J. Nonlinear Sci. Vol. 1: pp. 313–343 (1991)

Journal of Nonlinear Science

Noise Reduction in Chaotic Time Series Using Scaled Probabilistic Methods

P. F. Marteau[1] and H. D. I. Abarbanel[1,2]

[1] Institute for Nonlinear Science, University of California, San Diego, Mail Code R-002, LaJolla, CA 92093-0402, USA

[2] Department of Physics and Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, Mail Code R-002, La Jolla, CA 92093-0402, USA

Received October 8, 1990; accepted for publication April 17, 1991
Communicated by Stephen Wiggins

Summary. We present a probabilistic approach to the problem of additive source separation characterized by wide band power spectra when one of the sources is deterministic and low-dimensional (regular or chaotic). The algorithm is based upon a probabilistic analysis of local behavior in the phase space in which the deterministic source is embedded. It creates a kind of time domain "filter" for the noisy data from a *reference orbit* of the system, and it then corrects the observed noisy data in an incremental fashion to adjust it to match this filter in the sense of having as close as possible the same invariant distribution in the phase space. After describing the general method, we demonstrate its use on three familiar chaotic systems: first, two examples of chaotic maps (Hénon and Ikeda) in two dimensions and then a three-dimensional flow (the Lorenz system). The data from each of these is perturbed by uniformly distributed noise or by a spike signal, and we show that the method gives excellent results in recovering the original clean signal even when noise levels are quite large. Our present method appears to work even when the signal-to-noise ratio is quite small. The method works in cases where classical linear methods fail because both "signal" and "noise" are broadband spectrally.

Key words. signal separation in chaos, noise reduction, scaled probabilistic cleaning, maximum a posteriori probability methods

1. Introduction

One of the remarkable results of work in nonlinear dynamics and chaos over the past decade has been a reexamination of what we mean by "noise." The view is reasonably well established that "noise" is a high-dimensional system the dynamics of which are

probably not particularly interesting and which we characterize in some conventional statistical sense. The deterministic dynamics of a nonlinear system can look like conventional noise when displayed in conventional ways, namely, as time series or as the Fourier power spectrum of such time series. In experimental settings where deterministic chaotic signals are contaminated by high-dimensional dynamics (noise) of no interest to the observer, we require unconventional methods to separate the two types of broadband signal. This article presents a method for separating these sources using a recursive probabilistic, maximum likelihood approach. It appears in practice that the method works well even when the signal-(deterministic chaos) to-noise ratio is rather low. We have examined cases where the original signal-to-noise ratio is as low as a few dB, and see no particular barrier to applying the method to even lower signal-to-noise ratios.

Earlier work on this problem by Hammel [1], Farmer and Sidorowich [2], and Abarbanel, Marteau, and Sidorowich [3] has concentrated on the idea of recursively trying to satisfy the deterministic mapping governing the evolution of the system in its phase space. The map may be known beforehand or have been learned from either "clean" data, noisy data, or "on the fly." Another approach has been presented by Kostelich and Yorke [4] who utilize the observed data to make a local polynomial (in practice, linear) fit to the evolution of neighborhoods into themselves, and then by adjusting parameters in the map and in the best guess for the clean signal, they achieve a least squares estimate of the clean signal. This paper also treats some experimental data where the dynamics on the attractor has to be deduced "on the fly."

In this paper we present a quite different point of view on the problem of extracting a clean deterministic signal from an observed signal which is contaminated by additive noise. It takes into account the probabilistic properties of the attractor generated by the dynamical signal. These probabilistic properties are expressed in effect in terms of the invariant distribution of data points on the attractor [5,6]. We use the word probability here only in the sense that these invariant distributions act like probability distributions though they are sums over delta functions at the locations of the data and are thus quite singular in a mathematical sense. In the sense of useful quantities for numerical work we have no mathematical problems. Finite resolution of computations or observations smooths out the singular distributions quite nicely. The distributions are evaluated using a *reference orbit* which comes from earlier knowledge of the system or may be part of the noisy data set aside for this purpose. In this paper we work only with reference orbits coming from clean data and, not surprisingly, we find that the longer the reference orbit, the better we can do in separating the sources of signal and noise. In the case that the reference orbit comes from noisy data, we suggest that one might clean the data in sections using one section to act as the reference orbit and then substituting the newly cleaned orbit in its place. We have not tried this yet, but will return to it in the future.

We address two main issues in this paper:

- Separation of the sources which combine to produce the observed signal. In this case the task consists of recovering a signal which could be regular which is being observed in the presence of a low-dimensional chaotic signal. The contamination is the chaos, and we will be concerned with the case when the signal to noise ratio is quite low.

- Cleaning up the signal of the dynamical system itself by separating out the "noise" (high-dimensional dynamics). Experience demonstrates that the determination of invariants of the deterministic dynamics such as Lyapunov exponents [7], fractal dimensions, etc are quite sensitive to noise. These quantities are used to classify or characterize the dynamics, so if the problem at hand is identification of the system, the cleaning up of the signal, namely separating off the high-dimensional "noise," is critical.

In addressing both issues our procedure is essentially the same. We use the probability rules given in detail below to separate out a signal $\mathbf{S}(k); k = 1, 2, \ldots, N$, which satisfies some presumed, but not known, dynamics $\mathbf{S}(k+1) = \mathbf{F}(\mathbf{S}(k))$. The only knowledge of the dynamics is the observed reference orbit. In the first case listed we subtract the $\mathbf{S}(k)$ from the observations to recover the signal of interest. In the second case, $\mathbf{S}(k)$ is the signal.

We deal solely with *additive* noise in this paper. This means that the observed signal at "time" k, $\mathbf{O}(k)$, is composed of the clean signal $\mathbf{S}(k)$ plus some "noise" $\boldsymbol{\eta}(k)$

$$\mathbf{O}(k) = \mathbf{S}(k) + \boldsymbol{\eta}(k). \tag{1}$$

Other forms of noise are either additive in the dynamics itself or represent fluctuations in the parameters of the dynamics. At this time we have no methods for dealing with this kind of noise and restrict ourselves to additive noise. Additive noise is important in experimental settings as it represents the effect of the environment through which a signal must travel and/or the effects of the measurement instrument on the clean transmitted signal.

Our paper is organized as follows: In Section 2 we discuss the procedure in some generality with reference to and motivation from the application we have outlined. In Section 3 we give some details of the application of the method to the problem at hand. In Section 3 we also address the important issue of how we estimate the probabilities required for the algorithm. In Section 4 we present the details of our numerical experiments on two chaotic maps in two dimensions and on the Lorenz flow in three dimensions. In the final section we summarize this work and stress the open questions not addressed by our results.

It is useful to note now that the methods we describe here are likely to be complementary to the more precise methods which follow the stable and unstable manifolds of the dynamics [1,2,3]. Our present techniques will probably be useful when the signal-to-noise ratio of the observed data is rather low, and after these probabilistic algorithms have cleaned up the signal significantly, the sharper methods can take over and make the job as accurate as can be achieved by the machines available for the task.

2. Recursive Maximum Likelihood Principle for Noise Reduction

As indicated we are interested in a multivariate signal $\mathbf{S}(k)$ to be measured at a set of times $t_0 + k\Delta t$. The signal is in general a vector in a d-dimensional space

$$\begin{aligned} \mathbf{S}(t_0 + k\Delta t) &= \mathbf{S}(k) \\ &= [S_1(k), S_2(k), \ldots, S_d(k)], \end{aligned} \tag{2}$$

which, following standard practice in dynamical systems, we will usually reconstruct by time delay embedding from a scalar observation [5]. The full set of signals for $k = 1, 2, \ldots, N$ we call $\mathcal{S}_1^N = [\mathbf{S}(1), \mathbf{S}(2), \ldots, \mathbf{S}(N)]$. We assume that underlying the sequence of points $\mathbf{S}(k)$ is a deterministic evolution equation which is a map from R^d to itself

$$\mathbf{S}(k+1) = \mathbf{F}(\mathbf{S}(k)), \tag{3}$$

though we do not actually know the map or need to know it in what follows.

Our instruments do not report the sequence $\mathcal{S}_1^N$ since the observations are contaminated by noise $\boldsymbol{\eta}(k)$. The actual observations are the d-dimensional vectors $\mathbf{O}(k)$ related to the clean signal by $\mathbf{O}(k) = \mathbf{S}(k) + \boldsymbol{\eta}(k)$. In what follows we assume the noise to be bounded by some magnitude σ_η for all times, and we take the noise to be uniformly distributed and of zero mean. This assumption can easily be replaced by other assumptions about the distribution of the noise, but the boundedness, which is physically sensible, is critical. Further we assume the $\boldsymbol{\eta}(k)$ represent identical but statistically independent quantities at each "time" k. Our noise is what is usually called iid. For measurement time step Δt large compared to response times of the medium which contaminates the signal or large compared to the recovery time of an instrument, this iid assumption is quite sensible. It will become clear that the iid assumption could be relaxed, but we would need to specify something like correlation times among the measurements in order to proceed.

Now we are interested in analyzing the conditional probability of the state of the system actually being $\mathcal{S}_1^N$ when the observations are reported as $\mathbb{O}_1^N$. We call this conditional probability $P(\mathcal{S}_1^N \mid \mathbb{O}_1^N)$, and we seek the maximum of this conditional probability over the possible values of the states $\mathcal{S}_1^N$. That is, we are seeking the most probable state sequence Σ_1^N which satisfies

$$P\left(\Sigma_1^N \mid \mathbb{O}_1^N\right) = \max_{\mathcal{S}_1^N} P(\mathcal{S}_1^N | \mathbb{O}_1^N). \tag{4}$$

To determine the maximally likely sequence we use the definition of conditional probability to write

$$\begin{aligned} P(\mathcal{S}_1^N | \mathbb{O}_1^N) &= \frac{P(\mathbb{O}_1^N | \mathcal{S}_1^N) P(\mathcal{S}_1^N)}{P(\mathbb{O}_1^N)} \\ &= \frac{P(\mathcal{S}_1^N, \mathbb{O}_1^N)}{P(\mathbb{O}_1^N)}, \end{aligned} \tag{5}$$

where the joint probability for the observations and the actual sequence has been written as $P(\mathcal{S}_1^N, \mathbb{O}_1^N)$. Since the observations are given, we require expressions for the two probabilities $P(\mathbb{O}_1^N \mid \mathcal{S}_1^N)$ and $P(\mathcal{S}_1^N)$ which enter $P(\mathcal{S}_1^N, \mathbb{O}_1^N)$, and it is this joint probability we seek to maximize over the possible choices for $\mathcal{S}_1^N$ to find Σ_1^N.

For this purpose we require two properties for the process we are considering. *First* we note that the probability of the sequence $\mathcal{S}_1^m$, which is just the invariant distribution on the attractor evaluated on the orbit $\mathcal{S}_1^m$, can be written, by definition, as

$$P(\mathcal{S}_1^m) = P(\mathbf{S}(m) \mid \mathcal{S}_1^{m-1}) P(\mathcal{S}_1^{m-1}), \tag{6}$$

and by the assumption that the clean data sequence satisfies the deterministic recursion relation we see that the conditional probability here takes the simpler form

$$P(\mathbf{S}(m) \mid \mathcal{S}_1^{m-1}) = P(\mathbf{S}(m) \mid \mathbf{S}(m-1)), \tag{7}$$

so we have

$$P(\mathcal{S}_1^m) = P(\mathbf{S}(m) \mid \mathbf{S}(m-1)) P(\mathcal{S}_1^{m-1}). \tag{8}$$

Second, by the assumption of *additivity* of the noise to the signal and the iid nature of the noise we may write

$$\begin{aligned} P(\mathcal{O}_1^m \mid \mathcal{S}_1^m) &= P(\mathbf{O}(1), \mathbf{O}(2), \ldots, \mathbf{O}(m) \mid \mathbf{S}(1), \mathbf{S}(2), \ldots, \mathbf{S}(m)) \\ &= \prod_{k=1}^{m} P(\mathbf{O}(k) | \mathbf{S}(k)) \\ &= \prod_{k=1}^{m} P_\eta(\mathbf{S}(k) - \mathbf{O}(k)) \\ &= P(\mathbf{O}(m) \mid \mathbf{S}(m)) P(\mathcal{O}_1^{m-1} \mid \mathcal{S}_1^{m-1}) \end{aligned} \tag{9}$$

where P_η is the probability distribution assumed for the noise.

These properties lead us to the recursion relation for the joint probabilities

$$P(\mathcal{S}_1^m, \mathcal{O}_1^m) = \{P(\mathbf{O}(m) \mid \mathbf{S}(m)) P(\mathbf{S}(m) \mid \mathbf{S}(m-1))\} P(\mathcal{S}_1^{m-1}, \mathcal{O}_1^{m-1}), \tag{10}$$

and it is this we use to maximize over the $\mathcal{S}_1^m$ to find the most likely state Σ_1^m.

Now we divide the optimization procedure over the allowed states $\mathcal{S}_1^N$ into the effect at "time" m due to states in the past and the effect due to states at later times. Since we are using properties of the *stationary* time series which defines the time independent geometric structure of the attractor, this allows us to provide a balanced sense of the effect of all points on the transition probabilities at time m. To this end we define the optimal forward probability ending at state $\mathbf{S}(m)$

$$PF(\mathbf{S}(m)) = \max_{\mathcal{S}_1^{m-1}} P(\mathbf{S}(m) \mid \mathbf{S}(m-1)) P(\mathcal{S}_1^{m-1}, \mathcal{O}_1^{m-1}), \tag{11}$$

and using the recursion relation derived above among the joint probabilities $P(\mathcal{S}_1^m, \mathcal{O}_1^m)$, we can derive the recursion relation for the forward probabilities

$$PF(\mathbf{S}(m)) = \max_{\mathcal{S}(m-1)} [\{P(\mathbf{S}(m) \mid \mathbf{S}(m-1)) P(\mathbf{O}(m-1) \mid \mathbf{S}(m-1))\} PF(\mathbf{S}(m-1))], \tag{12}$$

From these forward probabilities we find the maximized joint probability as

$$P(\Sigma_1^m, \mathcal{O}_1^m) = \max_{\mathbf{S}(m)} [P(\mathbf{O}(m) \mid \mathbf{S}(m)) PF(\mathbf{S}(m))]. \tag{13}$$

The identification of $PF(\mathbf{S}(m))$ has allowed us to decompose the search among Nd variables into a sequence of searches over local d-dimensional spaces. This is the essence of the procedures called dynamic programming by Bellman [8].

Similarly, we can define the optimal backward state sequence probabilities $PB(\mathbf{S}(m))$ as a recursion relation starting from state N and moving backward to state m. This reads

$$PB(\mathbf{S}(m)) = \max_{\mathbf{S}(m+1)} [\{P(\mathbf{S}(m) \mid \mathbf{S}(m+1))P(\mathbf{O}(m+1) \mid \mathbf{S}(m+1))\} PB(\mathbf{S}(m+1))], \tag{14}$$

and the joint probability arises from $PB(\mathbf{S}(m))$ as

$$P\left(\Sigma_m^N, \mathbb{O}_m^N\right) = \max_{\mathbf{S}(m)} [P(\mathbf{O}(m) \mid \mathbf{S}(m)) PB(\mathbf{S}(m))] . \tag{15}$$

We will actually use a combination of these two methods and derive the maximum likelihood state sequence Σ_1^N from the forward-backward algorithm which maximizes at each "time" m the product

$$P(\mathbf{O}(m) \mid \mathbf{S}(m)) PF(\mathbf{S}(m)) PB(\mathbf{S}(m)) , \tag{16}$$

and the maximum over this gives $P(\Sigma_1^N, \mathbb{O}_1^N)$.

The use of the forward and backward information is motivated by two aspects of the dynamics we are studying:

- The geometric structure of the attractor which defines the invariant probability distribution is a time independent quantity. Information about local structure in the state space comes from the neighbors of points along the orbits, and those neighbors have no particular temporal order to their appearance in a neighborhood. Iterating forward and backward to a "time" m to determine the optimal value of $\mathbf{S}(m)$ makes use of all the data which is important for the neighborhood of $\mathbf{S}(m)$. This would not be the case if we used only forward or only backward iterations. It may be that combinations other than the one we choose of forward and backward information are as useful or even more so in this regard, and they are certainly worth exploring.
- In dynamical systems which exhibit chaos and strange attractors forward and backward iterations are not equivalent. The Lyapunov exponents which govern the growth of perturbations forward and backward are generally different. Since we must, as we shall see in the next Section, make an explicit choice on how we discretize the phase space in order to carry out our procedures in practice, the errors which inevitably result from this finite resolution behave differently under forward and under backward iterations. The quantity which determines the ability to accurately predict ahead is the Kolmogorov-Sinai entropy, and its relation to Lyapunov exponents is discussed by Pesin and by Ruelle [5, 9, 10]. Our present choice is a kind of compromise between the forward and backward procedures.

The three methods—forward, backward and both—for passing through the data are formally the same in the sense they represent rearrangements of the same overall formula for $P(\Sigma_1^N, \mathbb{O}_1^N)$. Nonetheless, one might suspect that they could give different answers, if the the data set is not very long. In the limit that the data set is very long, $N \to \infty$, they should all agree, since any dependence on "end" effects should go away. We used the forward/backward algorithm in all the work reported in this paper, but have used both the forward and the backward in some of our investigations, and cannot report any significant differences in their performance. Of course, if the data

set is too short, the effects of the beginning points are important, and then the three methods may give substantially different results.

3. Implementing the Algorithm for Observed Time Series of Dynamical Systems

3.1 Estimating the Probabilities

To apply our method to a time series from a dynamical system we require a way to estimate the two ingredients in the recursion formulae. These are the conditional probability for the state $\mathbf{S}(m)$ given $\mathbf{S}(m \pm 1)$ and the conditional probability for the observation $\mathbf{O}(m)$ given the state $\mathbf{S}(m)$. The latter, $P(\mathbf{O}(m) \mid \mathbf{S}(m))$, we have discussed above and indicated how it is just the distribution for the noise evaluated at $\mathbf{S}(m) - \mathbf{O}(m)$

$$P(\mathbf{O}(m) \mid \mathbf{S}(m)) = P_\eta(\mathbf{O}(m) - \mathbf{S}(m)), \tag{17}$$

since the noise is additive.

For the conditional probability $P(\mathbf{S}(m) \mid \mathbf{S}(m-1))$ we require a way to evaluate the joint probability for finding $\mathbf{S}(m)$ on the attractor at time m and finding $\mathbf{S}(m-1)$ on the attractor at time $m-1$. This is to be determined from the definition of the invariant distribution on the attractor which states that at a point $\mathbf{x}$ the density is

$$\rho(\mathbf{x}) = \frac{1}{N_R} \sum_{j=1}^{N_R} \delta^d(\mathbf{x} - \mathbf{R}(j)) \tag{18}$$

as the number of points on the *reference* orbit $N_R \to \infty$, that is, becomes very large. $\mathbf{R}(j)$ is a *reference* observed orbit of the system. The joint probability would then be

$$\rho(\mathbf{x}, \mathbf{y}) \propto \sum_{j',j=1}^{N_R} \delta^d(\mathbf{x} - \mathbf{R}(j)) \delta^d(\mathbf{y} - \mathbf{R}(j')). \tag{19}$$

In practice, we cannot use the delta functions to evaluate these probability distributions, so we estimate them using a kernel density estimate [11] which replaces the delta function by a smooth "kernel" $K(\mathbf{x})$ which has the same integral and has some width which is in the hands of the user. This width reflects the graininess of our knowledge of phase space. The joint density then becomes proportional to

$$\sum_{j',j=1}^{N_R} K(\mathbf{x} - \mathbf{R}(j)) K(\mathbf{y} - \mathbf{R}(j')), \tag{20}$$

and the conditional probability we need to estimate is

$$P(\mathbf{S}(m) \mid \mathbf{S}(m \pm 1)) \propto \frac{\sum_{j',j=1}^{N_R} K(\mathbf{S}(m) - \mathbf{R}(j)) K(\mathbf{S}(m \pm 1) - \mathbf{R}(j'))}{\sum_{n=1}^{N_R} K(\mathbf{S}(m \pm 1) - \mathbf{R}(n))}. \tag{21}$$

The main contribution to the sum in this numerator will come when m is near one of the "times" j, and the "time" j is near $m \pm 1$.

We have chosen to use for our kernel the exponential function of distance

$$K(\mathbf{x} - \mathbf{y}) = \exp\left(\frac{-|\mathbf{x} - \mathbf{y}|^2}{\epsilon}\right), \tag{22}$$

where ϵ is a scaling of the rms error size σ_η as we will discuss in a moment. The normalization of this kernel is absorbed in the constant of proportionality above.

Now we have assumed that the noise level is bounded which, while not true for a Gaussian distribution of noise, must be true physically. Our choice of a uniform distribution of noise is not particularly limiting, but clearly may be exchanged for other assumptions as the circumstance warrants. The boundedness is critical for us. Because of this boundedness we can easily discretize the phase space locally and capture the effects of noise while making sufficient opportunity for choice among possible actual states of the system. We establish a region or ball B_k centered around the observed data point $\mathbf{O}(k)$. It has radius σ_η, and we divide it into M subballs of size $\epsilon = \sigma_\eta / M^{1/d}$ in d-dimensional phase space. The centers of these ϵ sized balls are the possible values $\mathbf{S}(k, i); k = 1, 2, \ldots, N; i = 1, 2, \ldots, M$ allowed to the states of the system.

3.2 Details of the Algorithm

We implement the algorithm described in the previous section as follows.

- First we evaluate the set of $M(N-1)$ forward probabilities $PF(\mathbf{S}(k, i))$ using the recursion relation

$$PF(\mathbf{S}(k, i)) = \max_{\mathbf{S}(k-1, i')} \left[\{P(\mathbf{S}(k, i) \mid \mathbf{S}(k-1, i'))P(\mathbf{O}(k-1, i') \mid \mathbf{S}(k-1, i'))\}PF(\mathbf{S}(k-1, i'))\right]. \tag{23}$$

  We have to start with an assumed distribution $PF(\mathbf{S}(1, i))$ at the initial time $k = 1$, and we take this to be uniform over the $i = 1, 2, \ldots, M$ subballs around the observation point $\mathbf{O}(1)$. Now we evaluate M^2 connection probabilities from the M subballs near $\mathbf{O}(1)$ at locations $\mathbf{S}(1, i); i = 1, 2, \ldots, M$ to the subballs near $\mathbf{O}(2)$ at locations $\mathbf{S}(2, i'); i' = 1, 2, \ldots, M$. For each of the M, $\mathbf{S}(2, i')$ we record the maximum values and then have our $PF(\mathbf{S}(2, i'))$, which we tabulate and proceed. This set of actions is represented in a pictorial fashion in Figure 0.
- The backward probabilities $PB(\mathbf{S}(k, i))$ are computed by starting at time $k = N$ with an assumed uniform distribution in the subballs near $\mathbf{O}(N)$. Using the backward recursion formula

$$PB(\mathbf{S}(k, i)) = \max_{\mathbf{S}(k+1, i')} \left[\{P(\mathbf{S}(k, i) \mid \mathbf{S}(k+1, i'))P(\mathbf{O}(k+1, i') \mid \mathbf{S}(k+1, i'))\}PB(\mathbf{S}(k+1, i'))\right], \tag{24}$$

  step by step leads to $M(N-1)$ probabilities $PB(\mathbf{S}(N-1, i)), \ldots, PB(\mathbf{S}(1, i'))$.

- Altogether now we have recorded $2M(N-1)$ probabilities counting the forward and backward lists. So far we have made two passes through the time series.
- Our final pass through the time series now looks at each time not equal to the forward starting time $k = 1$ or the backward starting time $k = N$. We establish the most likely set of $\mathbf{S}(k, i); k = 2, 3, \ldots, N-1; i = 1, 2, \ldots, M$ by maximizing the product

$$P(\mathbf{O}(k) \mid \mathbf{S}(k, i))PF(\mathbf{S}(k, i))\, PB(\mathbf{S}(k, i)), \tag{25}$$

over the set of M, $\mathbf{S}(k, i); i = 1, 2, \ldots, M$ at each $k = 2, 3, \ldots, N-1$. This will give us the optimum or maximum likelihood value $\Sigma(k)$. The collection of these vectors for $k = 2, 3, \ldots, N-1$ is our cleaned up orbit, $\Sigma(k)$.

This completes one pass through the data. Now the procedure is repeated with the original data $\mathbf{O}(k)$ replaced by the values Σ_1^N with $\Sigma(1) = \mathbf{O}(1)$ and $\Sigma(N) = \mathbf{O}(N)$, since no cleaning was done on the ends. In this pass through the algorithm we scale the noise level σ_η by a factor we call s which we have chosen to slowly reduce the size of our estimate of σ_η as we repeat the procedure. Typically we have chosen s to be 1.1, so our reduction in the size of the phase space in which we allow the clean data point to be located only slowly decreases. If we took larger values of s in an attempt to decrease the number of iterations of the recursion relations, we found that there was a tendency to get stuck in a corner of phase space not much connected with the original orbit. Slowly approaching smaller and smaller noise balls worked quite well as we shall report in the next section.

The search algorithm outlined is thus repeated with a new set of $\mathbf{O}(k)$ as the observations, namely the set of cleaned data except for the end points, and the size of the noise is scaled down $\sigma_\eta \to \sigma_\eta / s$. We repeat this scaled probabilistic procedure until an acceptable level of clean data is achieved or until changes in the cleaned orbit $\Sigma(k)$ are small from pass to pass. In this paper we fixed the number of passes by trial and error and did not investigate the possibility of varying the scaling factor and the number of passes. Our choice was more than adequate, as we shall see.

This sequence of steps we call *scaled probabilistic noise reduction.*

3.3 Phase Space Reconstruction and Markov Properties of the Method

It is often true that we are given only measurements of a single scalar variable, $x(k)$, from the system. We must then reconstruct the phase space or state space of the d-dimensional vectors $\mathbf{S}(k)$, and we do this in the familiar fashion suggested by Ruelle and given firm mathematical foundations by Takens and Mañé [12, 13, 5] using a time delay embedding method which gives $\mathbf{S}(k)$ as

$$\mathbf{S}(k) = [x(k), x(k-T), \ldots, x(k-T(d-1))]\,. \tag{26}$$

In this paper we do not address methods for choosing either the time delay T or the embedding dimension d. A thorough discussion can be found in [7] and the references there.

This choice means that the one step transition probability $P(\mathbf{S}(k) \mid \mathbf{S}(k-1))$ as seen in d-dimensional space is (setting $T = 1$)

$$P(\mathbf{S}(k)|\mathbf{S}(k-1)) = P(x(k), x(k-1), \ldots, x(k-d+1)|x(k-1), \ldots, x(k-d)) = P(x(k)|x(k-1), x(k-2), \ldots, x(k-d)), \quad (27)$$

and this makes the probability $P(\mathcal{S}_1^N)$ factorize as a d-step Markov process as seen in the space of the scalar variables $x(k)$. This causes no special problem, of course, and needs to be tracked when we are making our search.

If the embedding dimension is too small to capture the geometry of the attractor, that is, to unfold it completely, then we will be misled as far as distances go in computing the kernels required for the probability distributions. There is, of course, the sufficient condition (for clean data) that d be the integer greater than or equal to twice the dimension of the attractor plus unity, but this may often be larger than is required. Indeed, it is clear that as the embedding dimension approaches the correct minimum dimension to unfold the attractor completely, the part of phase space where distances are being computed incorrectly becomes quite small, so it may be that reaching even this minimum dimension is not required for practical results. In our work on the Ikeda map, for example, it is known that the minimum embedding dimension should be $d = 4$ [7], but at $d = 2$ significant noise reduction is achieved, and at $d = 3$ nearly 70% of the full noise reduction at $d \geq 4$ is achieved. In this paper we are working with known systems, and we do not independently determine the value of d here but take it from previous work on these systems.

In practice, the real "clean" data does not exactly satisfy the deterministic mapping $\mathbf{S}(k+1) = \mathbf{F}(\mathbf{S}(k))$, and correlations among the data points beyond that introduced by the dynamics surely exist. A clear future direction of investigation would be to relax the strict one-step Markov property of the strictly deterministic data and to investigate how well one can do by allowing it to be two-step, three-step, and so forth.

4. Numerical Experiments

Our procedure in each of the cases we present now has been to generate a clean reference orbit, $\mathbf{R}(j)$, which we need for the estimates of the conditional probabilities or transition probabilities required in the algorithm. Then we generate another orbit which we call the clean orbit, $\mathbf{Sc}(k)$, starting from different initial conditions. To this clean orbit we add a second source signal, the contaminant, which is our iid bounded noise or either a "spike" or pulse or similar regular signal. In the case of the iid noise our goal is to remove it from the observed signal, now $\mathbf{Sc}(k) + \boldsymbol{\eta}(k)$, and find $\Sigma(k)$ which is our approximation to the original clean signal. In the case of a spike or other regular signal, the "noise" is the chaotic signal. The task we set ourselves is the same in each case: *Separate the two sources when only the sum has been observed, given the reference orbit.*

4.1 The Computing Cost of the Algorithm

In estimating the transition or conditional probabilities we require a computation of the distances in d-dimensional space between reference vectors and our candidates $\mathbf{S}(k, i)$

for the optimum state vector. We are using an exponential kernel for estimating the probabilities, and we retain in the sum involving the kernel only those terms with points in the immediate neighborhood of the $\mathbf{S}(k)$. We used a kd-tree algorithm [14] for identifying these neighborhoods of the reference orbit, and this allows us to make the search for neighbors in logarithmic time.

The local state space is partitioned by $M = L^d$ subballs where L is the number of subdivisions in each of the d dimensions of the space. If σ is the radius of the noisy ball around each noisy point, the subballs are of radius $\epsilon = \sigma/L$.

The most time-consuming part of the procedure is the estimation of the transition probabilities. If the time series is of length N, then about $dN\log(N)$ time is spent in the search for neighbors. We must evaluate M^2 transition probabilities at each stage, but we only have to search the neighbors for M states. If we want to clean $N' \leq N$ points on the orbit, the computation time will be proportional to $2N'L^d dN\log(N)$, with the factor of two coming because we search backward and forward. The most costly factor in this estimate is the term L^d which may prevent the use of our method in very large dimensions since, obviously, $L > 1$. In this paper we are working with $d \leq 4$, and small L, so this has not been an issue.

4.2 Various Parameters, Error Measures, and Cleaning Criteria

In all the numerical work which follows we chose the following values for various parameters in the algorithm:

- $L = 3$
- The number of neighbors used to estimate the transition probabilities is set at 20.
- The noise $\boldsymbol{\eta}(k)$ is uniformly distributed and has zero mean.
- The scale factor used in reducing the size of the search region in state space after each pass through the algorithm was $s = 1.1$ We have used 35 passes through the data to allow this small scaling factor to have its effect. We have not systematically explored the possibility of increasing s and decreasing the number of passes to achieve the same level of cleaning. As we noted above, however, larger values of s tend to trap one in unwanted parts of local phase space too early in the recursion procedure, and escape from these traps appears infrequent.
- Unless specified otherwise we use 10,000 points on our reference orbits. In practice, we expect one could easily use far fewer points, but we have not experimented with the behavior of the procedure as this number is reduced.

To test how well we have done in cleaning the noisy data we evaluate two kinds of deviation from known values. Because we know the original dynamics $\mathbf{S}(k+1) = \mathbf{F}(\mathbf{S}(k))$, we can ask how well the cleaned data satisfies this map. The error here we call the *deterministic error*. Locally in "time" it is

$$DE(k) = \left|\Sigma(k+1) - \mathbf{F}(\Sigma(k))\right|, \tag{28}$$

and Euclidian distance in d-dimensional space is used. We evaluate the average of this error over the whole data set, or, equivalently, across the whole attractor. The

experience we have had with our cleaning procedure and the experience we and others [1, 2, 3] have had with more dynamically based methods produces deterministic errors which are very small except at a few locations around the attractor. This means that the usual arithmetic average

$$\frac{1}{N}\sum_{k=1}^{N} DE(k) \tag{29}$$

can be quite misleading about how well we have done in our cleaning process, so we also compute the geometric average

$$\left\{\prod_{k=1}^{N} DE(k)\right\}^{\frac{1}{N}}, \tag{30}$$

which tells us much more about the trend of series such as $DE(k)$ which has enormous excursions from its general behavior. We shall actually report the logarithms of these averages

$$\begin{aligned} ADE &= \log\left[\frac{1}{N}\sum_{k=1}^{N} DE(k)\right] \\ GDE &= \log\left[\left\{\prod_{k=1}^{N} DE(k)\right\}^{\frac{1}{N}}\right] \\ &= \frac{1}{N}\sum_{k=1}^{N} \log[DE(k)]. \end{aligned} \tag{31}$$

We are also in the fortunate situation in these numerical experiments of knowing the clean data $\mathbf{Sc}(k)$, so we can compare the cleaned orbit $\Sigma(k)$ with the original clean data using the local *absolute error*

$$AE(k) = \left|\Sigma(k) - \mathbf{Sc}(k)\right|, \tag{32}$$

and its averages

$$\begin{aligned} AAE &= \log\left[\frac{1}{N}\sum_{k=1}^{N} AE(k)\right] \\ GAE &= \log\left[\left\{\prod_{k=1}^{N} AE(k)\right\}^{\frac{1}{N}}\right] \\ &= \frac{1}{N}\sum_{k=1}^{N} \log[AE(k)]. \end{aligned} \tag{33}$$

To give a quantitative statement about the success of the cleaning method we evaluate the following as a measure of how much noise is removed. Define the local deterministic error of the noise $DEN(k) = |\mathbf{Sn}(k+1) - \mathbf{F}(\mathbf{Sn}(k))|$, where $\mathbf{Sn}(k) = \mathbf{Sc}(k) + \boldsymbol{\eta}(k)$, and the local absolute error of the noise $AEN(k) = |\mathbf{Sc}(k) - \mathbf{Sn}(k)|$. The latter is just the original noise level. Now we report the following numbers [4]

$$RAE = 1 - \frac{\sqrt{\sum_{k=1}^{N} AE(k)^2}}{\sqrt{\sum_{k=1}^{N} AEN(k)^2}}, \tag{34}$$

which is the measure of how the absolute error has been improved, and

$$RDE = 1 - \frac{\sqrt{\sum_{k=1}^{N} DE(k)^2}}{\sqrt{\sum_{k=1}^{N} DEN(k)^2}}, \tag{35}$$

which is a measure of the effectiveness of the cleaning in removing deterministic error.

4.3 Results of Computations

4.3.1 Hénon Map The Hénon map [15] of the plane to itself is

$$\begin{aligned} x(k+1) &= 1 + y(k) - ax(k)^2 \\ y(k+1) &= bx(k), \end{aligned} \tag{36}$$

where we use the traditional values $a = 1.4$ and $b = 0.3$. The phase portrait of the attractor is familiar, and we show it in Figure 1 for convenient comparison with our noisy orbit and our optimally cleaned orbit.

We added noise uniformly distributed in the interval $[-0.2, 0.2]$, which means a 15.8% noise perturbation, computed as the ratio of the rms value of this noise sequence (0.115) and the rms value of the clean data sequence (0.73). This is a signal to noise ratio of about 8 dB. We compute dB values as $10 \log_{10}$ of the ratio of relevant *amplitudes*. Power dB values are gotten by multiplication by 2. Figure 2 shows the orbit after the noise has been added, this is $\mathbf{Sn}(k)$, and Figure 3 shows the phase plot after we have applied the cleaning algorithm our standard 35 times. In the cleaning process used here we worked directly with the two-vectors $[x(k), y(k)]$ rather than using phase space reconstruction vectors $[x(k), x(k+1)]$. This is a trivial difference for the Hénon map since $y(k+1) = bx(k)$.

In Figure 4 we show the absolute error and the deterministic error after the 35 cleaning passes for a limited section of the 1000 step data sequence which we cleaned. Again we worked here with the original two-vectors from the Hénon data. The comparison between the two kinds of error shows that even when the absolute error (shown with a solid line) is large, the deterministic error can remain quite small. In Figure 5

we exhibit the local absolute error $AE(k)$ which results from the cleaning procedure, and in Figure 6 is the local deterministic error $DE(k)$. The various averaged error measures for this process are $ADE = -3.61$; $GDE = -2.51$; $AAE = -3.09$; $GAE = -2.41$; $RDE = 97.3\%$; and $RAE = 78.3\%$. We can translate the numbers *AAE,* and *GAE* into more familiar terms recognizing that they are measures of the residual error. For $AAE = -3.09$ as here, this means the residual error is 0.000813 compared to the original error of 0.115. This is a noise reduction or gain in signal-to-noise ratio of 21.5 dB, so the final signal to noise ratio is 29.5 dB compared to the original 8.0 dB. If we use the geometric mean as the measure of noise reduction, the noise reduction or gain in signal against noise is found to be 14.7 dB for a final signal to noise ratio of 22.7 dB. Recall we use $10\log_{10}$ in amplitude as our measure of dB. The RAE figure of merit is discussed in [4].

Next we perturbed the Hénon map with a spike signal of magnitude 0.2 localized around time step 50. In Figure 7 we show as a solid line the original spike, which was then used as an additive contamination of the $x(n)$ part of the Hénon orbit, and as a solid line with stars we have in the same Figure the spike reconstructed after our cleaning process. The reconstructed spike is found by cleaning the "observed" spike plus clean Hénon data using our algorithm and then subtracting the cleaned data from the observations. In Figure 8 we show the time series for the Hénon data without the spike and with the spike of amplitude 0.2 at step 50 added to the clean data. Clearly in the vicinity of step 50 there is a small but visible perturbation of the original signal. Considering the spike as the signal and the Hénon orbit as the noise gives an original signal-to-noise ratio of -5.6 dB, using the maximum signal strength of 0.2 as the amplitude of the signal and the rms level of the Hénon orbit as the "noise" level. Even lower signal-to-noise ratios give much the same result as long as the reference orbit is long enough to yield a good representation of the attractor at small scales in the phase space. The various averaged error measures for this process are $ADE = -3.01; GDE = -3.29; AAE = -3.17; GAE = -3.99; RDE = 91.1\%$; and $RAE = 80.0\%$. Using the *AAE* as the measure of the accuracy with which we recover the spike, we could now say that the signal to noise level is 23.1 dB at the end of our cleaning process. This means a gain of 28.7 dB in the signal against the Hénon "noise."

In the next test of our algorithm on the Hénon map we added a pulse signal to one generated from the Hénon map. Again the pulse was added to the $x(n)$ component of the Hénon data. This pulse was localized around "time" step 40 and had amplitude 0.2. In Figure 9 we show the original pulse as a solid line, and the resulting pulse after the usual 35 passes through our cleaning algorithm is shown as the solid line with stars. In this case we used only $x(n)$ data from the Hénon orbit and reconstructed the attractor using time delay embedding in a dimension of 2. We also used 20,000 points on the reference orbit. The original signal to noise ratio was, as above, -5.6 dB and, after cleaning the signal strength relative to the residual error, is 32.9 dB for a gain in signal to noise ratio of 38.5 dB. In terms of the error measures defined above we find after cleaning: $AAE = -3.99; GAE = -3.18$; and $RAE = 84\%$. The measures of error or improvement for deterministic error for a pulse or spike are not so relevant as when we are cleaning up a chaotic signal to see how we are doing in satisfying the map. We could, of course, evaluate how well we recovered the original

Hénon "noise" source, but our emphasis in this set of experiments is in recovering the spike signal.

For an illustration of the importance of the quality of the reference orbit used in determining the conditional probability densities we show in Figure 10 the same pulse as an original (solid line) and cleaned pulse (solid line plus stars) when only 10,000 points on the reference orbit are used. Clearly one does much better when more points on the reference orbit are available.

4.3.2 The Ikeda Map The Ikeda [16] map of the plane to itself arose in studies of properties of a laser in a cavity. It is given in terms of a complex number $z(k)$ at each time step

$$z(k+1) = p + Bz(k)\exp[i(\kappa - \frac{\alpha}{1+|z(k)|^2})], \tag{37}$$

and we have used the parameter values $p = 1.0, B = 0.9, \alpha = 0.4$, and $\kappa = 6.0$. This attractor is much richer than the Hénon attractor, and provides a more complex challenge for our cleaning algorithm. The dimension of the attractor is about 1.7, and the dimension of the embedding space required using time delay embedding is at least 4 [7]. Again for convenience we show the original orbit in Figure 11. This figure is not original, of course, and is here for ease of comparison with our cleaning results.

To this clean orbit we added uniform noise with maximum magnitude 0.1. This is a noise perturbation of about 11.8% when rms levels of clean data (0.489) and rms levels of noise (0.0577) are compared, or an original signal-to-noise ratio of about 9.3 dB. Figure 12 and Figure 13 show respectively the noisy phase portrait and the phase portrait resulting from 35 passes of the cleaning procedure. In this cleaning procedure we used the original two-vectors $[x(k), y(k)]$ from the Ikeda map. In Figure 13 we have cleaned 3000 data points rather than our usual 1000 to be able to expose the way in which the cleaning procedure reveals the details of the structure of the Ikeda attractor. In Figure 14 we have the local absolute error and in Figure 15, the local deterministic error. The various averaged error measures for this process are $ADE = -1.71$; $GDE = -2.21$; $AAE = -3.23$; $GAE = -1.91$; $RDE = 88.6\%$; and $RAE = 60.8\%$. The final signal-to-residual error ratio is now 29.1 dB (using AAE) for a gain in the signal to noise ratio of 19.9 dB.

To an orbit of the Ikeda map we added to $x(n)$ our pulse of magnitude 0.2 localized near time step 40. This is an original signal to noise ratio of -3.9 dB. In Figure 16 we reproduce the original pulse as the solid line and the reconstructed pulse after the scaled probabilistic cleaning has been done as the solid line with stars. This calculation was done using the real part of the Ikeda orbit ($x(k)$) with a phase space reconstruction done in $d = 4$ to calculate the distances required in the conditional probability estimations [7], and 20,000 points on the reference orbit were used. The error measures indicated above were $AAE = -3.06; GAE = -3.96$; and $RAE = 79.7\%$. Using AAE as the measure of the improvement in the noise level we compute a final signal to noise ratio of 23.6 dB or a gain in this ratio of 27.5 dB.

4.3.3 The Lorenz Attractor As our final example we studied the usefulness of our algorithm in recapturing an orbit on the Lorenz attractor [17]. The orbits are the solution of the three ordinary differential equations

$$\begin{aligned}
\frac{dx_1}{dt}(t) &= \sigma[x_2(t) - x_1(t)] \\
\frac{dx_2}{dt}(t) &= -x_1(t)x_3(t) + rx_1(t) - x_2(t) \\
\frac{dx_3}{dt}(t) &= x_1(t)x_2(t) - bx_3(t)
\end{aligned} \tag{38}$$

and we take $\sigma = 16$, $b = 4$, and $r = 45.92$. We integrated the equations using a fourth-order Runge-Kutta method and a time step $\Delta t = 0.02$, which means that the orbit goes about 400 times around the attractor. The dimension of the attractor is 2.06. Again for convenient comparison with our results we show the attractor projected onto the $(x(n), y(n))$ plane in Figure 17. In this case we worked directly with the (x, y, z) data from the Lorenz system rather than create three-vectors from $x(n)$ alone by phase space reconstruction.

To this orbit we added uniform noise in the interval $[-5.0, 5.0]$, which gives a noise level of 23% comparing rms levels of the clean orbit and of the noise, or a signal-to-noise ratio of 6.4 dB. The noisy orbit is shown in Figure 18, and one can see that the attractor is quite thoroughly masked. In Figure 19 we show the result of applying our algorithm 35 times. The recovery of the geometric structure is quite striking. Again we display the local absolute error (Figure 20) and the local deterministic error (Figure 21). The various averaged error measures for this process are $ADE = -0.93; GDE = -0.99; AAE = -1.49; GAE = -0.24; RDE = 93.5\%$; and $RAE = 79.8\%$. Using AAE as a measure of the residual noise level we have a final signal-to-noise level of 25.9 dB for a gain in this ratio of 19.5 dB. This is a substantial gain against noise and clearly manifests itself in the visual images of the figures.

To a signal from the Lorenz attractor we added to $x(n)$ a pulse localized around time step 50 of magnitude 5.2, and then using a phase space reconstruction to find vectors on the attractor in an embedding space of $d = 3$, we cleaned the Lorenz plus pulse data to extract the regular pulse signal. In Figure 22 we show the original pulse as a solid line and the cleaned pulse as the solid line plus stars. The original signal-to-noise level was -3.9 dB. In terms of the error measures we have defined, this cleaning procedure led to $AAE = -1.64; GAE = -2.06$; and $RAE = 78.1\%$. In our usual way we compute a final signal-to-noise ratio of 27.4 dB or a gain of 31.3 dB against the original signal-to-noise ratio.

5. Discussion and Conclusions

This paper has presented a method for separating two signals using a *scaled probabilistic* procedure. We outlined the procedure in Section 2 with motivation for our detailed assumptions coming from our planned application to having one of the signals

being a deterministic system with chaotic, and thus spectrally broadband orbits, and the other signal being iid noise. In Section 3 we discussed the details of implementing the algorithm in the phase or state space of the dynamics, and then in Section 4 we gave several examples with error measures for applying the procedure to two different kinds of settings:

- a chaotic signal immersed in uniformly distributed iid random noise. The job of the cleaning algorithm is to extract the chaotic signal from the noisy environment;
- a pulse signal—taken as typical of a regular signal—immersed in a signal from a chaotic source. The job of the cleaning algorithm is to extract the pulse from within the chaotic background. Equivalently one can view this as removing the pulse (a contaminant) from the good chaotic data. The point of these examples is to indicate that a message (the pulse) can be reliably extracted from the background of a chaotic signal when a good reference orbit of the chaotic signal is available. In some sense we have thus demonstrated a probabilistic matched filter method.

The algorithm works the same in each of these settings. We separate the part of the observations $\mathbf{O}(k)$ which "best" satisfies some (unknown) dynamics $\mathbf{S}(k+1) = \mathbf{F}(\mathbf{S}(k))$. "Best" is determined by our reference orbit and the stated probability rules. In the first case the signal $\mathbf{S}(k)$ is our goal. In the second case we subtract $\mathbf{S}(k)$ from $\mathbf{O}(k)$ to reconstruct the pulse $\mathbf{O}(k) - \mathbf{S}(k)$.

In each instance we were able to demonstrate quite cleanly that the separation of two sources can be accomplished with our scaled probabilistic algorithm even when the signal-to-noise ratio is quite small. In the case of the pulse or spike embedded in a chaotic signal we dealt with original signal-to-noise ratios of -4 dB or less, and were able to recover the input pulse or spike with a final signal-to-noise ratio which was quite large. In some quite loose sense this is a gain against "noise"—namely, the unwanted source—of about 25 to 30 dB (in amplitude). In each instance where we explored the working of the algorithm, we were able to deal with even lower original signal-to-noise ratios and successfully extract the desired signal from the observations. It was quite critical to have a long reference orbit to be able to achieve this high gain against noise.

The methods presented here are cruder that those which use details of the dynamics [1, 2, 3] in the sense that they have no hope of achieving machine accuracy, and we cannot guarantee they converge in an exponential fashion. This leads us to think of the present scaled probabilistic methods as the first pass one would take on noisy data. One would follow that first pass with the highly convergent dynamical methods. In our work here we require a rather good reference orbit to achieve the large gains in signal-to-noise ratio we have observed.

Our use of the reference orbit is reminiscent of matched filtering in familiar linear theory. We use the density in phase space provided by the reference orbit as a template against which to test our local meanders in the vicinity of the observations $\mathbf{O}(k)$ as we seek the maximum likelihood for our probabilities. The matched filtering in linear theory is formulated in Fourier domain. Of course, in nonlinear problems Fourier space is generally not an improvement, so our matched filters have been presented in time-domain, and in the phase space of the system, to be precise. A most important

property of the filter is that it represents the invariant density [5, 6], so our observation of it by *any* orbit of the system tells us all we need to know. The details of the orbit itself are not important. If they were, we would be stopped by the usual positive Lyapunov exponents or sensitivity to initial conditions, and this method would be of little use.

By no means have we explored all possibilities for implementing the general scheme we outlined in Section 2. Section 3 of this paper represents our best numerical efforts to date, but we can imagine various kinds of improvements, including scaling strategies that are different in different parts of phase space, and generalizations of the basic one-step Markov assumption on the distributions of the clean state space vectors. The latter, in particular, may give degrees of freedom which allow greater gains in signal strength in the presence of non-white noise.

Other issues not explored by us but well worth investigation based on the pleasant success of the method as reported here include (1) exploration of the dependence of the method on the length and quality of the reference orbit (for too short a reference orbit, we surely must be defeated by relatively low noise levels); (2) exploration of the dependence of the method on the dimension and time lag used for embedding the scalar observed data into a multivariate phase space; (3) exploration of the dependence of the method on the number of subballs which we use to divide the noise neighborhood σ_η at each iteration of the cleaning procedure; (4) exploration of the sensitivity to the number of passes through the algorithm and scale factor used at each pass; and (5) exploration of adaptive scaling methods which scale down the noise ball size in different regions of phase space in different fashions.

Even at the stage of development we have presented in this paper, we are able to envision applying these methods to quite noisy laboratory and field data where we suspect there is a chaotic signal buried in environmental or instrumental noise. We shall do that with fluid dynamics data and report on this in the future. One can also envision the transmission of a regular "message" covertly located in a chaotic environment which can be retrieved by the methods we have used. We shall explore some of the limits of this as well.

Acknowledgments

We are most appreciative for productive conversations with J. Rissanen about the material covered in this note. J. J. ("Sid") Sidorowich read the drafts of this article and made numerous suggestions on its improvement. This work was supported in part under the DARPA—University Research Initiative, URI Contract No. N00014-86-K-0758 and in part by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Engineering and Geosciences, under Contract N000014-89-D-0142 DO#15.

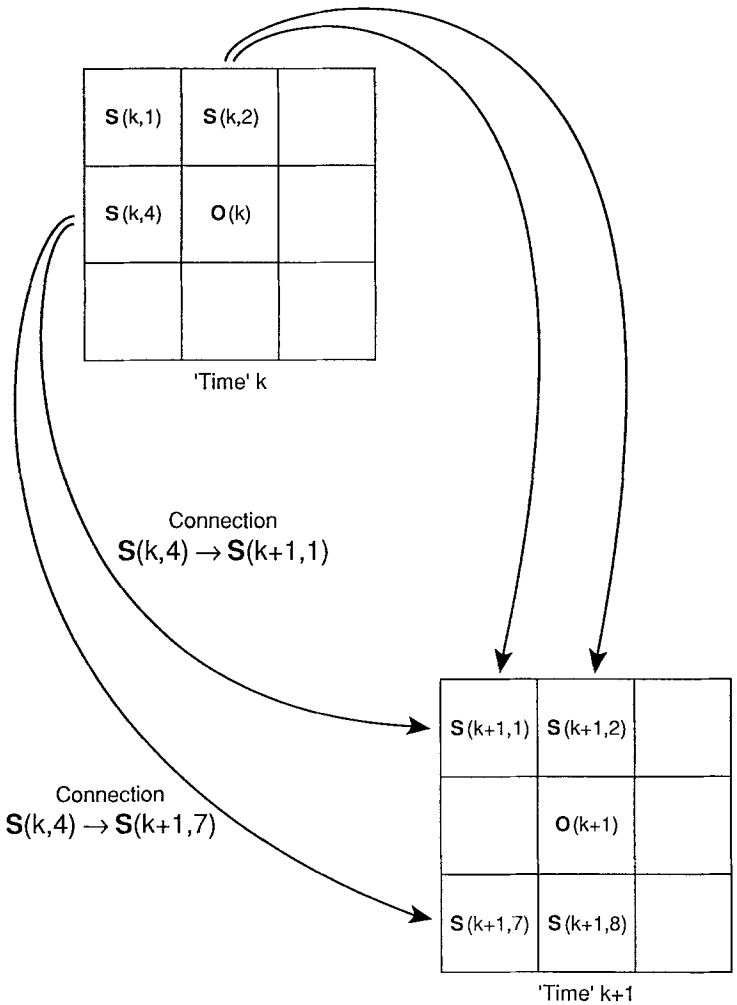

Fig. 0. An illustration of the M^2 connections to be evaluated in determining the $PF(\mathbf{S}(k+1, i))$ for $i = 1, 2, \ldots, M$ from the earlier values $PF(\mathbf{S}(k, i))$ for $i' = 1, 2, \ldots, M$. In this figure we show the local phase space for $M = 9$, $d = 2$. Each of the 9 squares around the observed points $\mathbf{O}(k)$ and $\mathbf{O}(k+1)$ is a possible location for a 'cleaned' data point. We must evaluate the 81 connection probabilities from "time" k to "time" $k+1$. The probabilities $PF(\mathbf{S}(k, i))$ are known, and we maximize at each $\mathbf{S}(k+1, i')$ over the values at $\mathbf{S}(k, i)$ to determine the values of $PF(\mathbf{S}(k+1, i'))$. Equation (12) with $m = k+1$ is used to determine the $PF(\mathbf{S}(k+1, i'))$. A similar picture holds for the backward recursion to evaluate $PB(\mathbf{S}(k, i))$.

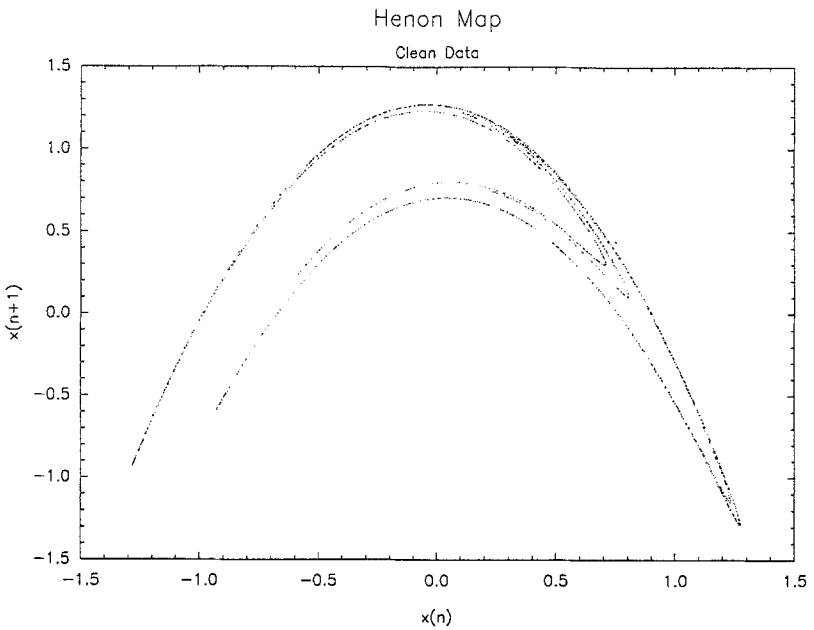

Fig. 1. The phase portrait for the Hénon map constructed from data on $x(n)$ only. Two-vectors $[x(n), x(n+1)]$ are plotted. These data have no noise except machine error. This figure is found in numerous places, and is reproduced here for ease of comparison with the noisy and cleaned Hénon data to follow.

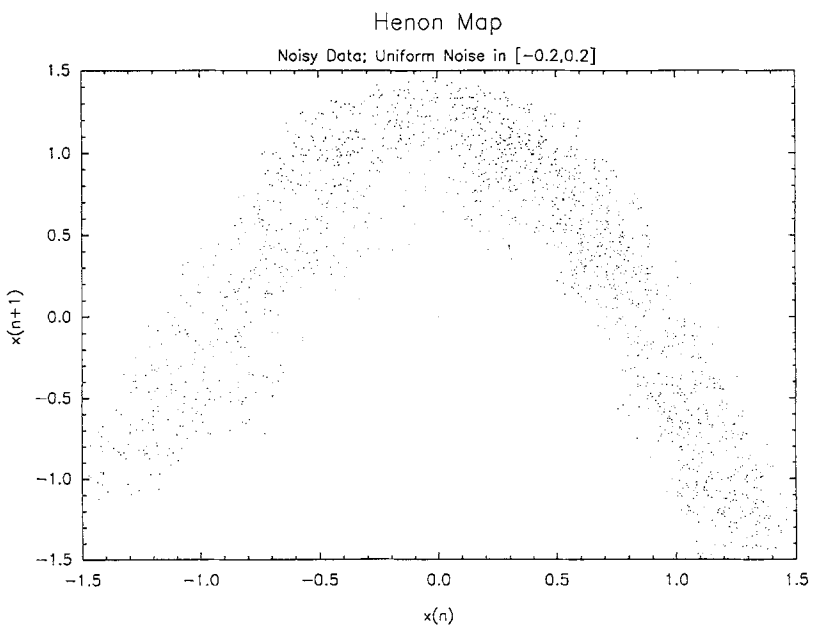

Fig. 2. Phase portrait for the Hénon map with noise uniformly distributed in $[-0.2, 0.2]$ added to the data.

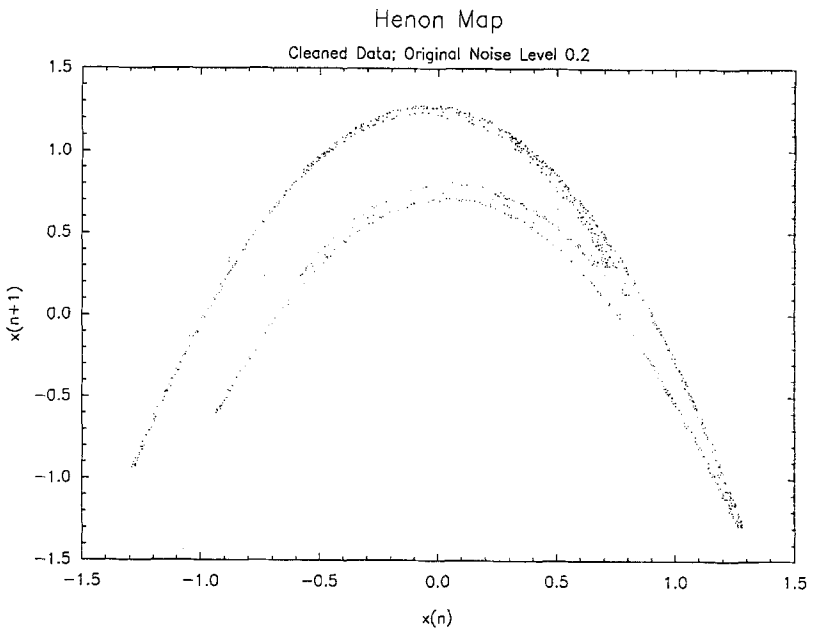

Fig. 3. Phase portrait of the Hénon map after 35 passes though the scaled probabilistic cleaning algorithm starting with a noise level of 0.2, as in Figure 2. 1000 points on the orbit have been cleaned.

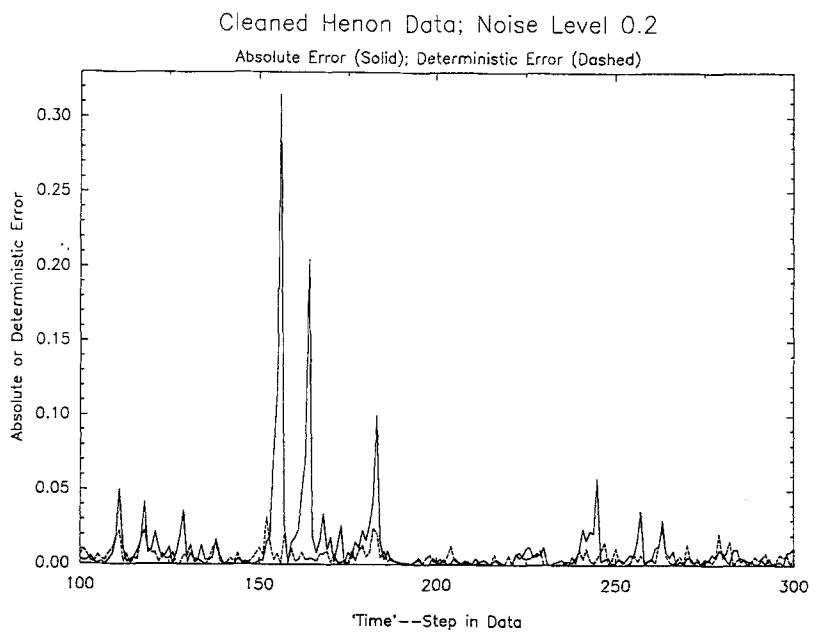

Fig. 4. The absolute error ($AE(k)$) and the deterministic error ($DE(k)$) in the cleaned data after the application of 35 passes of the scaled probabilistic cleaning procedure on Hénon map data. It is important to note that the deterministic error [how well the dynamics is satisfied] can remain small even when the absolute error [how close the cleaned data is to the original data] makes large excursions.

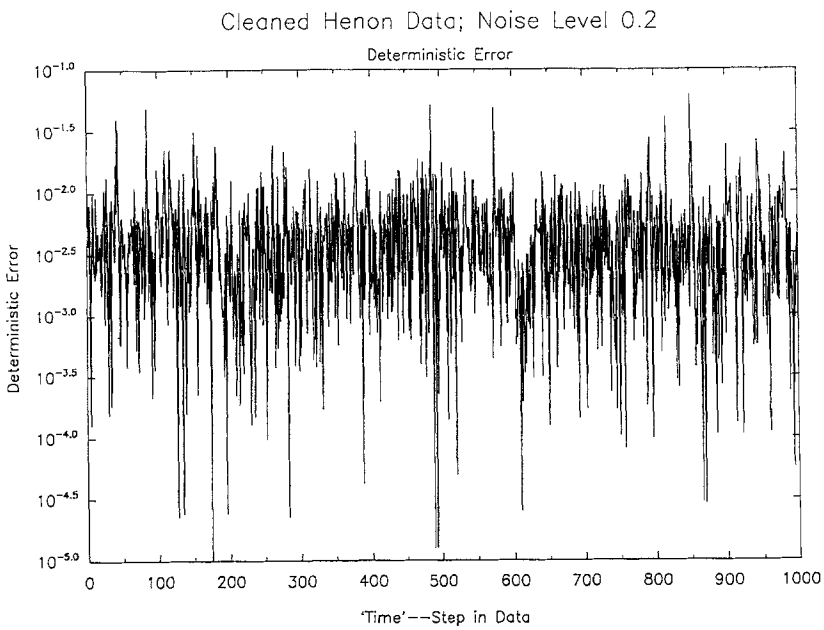

Fig. 5. The absolute error, $AE(k)$, for the cleaned Hénon data for $2 \leq k \leq 999$ for the cleaned orbit shown in Figure 3.

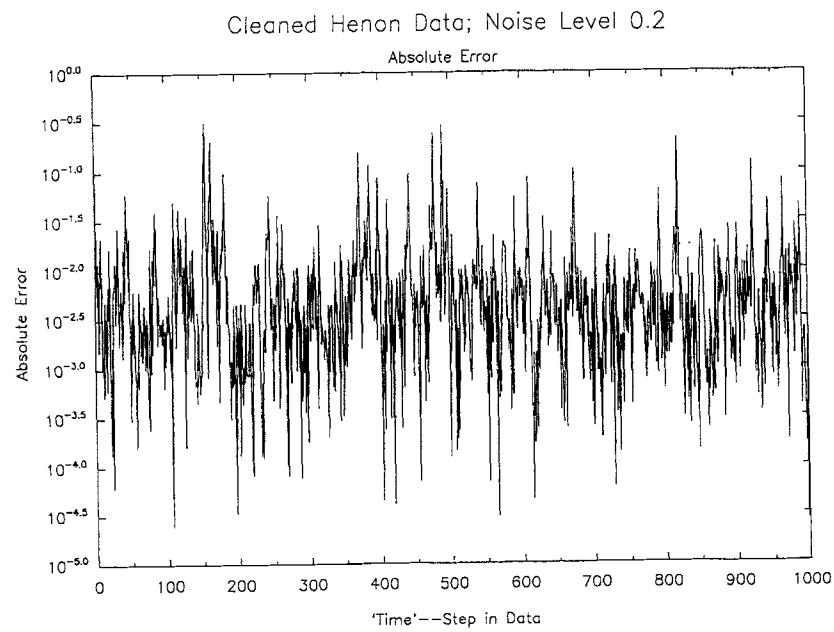

Fig. 6. The deterministic error, $DE(k)$, for the cleaned Hénon data for $2 \leq k \leq 999$ for the cleaned orbit shown in Figure 3.

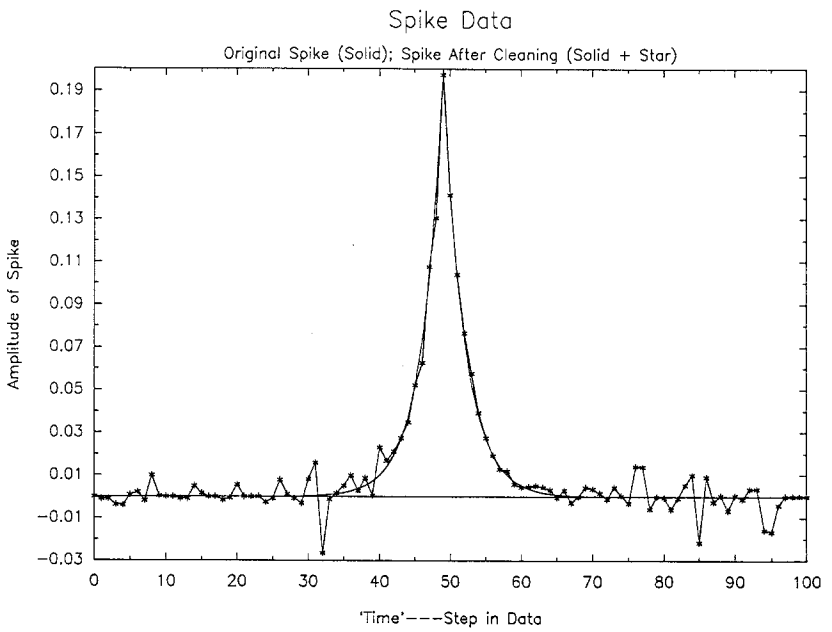

Fig. 7. The original (clean) spike signal—solid line—which is added to a signal from the Hénon map shown with the reconstructed signal—solid line with stars—from the scaled probabilistic noise reduction procedure.

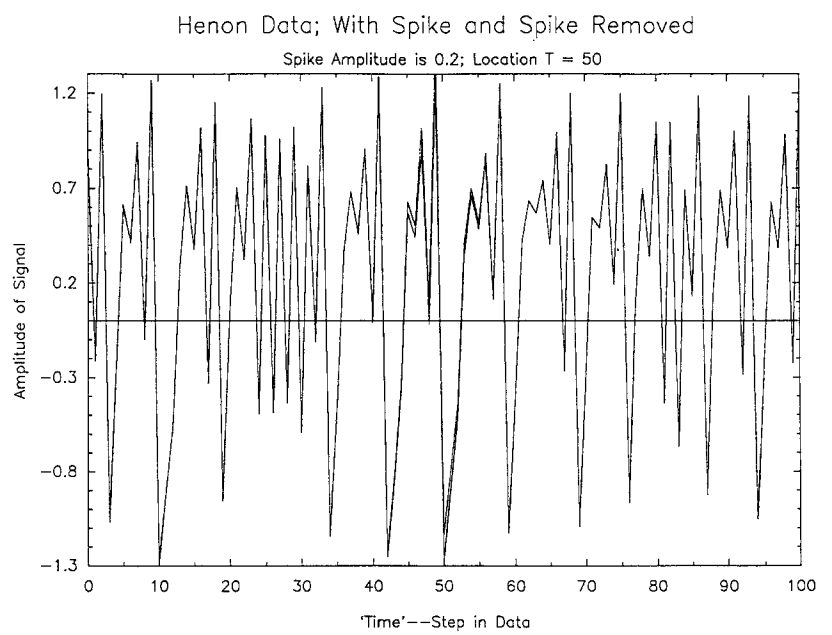

Fig. 8. The "noisy signal" which is the spike of amplitude 0.2 added to the signal from the Hénon map shown simultaneously with the Hénon signal alone. The time series differ only slightly in the neighborhood of time step 40 where the spike is localized.

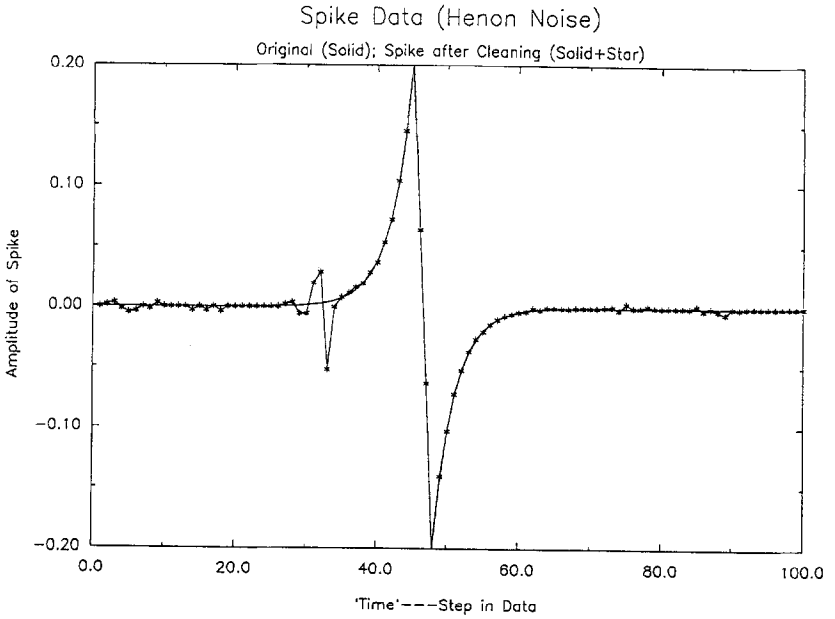

Fig. 9. The original pulse signal of amplitude 0.2 localized around time step 50—solid line—and the reconstructed signal—solid line with stars—from the scaled probabilistic noise reduction procedure. The noisy signal was an Hénon map orbit plus pulse signal. In this cleaning procedure 20,000 points for the reference orbit were used.

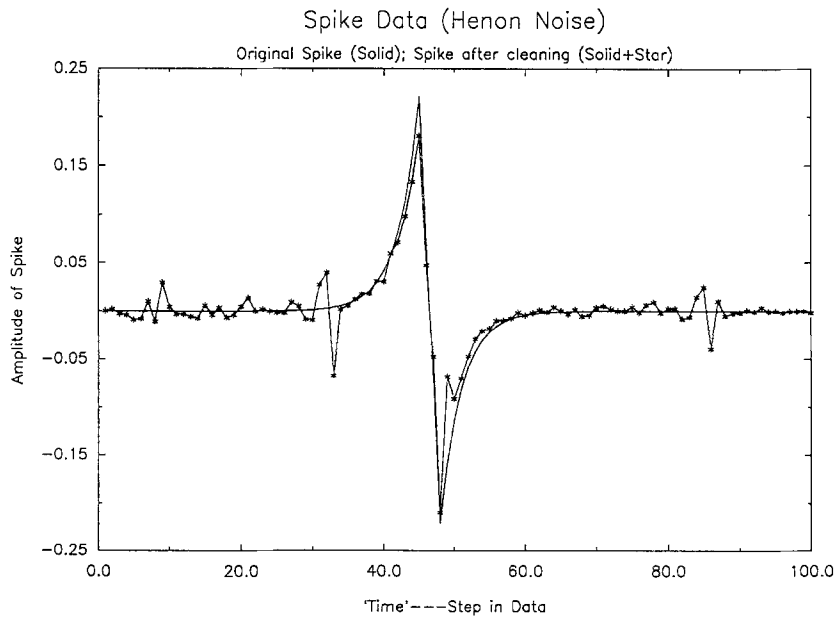

Fig. 10. The original pulse signal of amplitude 0.2 localized around time step 50—solid line—and the reconstructed signal—solid line with stars—from the scaled probabilistic noise reduction procedure. The noisy signal was an Hénon map orbit plus pulse signal. In this cleaning procedure 10,000 points for the reference orbit were used.

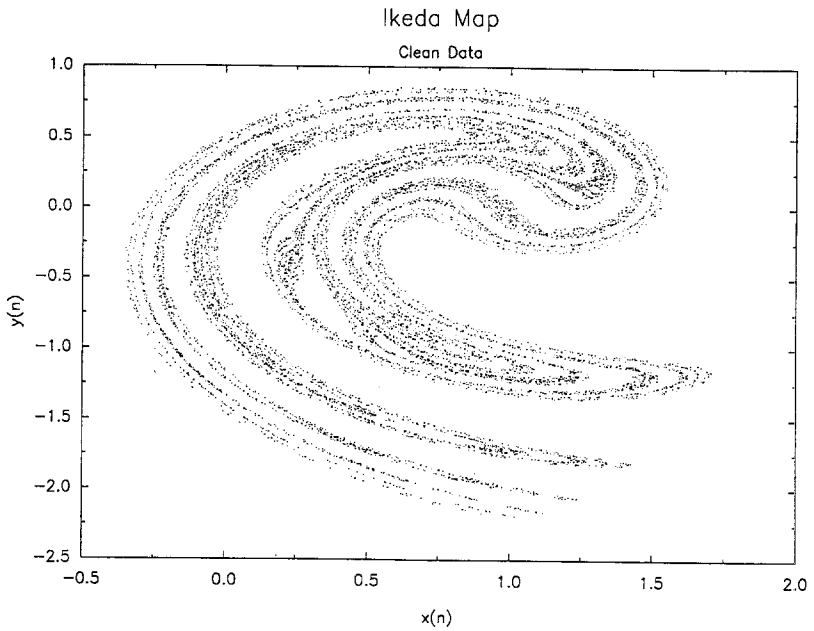

Fig. 11. The phase portrait for the Ikeda map. Two-vectors which are the real, $x(n)$, and imaginary part, $y(n)$, of the complex Ikeda amplitude are plotted. This data has no noise except for machine error. This figure is found in numerous places, and is reproduced here for ease of comparison with the noisy and cleaned Ikeda data to follow.

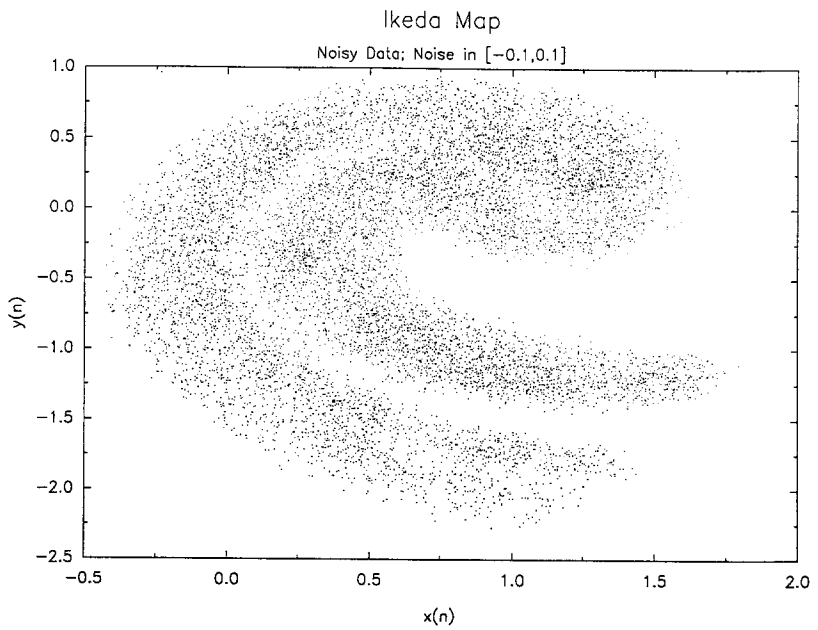

Fig. 12. Phase portrait for the Ikeda map with noise uniformly distributed in $[-0.1, 0.1]$ added to the data.

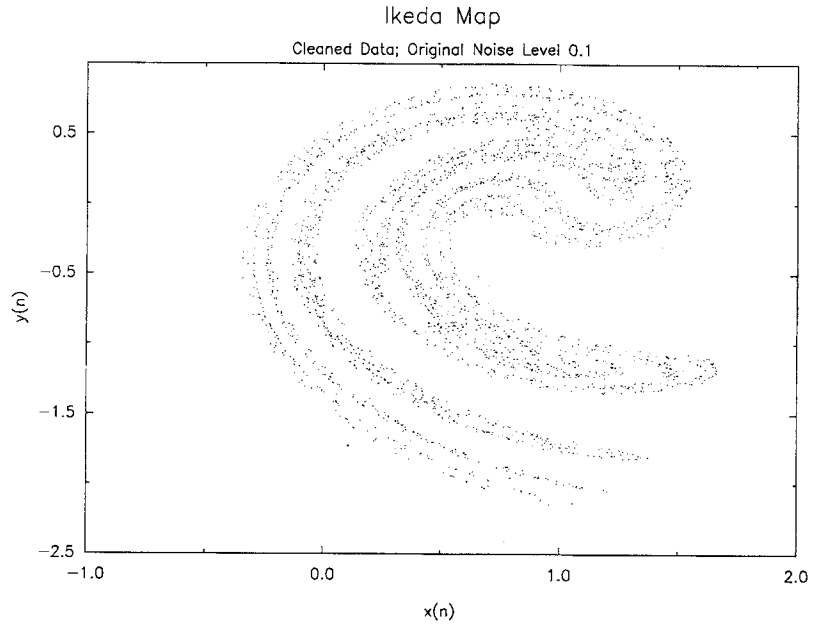

Fig. 13. Phase portrait of the Ikeda map after 35 passes though the scaled probabilistic cleaning algorithm starting with a noise level of 0.1, as in Figure 12. 1000 points on the orbit have been cleaned.

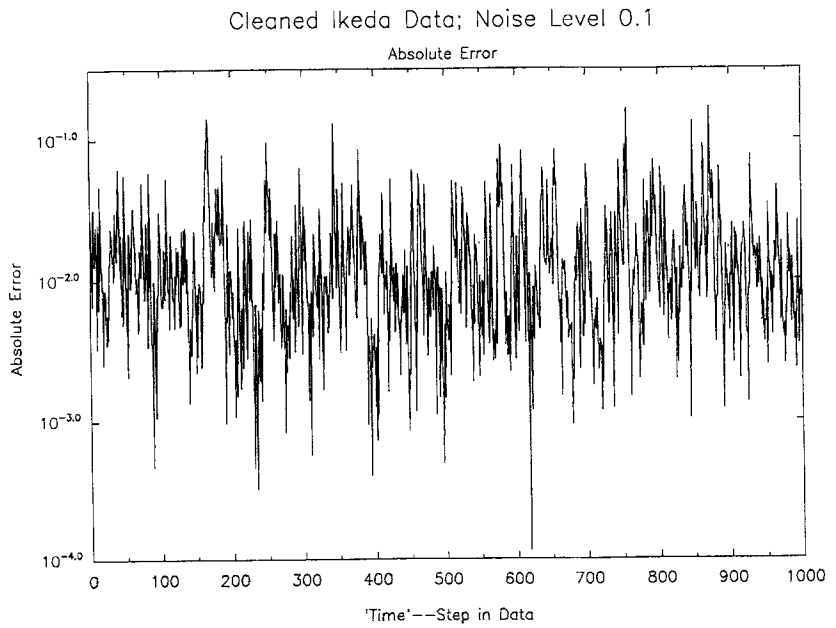

Fig. 14. The absolute error, $AE(k)$, for the cleaned Ikeda data for $2 \leq k \leq 999$ for the cleaned orbit shown in Figure 13.

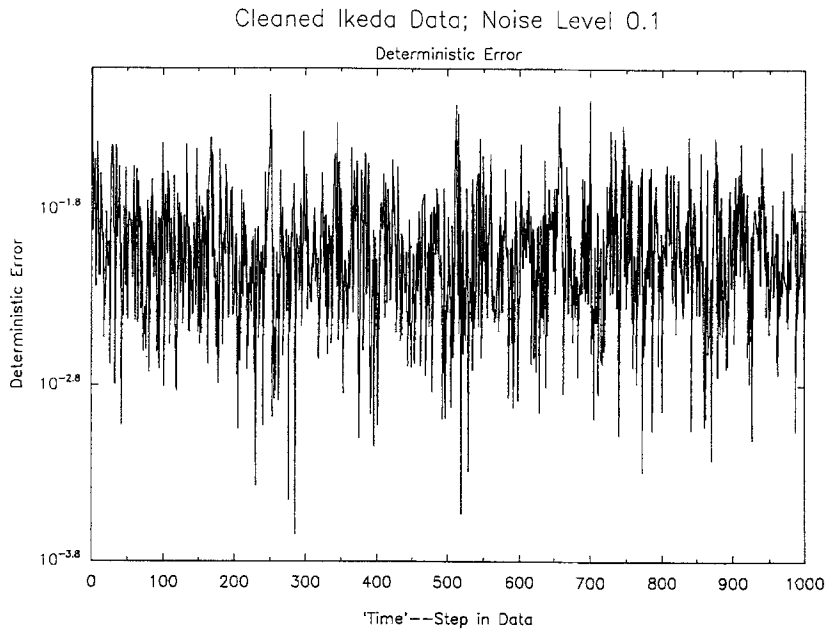

Fig. 15. The deterministic error, $DE(k)$, for the cleaned Ikeda data for $2 \leq k \leq 999$ for the cleaned orbit shown in Figure 13.

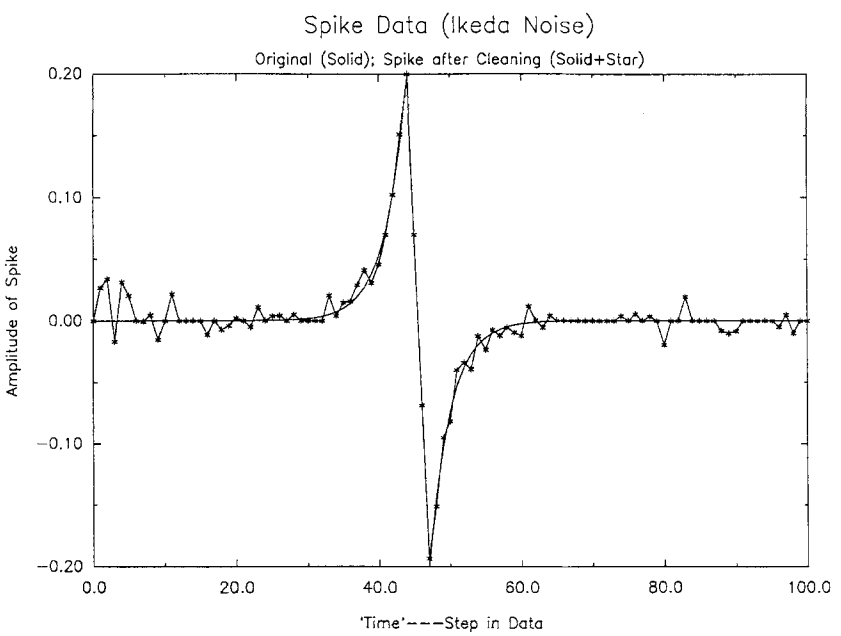

Fig. 16. The original pulse signal of amplitude 0.2 localized around time step 50—solid line—and the reconstructed signal—solid line with stars—from the scaled probabilistic noise reduction procedure. The noisy signal was an Ikeda map orbit plus pulse signal. In this cleaning procedure 20,000 points for the reference orbit were used.

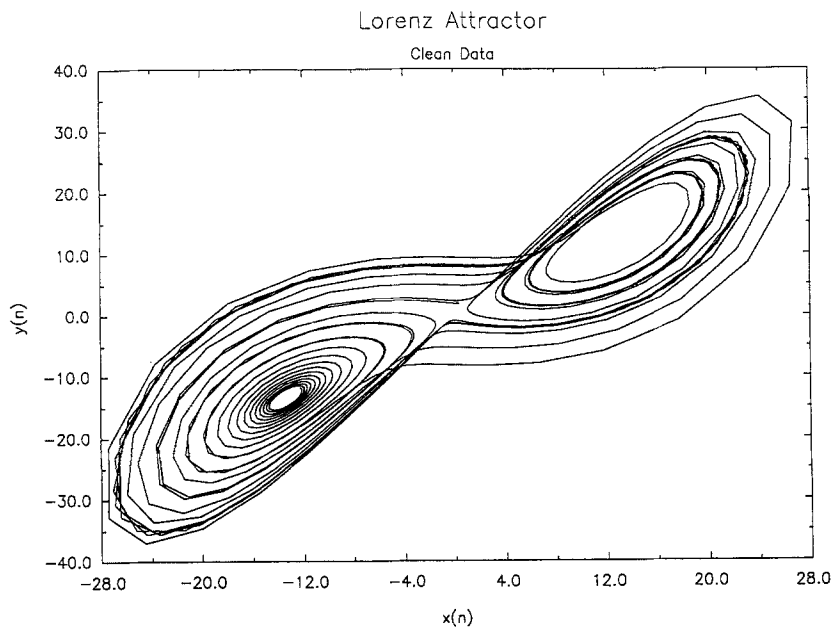

Fig. 17. The projected phase portrait of the Lorenz attractor. The projection is on the $[x(n), y(n)]$ plane. This figure is found in numerous places, and is reproduced here for ease of comparison with the noisy and cleaned Lorenz data to follow. Only 1000 points on the attractor are shown for comparison with what follows.

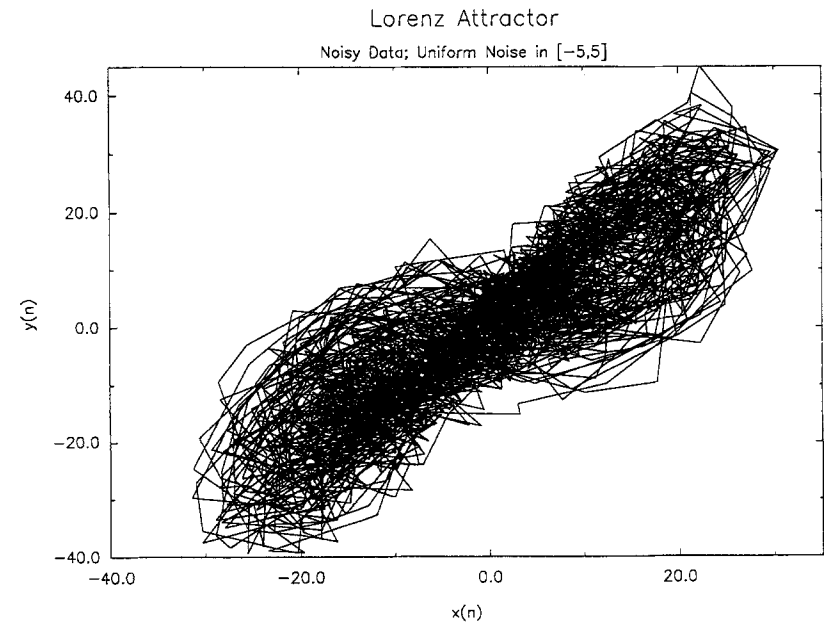

Fig. 18. The projected phase portrait of the Lorenz attractor with noise uniformly distributed in $[-5.0, 5.0]$ added to each component $[x(n), y(n), z(n)]$. The projection is on the $[x(n), y(n)]$ plane. Clearly with this noise level the Lorenz attractor has become quite obscured.

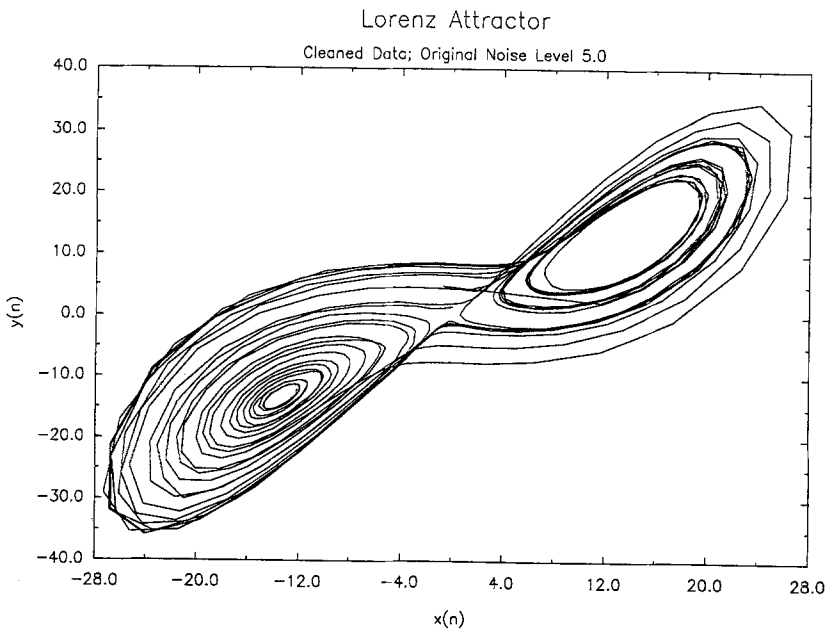

Fig. 19. The projected phase portrait for the data cleaned by our scaled probabilistic procedure starting from the data shown in Figure 18 for the Lorenz attractor. The projection is on the $[x(n), y(n)]$ plane.

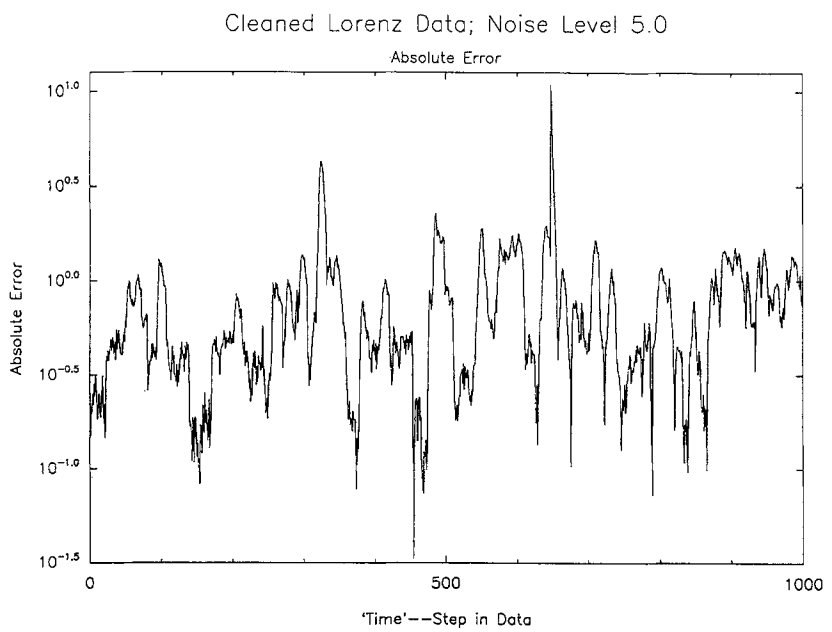

Fig. 20. The absolute error, $AE(k)$, for the cleaned Lorenz data for $2 \leq k \leq 999$ for the cleaned orbit shown in Figure 19.

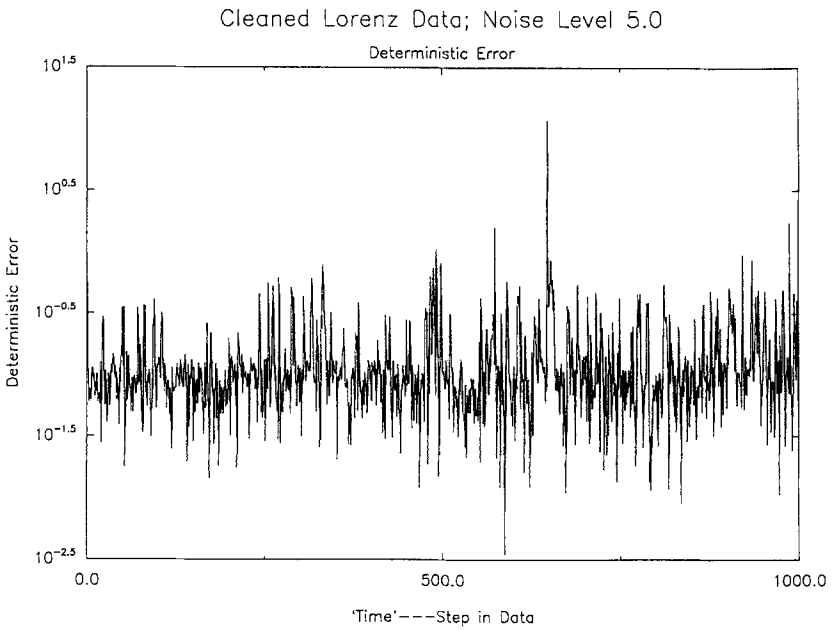

Fig. 21. The deterministic error, $DE(k)$, for the cleaned Lorenz data for $2 \leq k \leq 999$ for the cleaned orbit shown in Figure 19.

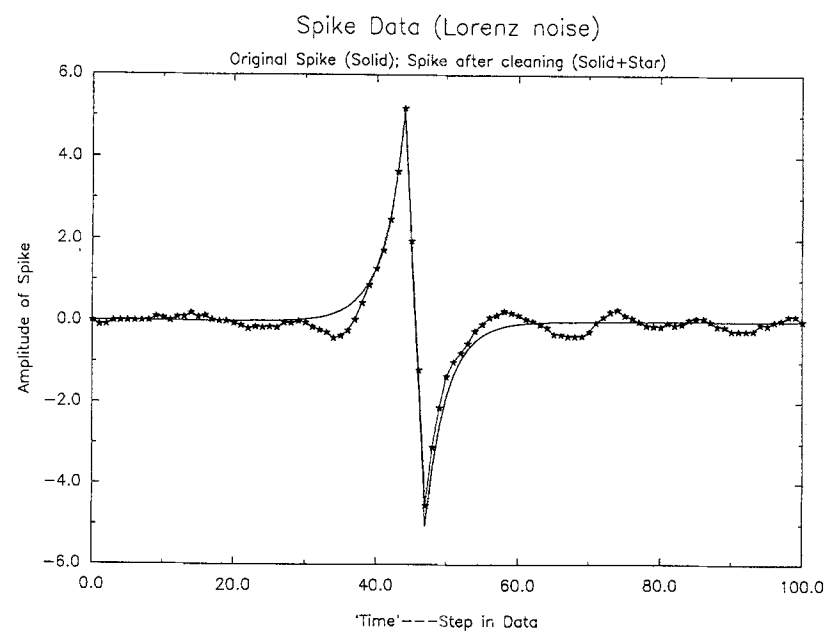

Fig. 22. The original pulse signal of amplitude 5.2 localized around time step 50—solid line—and the reconstructed signal—solid line with stars—from the scaled probabilistic noise reduction procedure. The noisy signal was a Lorenz attractor orbit plus pulse signal. In this cleaning procedure 10,000 points for the reference orbit were used.

References

1. Hammel, S. M., "A Noise Reduction Method for Chaotic Systems," *Phys. Lett. A* **148**, 421 (1990).
2. Farmer, J. D. and J. J. ("Sid") Sidorowich, "Exploiting Chaos to Predict the Future and Reduce Noise," in *Evolution, Learning, and Cognition*, ed. Y. C. Lee (World Scientific, 1988); "Optimal Shadowing and Noise Reduction," to appear in *Physica D*, 1990. The second paper has many similarities to the work of Hammel [1] and refines that work as well as treating some of the issues associated with "homoclinic tangencies" in a useful fashion. A review of work in noise reduction through Spring, 1990 can also be found in this second paper.
3. Abarbanel, H. D. I., P. F. Marteau, and J. J. ("Sid") Sidorowich, "Recovering Deterministic Signals from Noisy Data by Scaled Linear Cleaning," UCSD/INLS Preprint, October, 1990. To be submitted to *Phys. Rev. A*.
4. Kostelich, E. J. and J. A. Yorke, "Noise Reduction in Dynamical Systems," *Phys. Rev.* **A38**, 1649 (1988).
5. Eckmann, J. -P. and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.* **57**, 617–656 (1985).
6. Abarbanel, H. D. I., R. Brown, and J. Kadtke, "Prediction and System Identification in Chaotic Nonlinear Systems: Time Series with Broadband Spectra," *Phys. Lett.***138A**, 401 (1989). "Prediction in Chaotic Nonlinear Systems: Methods for Time Series with Broadband Fourier Spectra," *Phys. Rev.* **A41**, 1782 (1990).
7. Brown, R., P. Bryant, and H. D. I. Abarbanel, "Computing the Lyapunov Spectrum of a Dynamical System from Observed Time Series," *Phys. Rev. A* **43**, 2787 (1991); Bryant, P., R. Brown, and H. D. I. Abarbanel, "Lyapunov Exponents from Observed Time Series," *Phys. Rev. Lett.* **65**, 1523 (1990).
8. Bellman, R. and S. E. Dreyfus, *Applied Dynamic Programming*, Princeton University Press, 1962.
9. Pesin, Ya. B., "Lyapunov Characteristic Exponents and Smooth Ergodic Theory," *Usp. Mat. Nauk.* **32**, No. 4 (196), 55 (1977); *Russian Math. Survey* **32**, No. 4, 55 (1977).
10. Ruelle, D., "An Inequality for the Entropy of Differentiable Maps," *Bol. Soc. Bras. Mat.* **9**, 83 (1978).
11. Silverman, R., *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).
12. Takens, F., "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, eds. D. Rand and L. S. Young, *Lecture Notes in Mathematics* **898**, 366 (Springer, Berlin, 1981).
13. Mañé, R., "On the Dimension of the Compact Invariant Sets of Certain Nonlinear Maps," in *Dynamical Systems and Turbulence, Warwick 1980*, eds. D. Rand and L. S. Young, *Lecture Notes in Mathematics* **898**, 230 (Springer, Berlin, 1981).
14. Friedman, J. H., J. L. Bentley, and R. A. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," *ACM Trans. Math. Software*, **3**, 209–226, (1977).
15. Hénon, M., "A Two Dimensional Mapping with a Strange Attractor," *Commun. Math. Phys.* **50**,69 (1976).
16. Ikeda, K., "Multiple-Valued Stationary State and Its Instability of the Transmitted Light by a Ring Cavity System," *Opt. Commun.* **30**, 257 (1979).
17. Lorenz, E. N., "Deterministic, Nonperiodic Flow," *J. Atmos. Sci.* **20**, 130 (1963).