbol tree, is similar to that of observational noise. While the symbol sequence statistics are quite robust in the presence of moderate amounts of dynamical noise, substantial dynamical noise produces a bias in the symbol sequence statistics. The Shannon entropy, Eq. (5), was found to generally increase as we increase the magnitude of the dynamical noise that was coupled to the deterministic dynamics (see Fig. 9). The same increase was observed in the bias function, Eq. (6). The slight decrease in the Shannon entropy at small dynamical noise amplitude

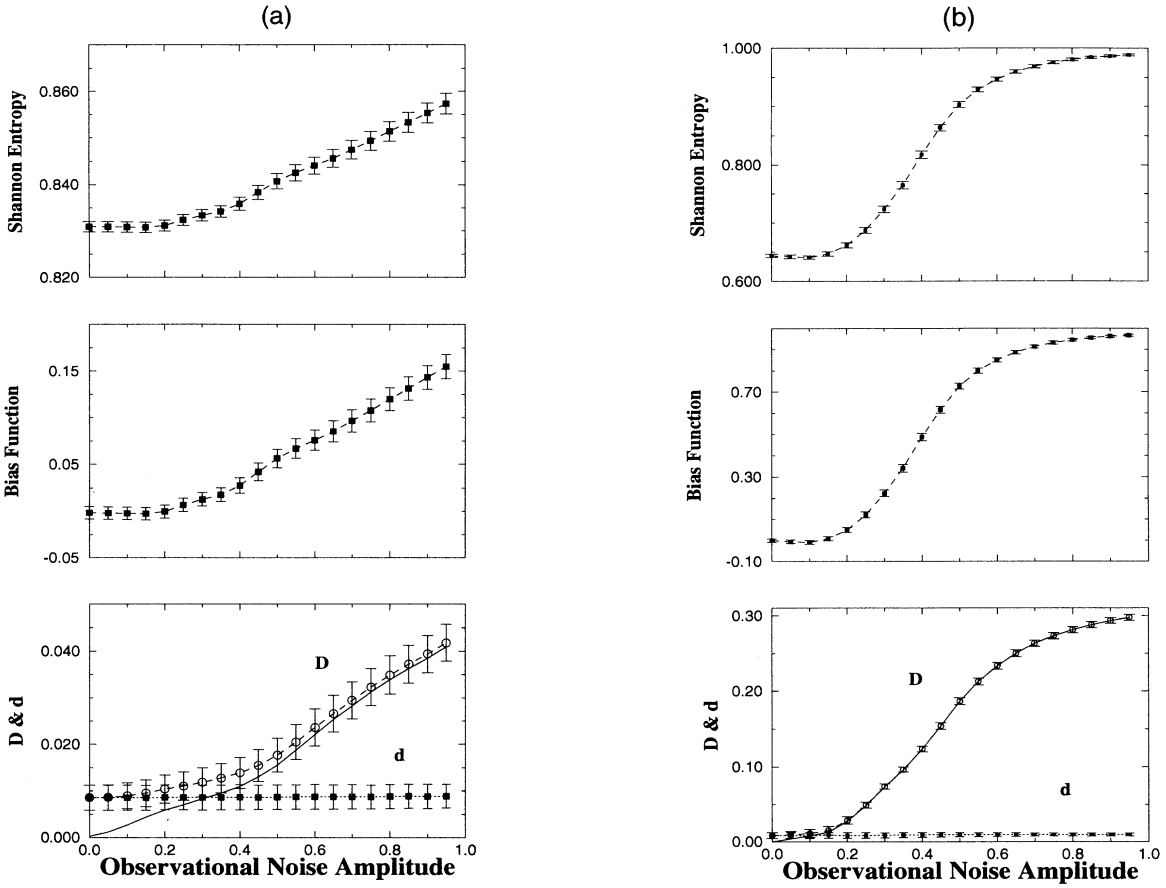


FIG. 6. The Shannon entropy and bias function in general increase upon the introduction of uncorrelated observational noise, but they drop slightly at low-noise amplitude. The effect of the observational noise on the symbol sequence statistics is limited to a bias in the tree branch space. (The means and standard deviations of the quantities D and d are defined in Eqs. (7), while the solid line in the third plot is $\mathcal{D}_L$ [see Eq. (8)].) (a) The Hénon map with $a=1.790$ and $b=-0.0260$; (b) the Ikeda map with $a=1.0$, $b=-0.7$, $\kappa=0.4$, and $\eta=6.0$.

might be associated with the so-called noise-induced order [28]. The dips in the Shannon entropy and bias function are insignificant compared with the global trend of increase. The effect of sample size on symbol sequence statistics is shown in Fig. 10.

E. Effects on the error landscape

Although the effect of noise (both observational and dynamical) on the symbol statistics is, crudely speaking, limited to a simple shift of the center of mass of the distribution, the ensuing effect on the error landscape can be more dramatic. Here, one must distinguish between *local* and *global* effects. Consider the following thought experiment: for a given model, with parameter values $\vec{\lambda}_0$, measure an infinite amount of noise-free data and compute the "true" symbol sequence statistics. Use this data to generate the true Lth-level tree, $\mathbf{T}_L(\vec{\lambda}_0)$ (where the parameter dependence is now explicitly displayed). Construct the "true" error landscape by using models from the same class to generate infinite amounts of noise-free data. The true error landscape is a graph of the following function:

$$\mathcal{E}^{(\text{true})}(\vec{\lambda};\vec{\lambda}_0)=\|(\mathbf{T}_L(\vec{\lambda})-\mathbf{T}_L(\vec{\lambda}_0)\| \ . \tag{10}$$

This landscape will possess a global minimum at $\vec{\lambda}=\vec{\lambda}_0$.

We now consider the effect of *small* amounts of noise. As discussed in the previous sections, the noisy symbol statistics will exhibit only a small deviation from the "true" symbol statistics. Hence $\mathbf{T}_L(\vec{\lambda}_0;\sigma\ll 1)\approx\mathbf{T}_L(\vec{\lambda}_0)$ and the error landscape becomes

$$\mathcal{E}(\vec{\lambda};\vec{\lambda}_0)\equiv\|\mathbf{T}_L(\vec{\lambda})-\mathbf{T}_L(\vec{\lambda}_0;\sigma)\| \ . \tag{11}$$

It is important to recognize that, although the noisy vector lies close to the true vector, in general no choice of parameter values $\vec{\lambda}$ will lead to an exact match of symbol statistics. Thus, $\mathcal{E}(\vec{\lambda})>0$.

Continuity of the invariant distribution implies that for small noise there will still be a global minimum, denoted by $\hat{\vec{\lambda}}(\sigma)$, in the neighborhood of $\vec{\lambda}_0$. Hence, the bias in the parameter estimate, $\delta\vec{\lambda}(\sigma)=\hat{\vec{\lambda}}(\sigma)-\vec{\lambda}_0$, satisfies $\delta\vec{\lambda}(\sigma)\to\vec{0}$ as $\sigma\to 0$. Because of this limiting behavior we call this a "local" effect.

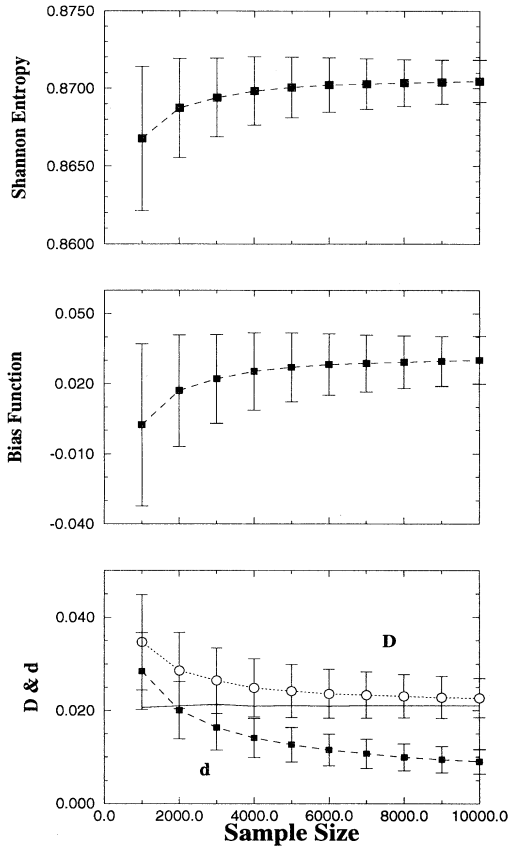


FIG. 7. The behavior of the symbol sequence statistics of the Hénon map ($a=1.790$ and $b=-0.0260$) with observational noise are plotted against the sample size in the tree branch space (fifth level). Each point is computed by using 5000 samples.

More damaging effects can occur at high-noise levels. As indicated above, high-noise levels produce substantial bias in the symbol statistics. Therefore, for high-noise levels it is possible for a new global minimum, far from the target parameter values, to appear. This "global" effect is a bifurcation in the sense that the parameter estimate has a qualitatively new solution. Clearly if such a bifurcation is present it will lead to a substantial error in the estimated value of $\vec{\lambda}_0$. In the next section we will show how to remove these effects by including noise as part of the model.

IV. RECONSTRUCTING THE DYNAMICS

In this section we finally address the issue of reconstructing a chaotic dynamical system using the symbol sequence statistics. This requires finding the global minimum of the error function. Because the error landscape is fractal, this minimum cannot be found using standard search routines (e.g., gradient descent). We used the simulated annealing code of Ingber [23] which performed well. Each of the topics of the preceding section (finite-N fluctuations, observational and dynamical

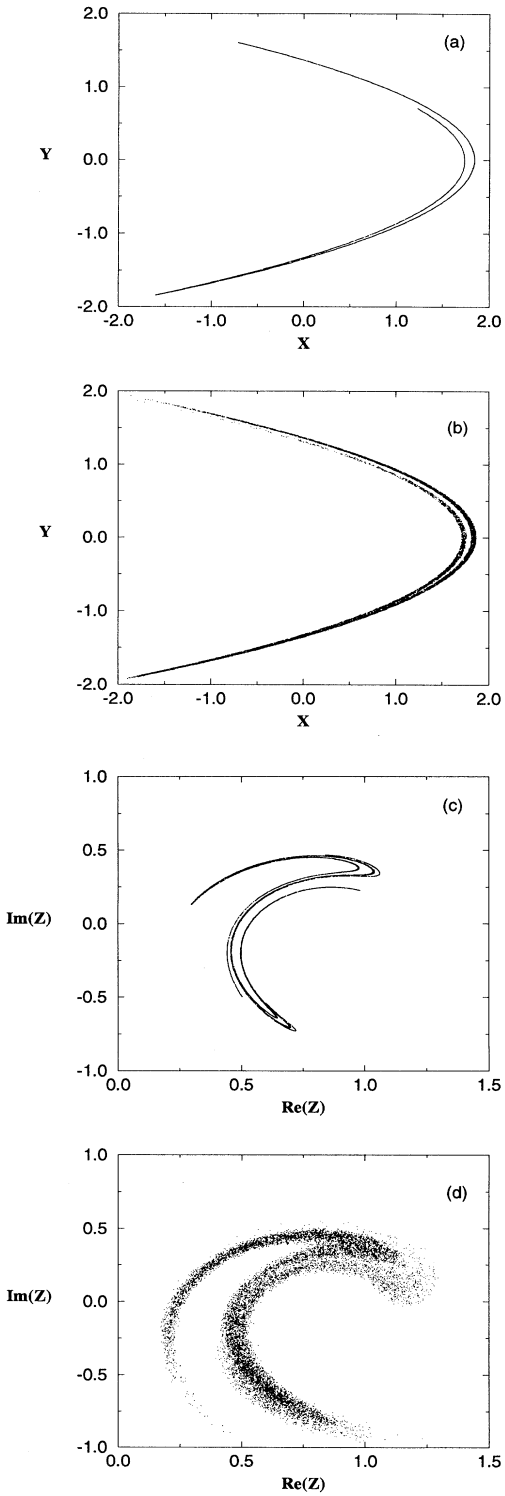


FIG. 8. The effect of dynamical noise on strange attractors (Hénon: $a=1.790$ and $b=-0.0385$; Ikeda: $a=1.0$, $b=-0.7$, $\kappa=0.4$, and $\eta=6.0$). (a) Clean Hénon attractor; (b) noisy Hénon attractor with dynamical noises of amplitude 0.06 (coupled to the x component) and 0.7 (coupled to the y component); (c) clean Ikeda attractor; (d) noisy Ikeda attractor with dynamical noises of amplitude 0.3 (coupled to the x component) and 0.5 (coupled to the y component).

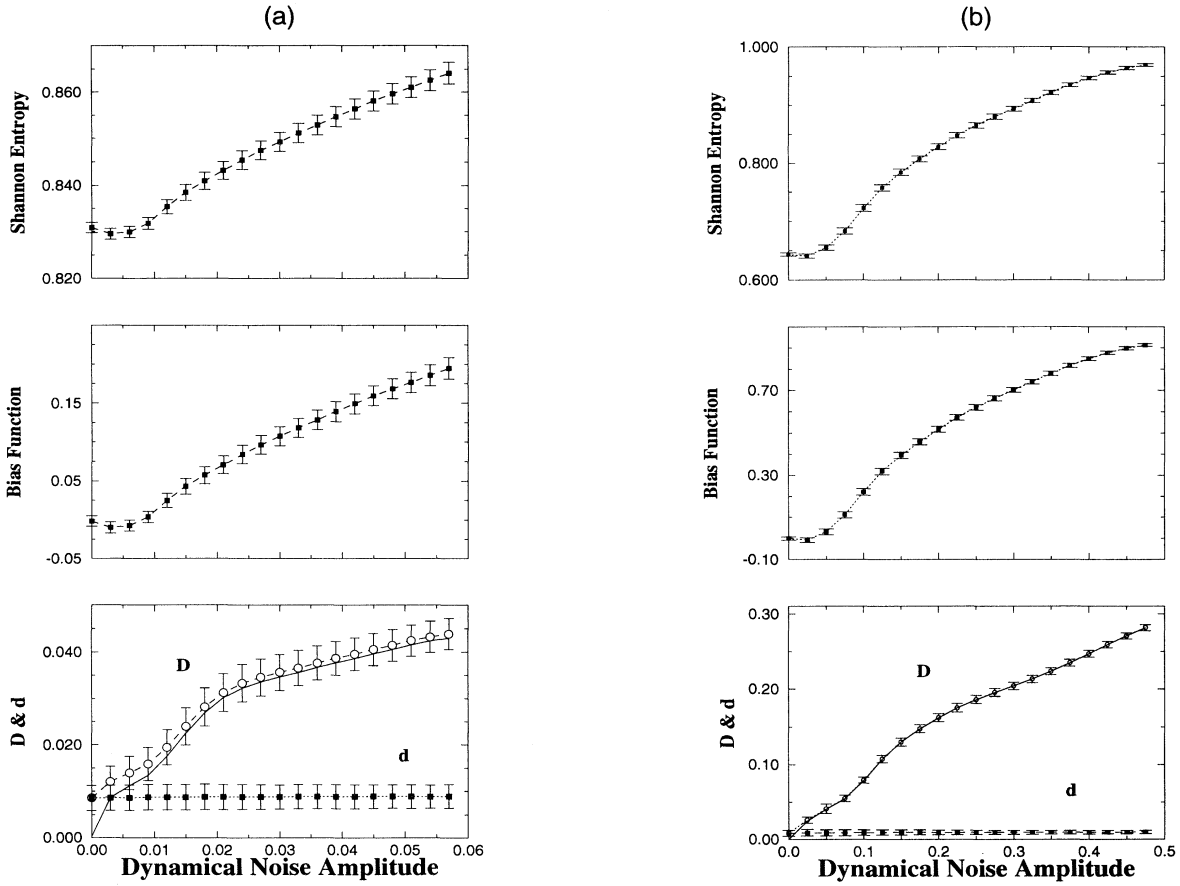


FIG. 9. The Shannon entropy and, hence, the bias function generally increase upon the introduction of uncorrelated dynamical noise, but they drop slightly at low-noise amplitude. The effect of the dynamical noise on the symbol sequence statistics is also limited to a bias in the tree branch space. (The means and standard deviations of the quantities D and d are defined in Eqs. (7), while the solid line in the third plot is $\mathcal{D}_L$ [see Eq. (8)].) (a) The Hénon map with $a=1.790$ and $b=-0.0385$; (b) The Ikeda map with $a=1.0$, $b=-0.7$, $\kappa=0.4$, and $\eta=6.0$.

noise, the error landscape) are revisited, now in the context of parameter fitting.

In any experimental situation, one is restricted to using noisy data to estimate the target tree. In the attempt to reconstruct the underlying dynamics, one can pursue two different strategies. At low-noise levels, the bias in the estimates of the symbol statistics should be small, hence the model dynamics used to fit the data should be strictly deterministic. At high-noise levels, the bias in the symbol statistics will be substantial, possibly resulting in a global change in the error landscape and subsequent degradation in the parameter estimates. This can be countered by computing the model tree using a combined deterministic/stochastic model. The noise characteristics then become a new set of parameters to be estimated during the fitting process.

A. Parameter fitting in a fractal error landscape

In this paper we are attempting to solve an inverse problem. Since inverse problems can be notoriously unstable, the following two questions must be addressed: (1) How robust is our ability to estimate parameters? (2) Do small changes in the target tree entries lead to small changes in the parameter estimates?

To answer these questions we used the Hénon map, with $a_0=1.79$ and $b_0=-0.0385$, to generate an ensemble of sample time series. The ensemble contained 3600 samples each of length $N=5000$. Next we used these samples, and a fixed partition, to form an ensemble of symbol trees. For each tree we selected an initial guess for the parameters and used the annealing code [23] to minimize Eq. (11). This procedure results in an ensemble of estimates for the parameters a and b. Initial guesses for the annealing code [23] were randomly chosen from the domain defined by $a\in(1.0,1.8)$ and $b\in(-0.2,-0.1)$, while the search range was specified to be $a\in(-2,2)$ and $b\in(-2,2)$ [29]. The annealing code was run for a fixed number of iterations (in this case 4×10^3), and those samples that passed the error condition $\mathcal{E}(a,b)\leq\delta\mathcal{E}=10^{-2}$ were retained.

The results of our noise-free study are shown in Fig. 11. We found that approximately 600 of the 3600 samples passed the error condition. The samples that passed the error condition produced parameter estimates that form a cloud which straddles the true target parameters.

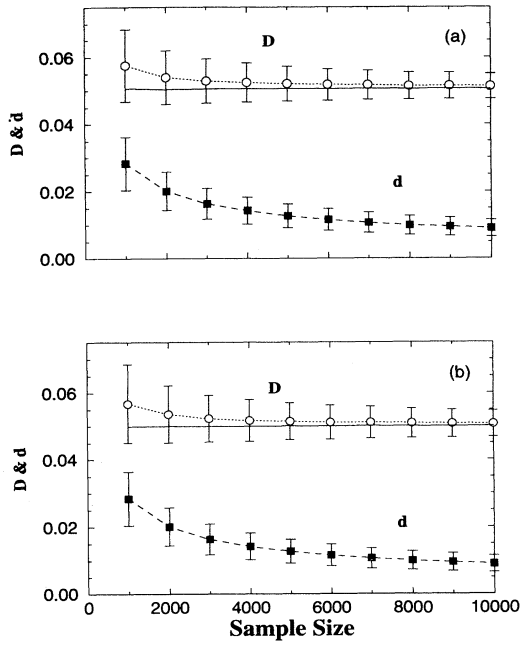


FIG. 10. The effect of the sample size on the symbol sequence statistics generated by the Hénon map ($a=1.790$ and $b=-0.0385$) with (a) dynamical noise of amplitude 0.04 in the x component; (b) dynamical noise (amplitude 0.03 in x and 0.5 in y) and observational noise (amplitude 0.5). (The means and standard deviations of the quantities D and d are defined in Eqs. (7), while the solid line in the third plot is $\mathcal{D}_L$ [see Eq. (8)].)

Our numerical experiments indicated that there are two ways to reduce the amount of scatter in this experiment. The first method involves increasing the number of annealing iterations. When this was done the rms scatter of the parameter estimates, Eq. (9), reduced and even more samples satisfied the error condition. The second method to reduce scatter involved using longer symbol sequences to construct the target trees. As stated above, finite sample lengths result in fluctuations in the tree entries. By increasing the sample lengths these fluctuations decrease,

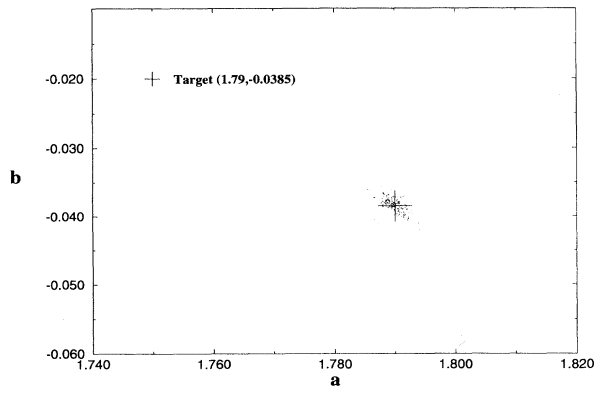


FIG. 11. Reconstruction results (parameter estimation) from clean time series produced by the Hénon map. Only those results with error function $\mathcal{E}(\vec{\lambda})<10^{-2}$ are retained.

which results in reduced scatter in the parameter estimates. This behavior suggests that there is no *systematic* bias in our fitting procedure. We conclude from this study that, in the absence of noise, the inverse problem is stable and convergent.

Although neighboring target trees produce neighboring parameter estimates, the distribution of parameter estimates in parameter space can be nonuniformly skewed along striations in the error landscape. This effect, for the Hénon map, can be seen by comparing the distribution of the parameter estimates (Fig. 11) with the error landscape (Fig. 2). The long axis of the distribution in Fig. 11 lies along the line given in Eq. (4) and corresponds to a direction in which the error landscape in Fig. 2 has a slow variation. The short axis of the distribution corresponds to a direction in which the error landscape has a sharp variation.

In an effort to quantify this behavior we note that, in the absence of noise, there are two major factors affecting the distribution of the parameter estimates. The first factor is the finite-N fluctuations previously discussed. The second factor is associated with different optimization trials. We found, not surprisingly, that using different initial guesses for the annealing code and/or changing the bounds of the search caused the annealing routine to produce different estimates for the parameter values (for a fixed number of iterations and the same target tree). We denote the standard deviation of the spread in the distribution of the parameter estimates due to finite-N effects by σ_1 and the spread due to different optimization trials by σ_2.

In order to examine σ_2 we constructed a *single* target tree and performed many optimization trials, each trial using different initial guesses or search bounds. We retained the parameter estimates only if the trial satisfied the error condition, $\mathcal{E}\leq\delta\mathcal{E}$. Since all of the trials used the same target tree the scatter in the parameter estimates is a measure of σ_2. The results are shown in Fig. 12(a) where we have plotted the mean values and the standard deviations of the parameter estimates vs the error bounds. We note that σ_2 decreases monotonically as the error constraint is made more stringent.

The asymptotic limits of a and b in Fig. 12(a) are $a=1.792$ and $b=-0.0255$ which differ from the target values of $a=1.79$ and $b=-0.026$. This difference is due to a finite-N fluctuation in the single target tree used. This finite-N effect can be eliminated by choosing a different tree from the ensemble for each new annealing run. When this is done, the standard deviations of the parameter estimates steadily shrink as one imposes stricter error constraints, as can be seen in Fig. 12(b), *and* the parameter estimates converge to the target values $a=1.79$ and $b=-0.0385$.

B. Parameter fitting in the presence of noise

Substantial amounts of noise (observational and/or dynamical) produce a notable bias in the symbol sequence statistics which induces a degradation in parameter estimations. Our numerical investigations suggest that the bias in parameter estimations saturates rather quickly as

the noise amplitude is increased (Fig. 13). We emphasize that these comments hold for the case where noise is not included in the modeling, i.e., the error function is given by Eq. (11). The bias can be removed by including noise in the modeling, as will be discussed in the next section.

To study the effect of noise on reconstruction, we once again generated a large number of parameter estimates and recorded those which satisfied specified error bounds. (The error bound was set higher if the target was polluted by higher noise.) As before, the finite-N fluctuations are dealt with by using different target trees for each optimization trial. Finally, the mean and the standard deviation of the parameter estimates were computed and plotted against the noise amplitude.

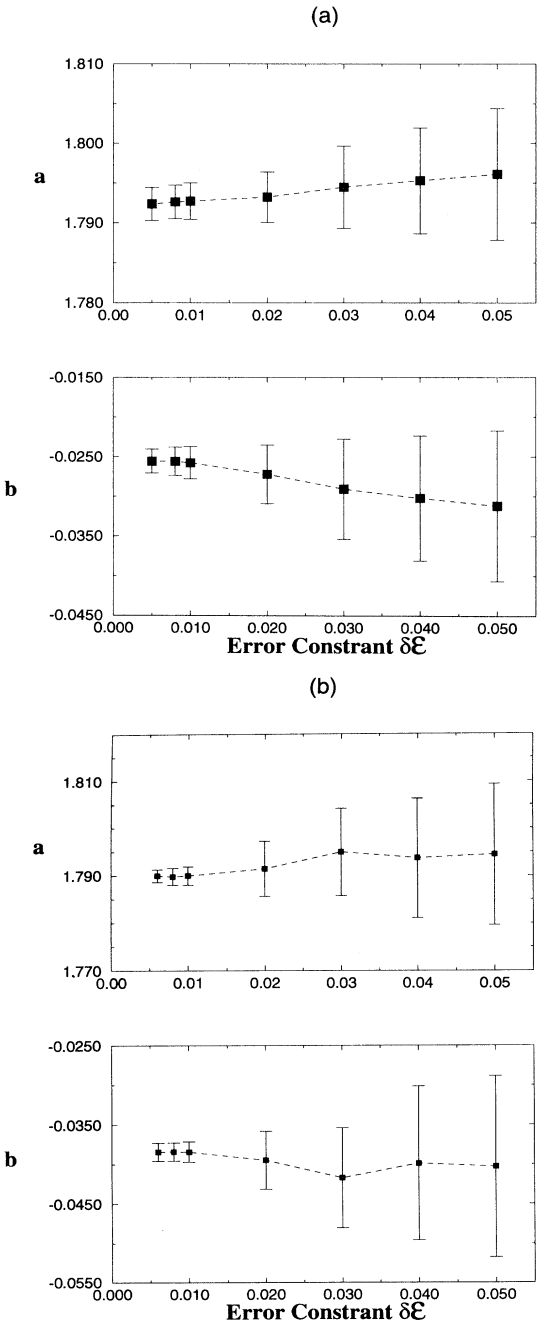


FIG. 12. The standard deviation of parameter estimates of the Hénon system shrinks as one imposes a stricter error function constraint. (a) Many reconstruction (annealing) attempts for one target ($a=1.79$ and $b=-0.026$) realization; (b) many reconstruction attempts for many target ($a=1.79$ and $b=-0.0385$) realizations.

In Fig. 13, we plot the results for the parameter estimates of a and b of the Hénon system as a function of observational noise amplitude. In Fig. 13(a) the target was set at ($a=1.790$, $b=-0.026$) and no dynamical noise is present (aside from numerical roundoff). In Fig. 13(b), the target values were moved to ($a=1.790$, $b=-0.0385$) and white dynamical nose of amplitude 0.01 was coupled to the x component, in addition to the observational noise. In Fig. 13(c) the target parameter values remains the same as in Fig. 13(b), but dynamical noise of amplitude 0.02 for the x component and 0.3 for the y component was present in addition to the observational noise.

The reconstruction results as a function of dynamical noise amplitude are presented in Fig. 14. In all these studies the target parameter values are fixed at ($a=1.790$, $b=-0.0385$). In Fig. 14(a) dynamical noise is coupled to the x component and no observational noise is present. In Fig. 14(b) dynamical noise is coupled to the y component and observational noise of amplitude 0.1 is present. In Fig. 14(c) dynamical noise is, once again, coupled to the x component and observational noise of amplitude 0.2 is present.

From the results shown in Figs. 13 and 14 one can see that symbol sequence statistics are reliable reconstruction constraints in the presence of moderate amounts of observational and/or dynamical noise. In particular, two- to three-digit accuracy was achieved for the Hénon system with observational noise of amplitudes up to 0.1 (about 10% of the signal amplitude) and/or dynamical noise of amplitude 0.01 (about 1% of the signal amplitude). For higher-noise levels the bias in parameter estimates saturate (i.e., cease to grow with increasing noise amplitude) even as the bias in the symbol sequence statistics continues to grow (see Figs. 6 and 9). As can be seen in Fig. 14, even when the noise was substantial (i.e., the observational noise becomes comparable to the signal, or the dynamical noise reaches its limit before inducing divergent motion), the parameter estimates are not too far off target. This is particularly true for the parameter a of the Hénon system.

The apparent success of this approach in recovering the true dynamics in an extremely noisy environment can be attributed to our choice of error function. The constraint of minimizing the difference in symbol sequence statistics is much stricter than many other constraints used in the past by other researchers (e.g., time evolution entropy, Lyapunov exponents, etc.). The emphasis in this earlier work was to use metric invariants as the targets for reconstruction, the argument being that this gave a target which was independent of the coordinates (or data variables) used. The symbol sequence statistics are *not* invariant, hence we must assume that the modeling is done

in a manner consistent with the observations (as is typically the case in physics applications.) Using the full symbol statistics (as opposed to the metric entropy, for example) leads to a sharper error landscape.

Take the first tree level as an example. $\mathbf{T}_1^{(i)}(N,\sigma)$ has to fall on the line segment connecting (1,0) to (0,1). There are two points, say, (p_0,p_1) and (p_1,p_0), having exactly the same entropy. In the case of deeper tree levels, say the Lth, $\mathbf{T}_L^{(i)}(N,\sigma)$ lies on a hyperplane defined by $\sum_{j=1}^{2^n-1}[\mathbf{T}_L^{(i)}(N,\sigma)]_j=1$. There are 2^L permutations of the entries of $\mathbf{T}_L^{(i)}(N,\sigma)$ which give rise to the same entropy. Hence, if the entropy were used as the target then each of these would also give a global minimum. This is particularly problematic in the high-noise situation when all the permuted vectors are converging in the neighborhood of the maximum entropy vector. Obviously, minimizing the difference in symbol sequence statistics is a much more discriminating measure.

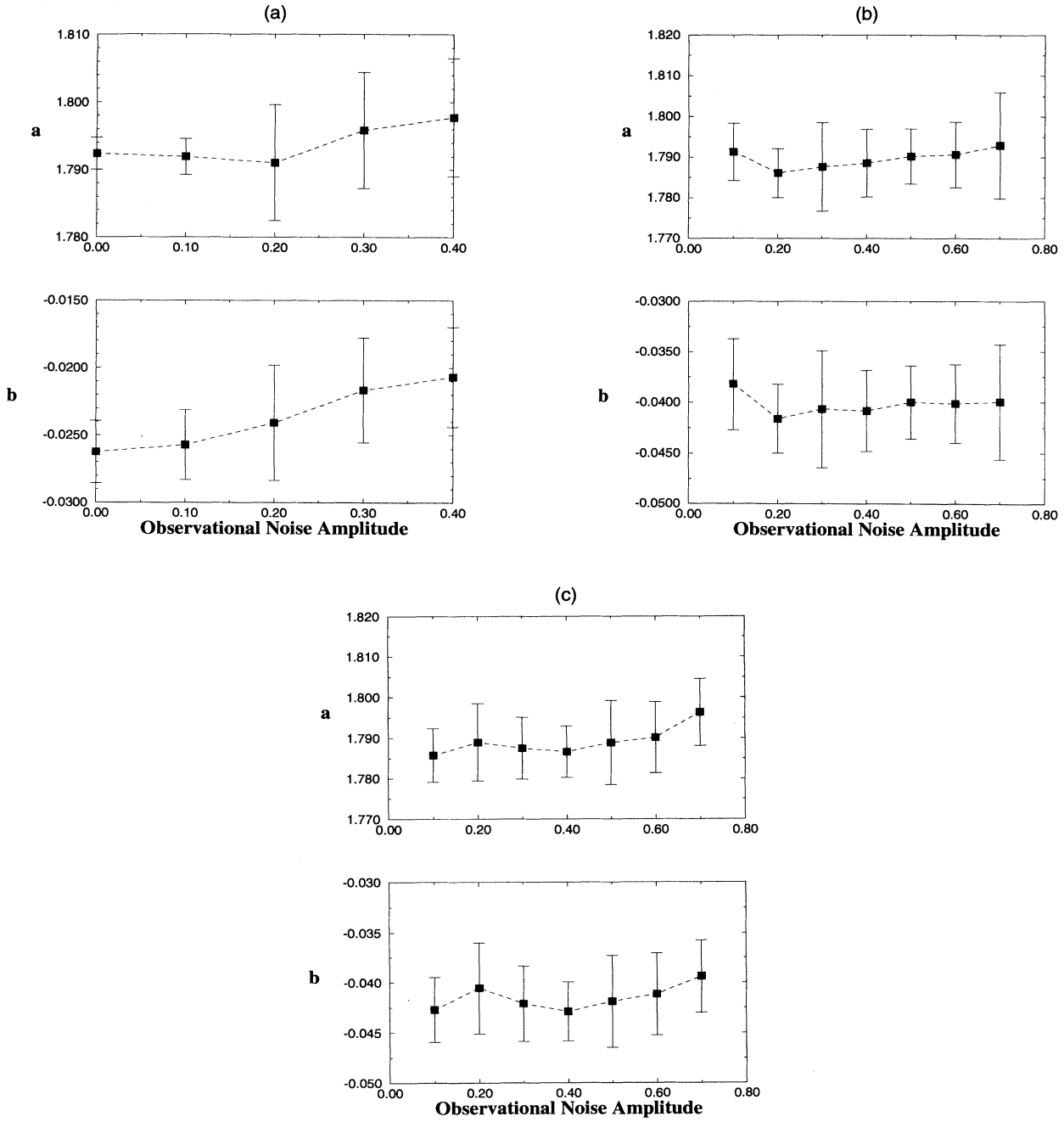


FIG. 13. Reconstruction results of the Hénon system under different observational noise amplitudes. The target parameters are $a_0=1.790$ and $b_0=-0.0385$. (a) No dynamical noise; (b) dynamical noise of amplitude 0.01 in the x component; (c) dynamical noise of amplitude 0.02 in the x component but amplitude 0.3 in the y component.

In the next section we will show that by including the noise characteristics into the model the bias in parameter estimates can be removed even when the noise dominates the signal.

C. Modeling the noise

Since the tree branch vectors tend to move as a group along a unique bias curve, the effects of noise on the tree can be taken into account by including a model of the noise in the computation of the model tree. To be precise, consider the following example: suppose our target tree was generated by the Hénon map with parameter values (a_0,b_0) and an observational noise level of σ_0. The target vector is $\mathbf{T}_L(\vec{\lambda}_0,N,\sigma_0)$. We will model the observed time series as a dynamical signal plus observational noise. The dynamical signal is $\mathbf{y}(n+1)=\mathbf{F}[\mathbf{y}(n);\vec{\lambda}]$

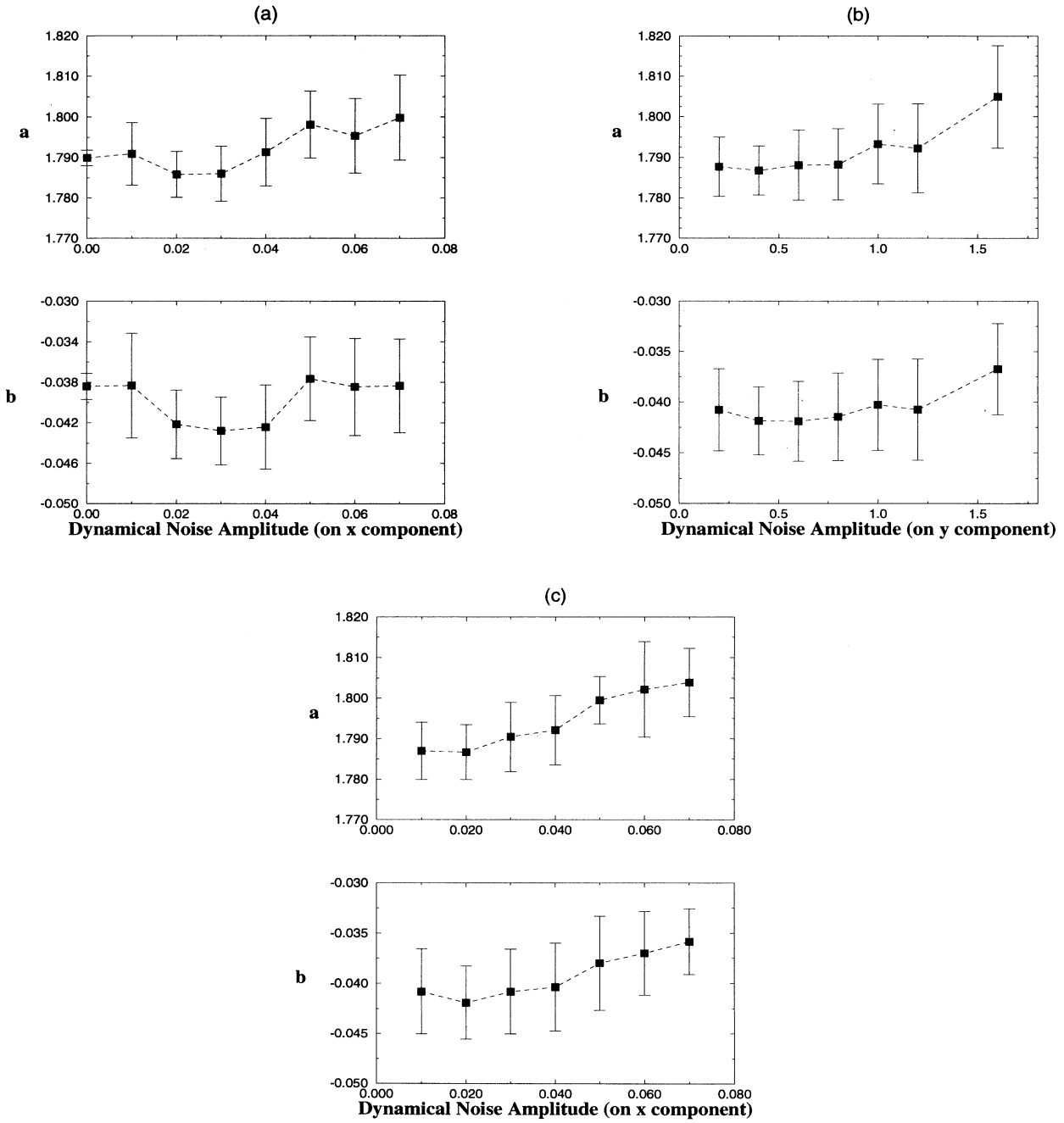


FIG. 14. Reconstruction results of the Hénon system under different dynamical noise amplitudes. The target is set to be $a_0=1.790$ and $b_0=-0.0385$. (a) Dynamical noise is in the x component and there is no observational noise; (b) dynamical noise is in the y component and there is an observational noise of amplitude 0.1; (c) dynamical noise is in the x component and there is an observational noise of amplitude 0.2.

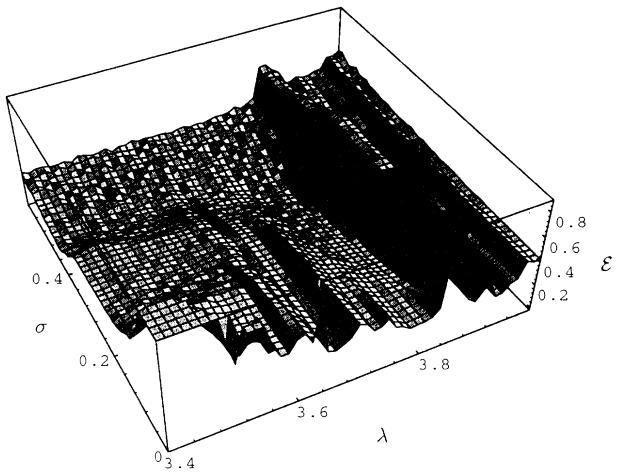


FIG. 15. The extended error landscape of the quadratic map is plotted in the range $\lambda \in (3.4, 4.0)$ and observational noise amplitude $\in (0, 0.6)$. The observed data was generated by the quadratic map with $\lambda_0 = 3.75$ and polluted by observational noise of amplitude 0.4.

where **F** is the Hénon map. The tree vector that results from evolving the model for the observed time series, $\mathbf{T}_L(\vec{\lambda}, N, \sigma)$, is computed using one component of the time series polluted by measurement noise, $x_n \equiv y_1(n) + \eta(n)$. The noise characteristics are another set of parameters to be varied. In this extended parameter space the error landscape is given by

$$\mathcal{E}^{(\text{noise})}(\vec{\lambda}, \sigma; \vec{\lambda}_0, \sigma_0) \equiv \| \mathbf{T}_L(\vec{\lambda}, N, \sigma) - \mathbf{T}_L(\vec{\lambda}_0, N, \sigma_0) \| \ . \quad (12)$$

This new error landscape will now have a global minimum at $(\vec{\lambda}, \sigma) = (\vec{\lambda}_0, \sigma_0)$ (i.e., we optimize the model over $\vec{\lambda}$ *and* σ). As mentioned in the preceding section, a new effect now comes into play: as the observational noise level increases *all* of the vectors are biased towards the maximum entropy vector $\mathbf{T}_L^{(\max)}$. This causes the entire error landscape to flatten and the global minimum to broaden (Fig. 15). With such a flat landscape if the sample length N is too short then finite-N fluctuations could lead to false global minima and to poor resolution in the parameter estimates. In principle this effect can be counteracted by using longer sample strings. Therefore, given computational resources, available data, and precision requirements there exists an upper bound on the noise level which can be tolerated.

Our numerical strategy was basically the same as before: a model of the observed time series was used to generate 640 sample strings of length $N = 2 \times 10^4$ (compared with $N = 5000$ used for the zero noise computation). The annealing code was run for a fixed number of iterations. Due to the larger parameter space, the annealing code was run longer (typically 15 000 iterations) in order to get a substantial number of samples to attain the error bound of $\mathcal{E}^{(\text{noise})}(a, b, \sigma; a_0, b_0, \sigma_0) \leq \delta\mathcal{E} = 10^{-2}$. Approximately 250 samples passed this error condition. As before, if the annealing code was run longer, more samples joined the convergent group and the rms scatter was reduced. Results are presented in Fig. 16 for a measurement noise level of $\sigma(\text{noise})/\sigma(\text{signal}) > 1$. Most importantly, notice that the cloud of parameter estimates straddles the neighborhood of the target parameters. Thus, even though noise produces a bias in the tree, by modeling the noise itself it is possible to remove this systematic error from the parameter estimates.

To test further the robustness, we added modest levels of dynamical noise in the generation of the target vector, in addition to the substantial level of observational noise. In this test only the observational noise was assumed in our model of the observed time series. A summary of the results is given in Table I. In Fig. 17, we show the phase space portraits of the noisy Hénon map (target) and the reconstructed Hénon map.

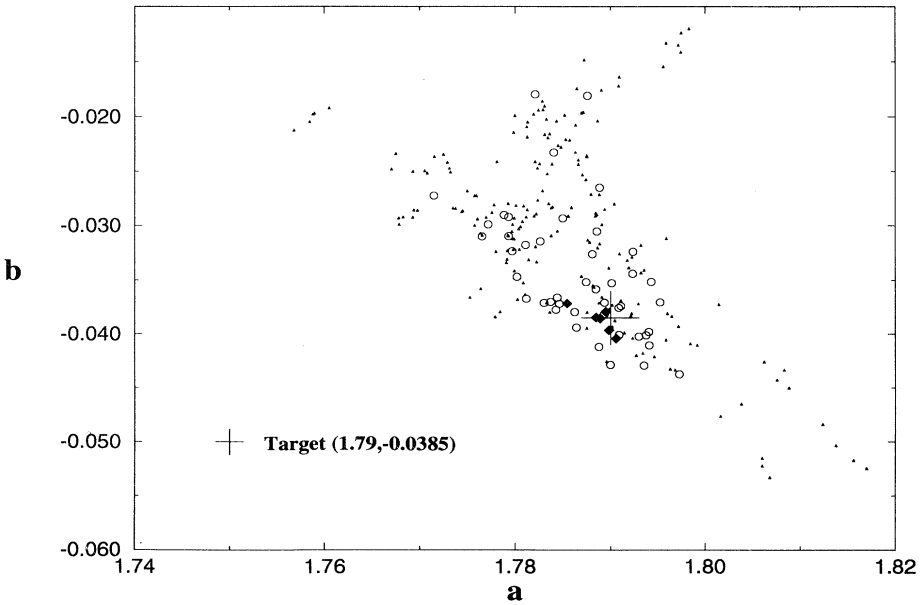
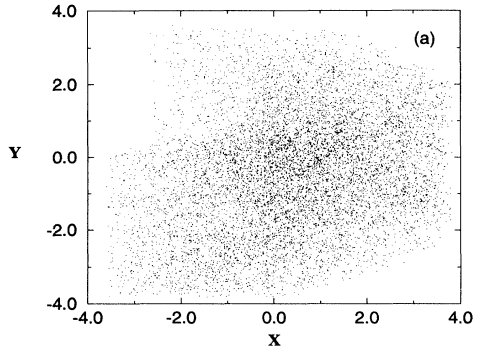


FIG. 16. Reconstruction results from noisy (observational noise only) Hénon time series $[\sigma(\text{noise})/\sigma(\text{signal}) > 1]$. Filled triangle, circle, and filled diamond correspond to those realizations satisfying error conditions: $6 \times 10^{-3} < h < 10^{-2}$, $4.5 \times 10^{-3} < h < 6 \times 10^{-3}$, and $h < 4.5 < 10^{-3}$, respectively.



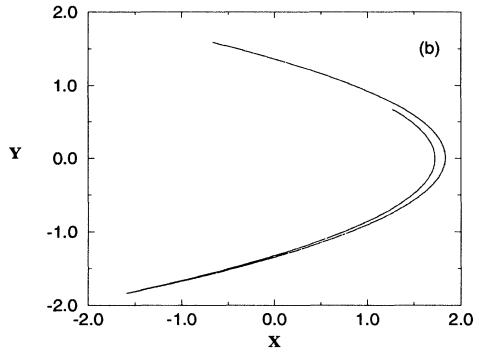
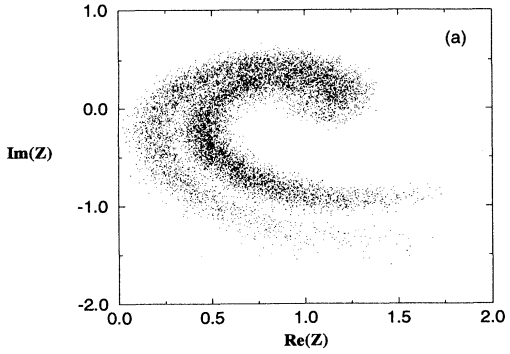


FIG. 17. (a) Noisy Hénon attractor with $a_0=1.790$ and $b_0=-0.0385$. The observational noise is of amplitude 2.0, while the dynamical noise is coupled to both x (0.01) and y (0.1) components of the map; (b) the reconstructed Hénon attractor. The parameters are found to be $a=1.78$ and $b=-0.041$, see Table I. The clean Hénon attractor of $a=1.790$ and $b=-0.0385$ is shown in Fig. 8(a).



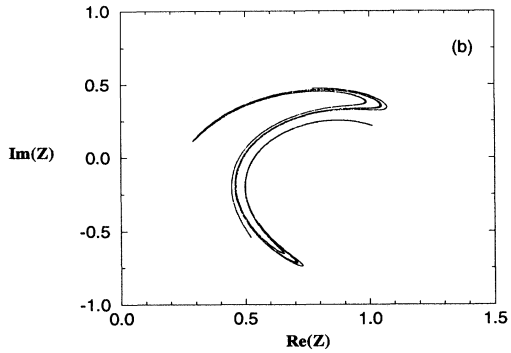


FIG. 18. (a) Noisy Ikeda attractor with $a=1.0$, $b=-0.7$, $\kappa=0.4$, and $\eta=6.0$. The observational noise is of amplitude 0.05, while the dynamical noise of amplitude 0.3 is coupled to both the real and imaginary parts of the Ikeda map; (b) the reconstructed Ikeda attractor with $a=1.006$ and $\eta=6.029$, see Table II. The clean Ikeda attractor with the target parameters is shown in Fig. 8(c).

The parameter fitting for the Hénon system is reasonably sound over the whole dynamical noise range in which the physical orbit does not diverge. This range, however, proved to be quite limited, therefore we proceeded to repeat these experiments with the Ikeda map, which can tolerate a much higher dynamical noise amplitude. This allows us to demonstrate the feasibility of also modeling substantial dynamical noise. We fit both a and η in the Ikeda map [30],

$$z_{n+1}=a-bz_n\exp\left[i\kappa-\frac{i\eta}{1+|z_n|^2}\right], \tag{13}$$

with $z=x+iy$. The target parameter values of the map were set to be $a_0=1.0$, $b_0=-0.7$, $\kappa_0=0.4$, and $\eta_0=6.0$. Its phase space portrait is given in Fig. 7(c). While the observational noise of amplitude 0.05 (about 10% of the signal) was modest, three different tests were run with dynamical noise amplitudes of $\sigma=0.1$, 0.2, and 0.3. This noise was coupled to both the real and imaginary parts of the Ikeda map, Eq. (13). At each separate noise level, we obtained a large number of realizations which satisfied the error constraint of $\mathcal{E}^{(\text{noise})}<0.01$. These three groups were then used to compute the mean and standard deviation of the parameter estimates and to estimate the

TABLE I. Reconstruction results for the Hénon system by including the observational noise level into the model. There is also dynamical noise of amplitude 0.01 (coupled to x component) and 0.1 (coupled to the y component of the Hénon map) in the target, which is responsible for the remaining small dc bias. The standard deviations of the reconstruction results are in parentheses.

Target			Reconstruction results		
a	b	Noise level	a	b	Noise level
1.790	−0.0385	1.0	1.781 260 (0.009 635)	−0.040 784 (0.003 054)	1.026 127 (0.145 554)
1.790	−0.0385	1.5	1.781 048 (0.014 619)	−0.040 805 (0.004 571)	1.507 154 (0.046 310)
1.790	−0.0385	2.0	1.781 131 (0.016 977)	−0.041 469 (0.006 114)	1.994 599 (0.037 629)

TABLE II. Reconstruction results for the Ikeda map by incorporating the dynamical noise into the model. There is also observational noise of amplitude 0.05 (about 10% of the signal amplitude) in the target. The standard deviations of the reconstruction results are in parentheses.

Target			Reconstruction results		
a	η	Noise level	a	η	Noise level
1.00	6.0	0.1	0.990 225 (0.012 101)	5.990 817 (0.061 517)	0.109 384 (0.011 739)
1.00	6.0	0.2	1.005 913 (0.013 724)	6.090 413 (0.097 861)	0.200 999 (0.008 038)
1.00	6.0	0.3	1.006 474 (0.017 266)	6.029 555 (0.049 497)	0.301 858 (0.025 179)

dynamical noise amplitude (Table II). The phase portrait of the noisy Ikeda map (target), and the reconstructed Ikeda map, are given in Fig. 18.

V. SUMMARY AND CONCLUSIONS

We have addressed the problem of reconstruction of chaotic dynamics from short and/or noisy data sets using the symbol sequence statistics. Our goal has been to develop a useful experimental technique. In particular, we have demonstrated how to extract parameter estimates for low-dimensional nonlinear dynamical systems using the symbol sequence statistics as the target. We find that this approach is highly robust even in the presence of observational and dynamical noise. Substantial noise produces a strong bias in the symbol sequence statistics, but the bias induced in the parameter estimates saturates quickly. We also showed that, even in the case where the noise is comparable to the signal, it is possible—by including the noise characteristics as part of the model—to produce robust reconstructions. Potential applications include nonlinear model validation, and nonlinear systems characterization, especially in circumstances where one has some knowledge of the class of model to be used and the data set is short and/or noisy.

In the more general *ab initio* or black box reconstruction problem, one begins only with a time series and a very general guess as to the class of models to be considered. Many techniques have been developed to identify the proper dimension of the system and eventually reconstruct a sensible phase space portrait [3]. Once this is done, one can expand the dynamics in some functional basis and the reconstruction amounts to finding the proper expansion coefficients. This is equivalent to a parameter fitting problem as discussed in the previous sections, but now in a much higher dimensional parameter space. In principal, therefore, our approach can be straightforwardly extended to this situation. In practice, of course, the computer CPU time required for convergence of the annealing calculation (or any other nonlinear optimization technique) quickly becomes prohibitive as the number of fitting parameters increases.

On the other hand, any expansion involves truncation at some order. This is not only a practical consideration, but also necessary to avoid overfitting. As one possible solution to choosing optimal models, Rissanen's minimum description length criterion [31] can be employed to guide such truncation. In fact, significant progress has been made recently by applying this criterion to chaotic signal modeling [2]. As an example of applying the symbolic approach described in this part to *ab initio* reconstruction, one could combine it with the measure-based orthonormal polynomial expansion method [32,33], while Rissanen's minimum description length criterion can be imposed to guard against overfitting. This is a topic for future study.

ACKNOWLEDGMENTS

We wish to acknowledge the support of the Physics Department of the College of William & Mary. X. Z. Tang would also like to thank Allen H. Boozer for discussions, support, and encouragement. A. D. Boozer and A. deBrauw were supported by the NSF Research Experience for Undergraduates Program.

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day, San Francisco, 1970).
[2] R. Brown, N. F. Rulkov, and E. R. Tracy, Phys. Rev. E **49**, 3784 (1994).
[3] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[4] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).
[5] A. B. Rechester and R. B. White, Phys. Lett. A **156**, 419 (1991).
[6] A. B. Rechester and R. B. White, Phys. Lett. A **158**, 51 (1991).
[7] U. Schwarz, A. O. Benz, J. Kurths, and A. Witt, Astron. Astrophys. **277**, 215 (1993).
[8] T. Elbert, W. J. Ray, Z. J. Kowalik, J. E. Skinner, K. E. Graf, and N. Birbaumer (unpublished).
[9] P. Cvitanović, G. H. Gunaratne, and I. Procaccia, Phys. Rev. A **38**, 1503 (1988).
[10] N. B. Tufillaro, P. Wycoff, R. Brown, T. Schreiber, and T. Molteno (unpublished).
[11] X. Z. Tang, E. R. Tracy, A. D. Boozer, A. deBrauw, and R. Brown, Phys. Lett. A **190**, 393 (1994).
[12] P. Grassberger and H. Kantz, Phys. Lett. A **113**, 235 (1985).
[13] J. P. Crutchfield and N. H. Packard, Physica D **7**, 201 (1983).
[14] Hao Bai-Lin, *Elementary Symbolic Dynamics* (World

Scientific, Singapore, 1989).
[15] V. M. Alekseev and M. V. Yakobson, Phys. Rep. **75**, 287 (1981).
[16] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[17] P. Grassberger and I. Procaccia, Physica D **13**, 34 (1984).
[18] A. Renyi, *Probability Theory* (North-Holland, Amsterdam, 1970).
[19] D. Ruelle, Bol. Soc. Bras. Math. **9**, 331 (1978).
[20] Ya. B. Pesin, Usp. Math. Nauk. **32**, 55 (1977).
[21] R. Bowen and D. Ruelle, Invent. Math. **29**, 181 (1975).
[22] S. V. Ershov, Phys. Lett. A **177**, 180 (1993).
[23] L. Ingber, Math. Comput. Modeling **8**, 29 (1993).
[24] M. Henon, Commun. Math. Phys. **50**, 69 (1976).
[25] J. A. C. Gallas, Phys. Rev. Lett. **70**, 2714 (1993).
[26] H. D. I. Abarbanel, R. Brown, and J. B. Kadtke, Phys. Rev. A **41**, 1782 (1990).
[27] T. Kai and K. Tomita, Prog. Theor. Phys. **64**, 1634 (1980).
[28] G. Mayer-Kress and H. Haken, Physica D **10**, 329 (1984); K. Matsumoto and I. Tsuda, J. Stat. Phys. **31**, 87 (1983); K. Matsumoto, *ibid.* **34**, 111 (1984); S. Doi, *ibid.* **55**, 941 (1989).
[29] Much broader search ranges and initial guesses were also tried. They yield similar results but required more computer time.
[30] S. Hammel, C. K. R. T. Jones, and J. Maloney, J. Opt. Soc. Am. B **2**, 552 (1985).
[31] J. Rissanen, Ann. Stat. **11**, 416 (1983); J. Rissanen, *Stochastic Complexity in Statistical Inquiry* (World Scientific, New Jersey, 1989).
[32] M. Giona, F. Lentini, and V. Cimagalli, Phys. Rev. A **44**, 3496 (1991).
[33] R. Brown, University of California, San Diego, 1992 (unpublished).

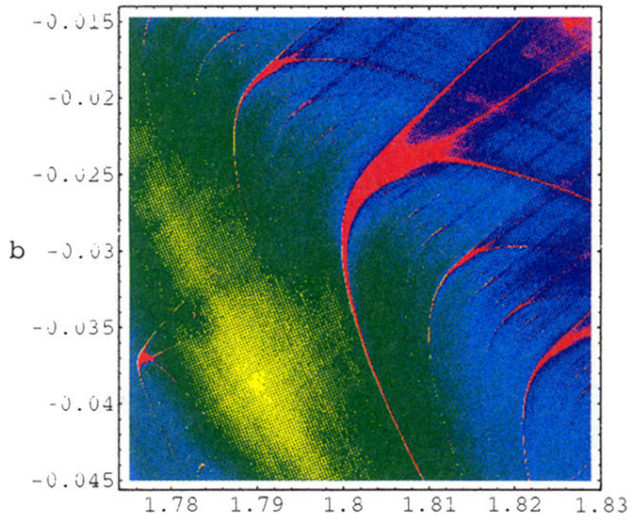


FIG. 2. The error landscape of the Hénon map ($a_0 = 1.79$, $b_0 = -0.0385$) is shown for the parameter range $a \in (1.775, 1.83)$ and $b \in (-0.045, -0.015)$. (a) Without noise, the 15th tree level is used (the minimum is at the center of the yellow area at the lower left while the maxima are red); (b) with observational noise of amplitude 0.5, the fifth tree level is used. The minima [now near the point around (1.784, −0.0365)] is in orange while red is still assigned for the maxima (shrimps).