Physica D 34 (1989) 391–404
North-Holland, Amsterdam

RECONSTRUCTING ATTRACTORS FROM SCALAR TIME SERIES: A COMPARISON OF SINGULAR SYSTEM AND REDUNDANCY CRITERIA

Andrew M. FRASER
Department of Physics, The University of Texas at Austin, Austin, TX 78712-1081, USA

Received 2 June 1988
Revised manuscript received 30 September 1988
Communicated by R.M. Westervelt

A delay-vector phase space reconstruction in which the delay time satisfies a minimum redundancy criterion is compared with a reconstruction obtained using a singular system approach. Minimum redundancy produces the better reconstruction. The reconstructions are compared using a *distortion functional* $\mathscr{D}$ which measures how well the location of a point in the original phase space can be determined on the basis of its image under the reconstruction process. The superiority of the redundancy analysis over the singular system analysis is found to arise from the former's foundation on the notion of general independence as opposed to the latter's foundation on the notion of linear independence.

1. Introduction

The observation that the dynamics of a system with many degrees of freedom can be investigated using time series of a single scalar observable has broadened the class of experiments in which complex behavior can be interpreted as manifestations of strange attractors. Packard et al. [1] suggested two schemes for reconstructing vector dynamics from scalar time series. Takens [2] suggested the same techniques and proved that reconstructions are generically diffeomorphic (one to one differentiable with a one to one differentiable inverse) to the original dynamics.

In the ensuing flood of applications of these ideas, experimenters observed that some reconstructions are *better* than others. Usually a *bad* reconstruction is one that is not invertible, i.e., points in the reconstructed phase space do not uniquely identify points in the original phase space. In experimental situations, noise makes all reconstructions non-invertible and whether a reconstruction is *good* or *bad* is a question of degree. It is harder to estimate diffeomorphic invariants using worse reconstructions. For instance Haucke et al. [3] studied the dependence of their estimates of fractal dimension (a diffeomorphic invariant) on the parameters of their reconstructions. One is led to ask, "Which reconstruction is 'most diffeomorphic' to the original phase space?"

Thus there is a need for techniques of obtaining good reconstructions that do not depend on trial and error or the taste of the operator. Takens' paper is mathematically authoritative, but it does not touch on many of the issues that must be considered by an experimenter in applying the techniques. From Takens' mathematical point of view a reconstruction is diffeomorphic to the original phase space or it is not, and an idea such as "more diffeomorphic" is meaningless. An experimenter is assured that he can extract diffeomorphic invariants (dimensions, Lyapunov exponents, etc.) of an attractor from an infinite amount of noise-free data, but he is given little help in selecting a reconstruction technique that will be robust to the limitations inherent in "real" data. In the paper by Packard et al. most of the issues involved in obtaining good reconstructions are discussed.

Their discussion is not rigorous; it focuses on describing problems rather than solutions.

Following a suggestion of Shaw's, Fraser and Swinney [4] implemented an information theoretic procedure for finding good reconstructions. Roughly the procedure requires that the *mutual information* between the components of a two-dimensional reconstruction should be as small as possible. In [5] we expanded the two-dimensional approach to higher dimensions by minimizing *redundancy*, which is a generalization of mutual information. Broomhead and King [6, 7] have introduced an alternative procedure based on singular value decompositions, which they call a singular systems approach.

In the present paper we try to define more precisely what *good* reconstructions are, and how one might determine whether a reconstruction is *good* without having an original phase space trajectory available. We apply the two suggested procedures to a quasiperiodic time series, and we evaluate the reconstructions by measuring how closely points in the original phase space can be approximated by a piecewise linear map applied to points in the reconstructed phase space. We find the reconstruction produced by the singular system approach to be inferior. We then use the example to illustrate the weaknesses of the singular system approach.

2. History

Takens [2] describes a technique for determining whether experimental behavior (such as in the transition to turbulence in Taylor–Couette flow) can be attributed to the presence of a strange attractor. The two characteristics that he uses to identify a strange attractor are small noninteger dimension and positive entropy. The paper explains how a scalar time series of measurements can be used to measure attractor dimension and entropy. As a first step in developing the technique, Takens shows that if the dynamics are governed by a finite-dimensional attractor, then it is a generic property that a time series of a scalar observable can be used to construct a vector time series that is diffeomorphic to the original dynamics.

2.1. *Experimenters' experience*

Since the publication of the papers by Packard et al. [1] and Takens [2], hundreds of experimenters have used delay coordinates to reconstruct phase space trajectories from scalar time series. While some experimenters have found that their reconstructions are robust with respect to the choice of coordinates (time delay T) [8], the more common experience is that the values of "invariants" (dimension, entropy, Lyaponov exponents, etc.) seem to depend on the reconstruction chosen. In the proceedings of one conference [3], five different papers stress the importance of choosing the correct parameters for a reconstruction.

In section 4 we suggest that the appropriate way to quantify the quality of a reconstruction is in terms of a distortion functional $\mathscr{D}(S, V)$ which is evaluated over the "original" phase space S and the reconstructed phase space V. One would like to choose a distortion measure in such a way that invariants are easy to estimate if a reconstruction has low distortion. The observation that experimenters can tell whether reconstructions are distorted even when the original phase spaces are not available, suggests that it might be possible to define a distortion measure that does not involve the original phase space S. We examine this question more closely in section 4.

2.2. *The informal analysis of Packard et al.*

Packard et al. discussed how one obtains good reconstructions from experimental data in terms of the probability densities that are assumed to arise from ergodic natural measures of attractors. Such probability densities can be estimated from experimental data. In order to discuss these densities we will use Gallager's notation [9]. An upper case letter Z indicates an *ensemble* which consists of a set of possible values (also indicated by the same upper-case letter) and an associated probability function p_Z. The corresponding lower-case letter z indicates an element of the set Z. To indicate a vector variable or a set of vectors we use boldface letters, e.g. $\boldsymbol{z}$ or $\boldsymbol{Z}$.

Using this notation, we can describe dynamical systems and their measurements as we conceive of them in this paper. We denote the phase space S with elements s. We suppose that the dynamics are given by a map $\phi: S \mapsto S$ with $s(t+1) = \phi(s(t))$ and that there is a scalar observable function $\psi: S \mapsto \mathbb{R}$. Before the values of the observable can be recorded, they are contaminated by measurement noise. Thus we model the measurements X by

$$x(t) = \psi(s(t)) + \eta(t),$$

where η is an independent noise term.

The simplest way to reconstruct phase space from a scalar time series is by using delay coordinates, i.e., $\boldsymbol{v}(t) = \boldsymbol{x}_m^T(t)$, where we have introduced the notational convenience

$$\boldsymbol{x}_m^T(t) \equiv (x(t), x(t+T), \ldots, x(t+(m-1)T)).$$

We can denote the composition of the observable function, measurement process, and method of delays by $\boldsymbol{\Phi}: S \mapsto V$, where

$$\boldsymbol{v}(t) = (v_0(t), v_1(t), \ldots, v_{m-1}(t))$$

and

$$v_i(t) = \psi(s(t+iT)) + \eta(t+iT) = x(t+iT).$$

$\boldsymbol{\Phi}$ is called the reconstruction function. We could use a more general reconstruction, e.g.,

$$v_i(t) = \sum a_{i,j} x(t+j). \qquad (1)$$

Here the coordinate functions are linear. Packard et al. suggested trying coordinate functions which are discrete approximations to derivatives, e.g.,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ -0.5 & 0 & 0.5 & \cdots \\ 1 & -2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

One can imagine generalizing even farther and considering nonlinear coordinate functions, but for now we consider only delay coordinates, i.e., $a_{i,j} = \delta_{iT,j}$.

The idea of a good reconstruction is that the conditional probability density $p_{S|V}$ is "extremely sharp". Since phase space is not directly accessible in an experiment, Packard et al. say that the conditional density which should be sharply peaked is $p_{X(t+mT)|\boldsymbol{X}_m^T(t)}$, and that the components of $\boldsymbol{X}_m^T$ should be "independent in an operational sense" to obtain such sharpness. They also say that the value of m required to get sharp conditional densities is the topological dimension of the attractor. In other words if m is the topological dimension of the attractor and a reconstruction is done in $m+1$ dimensions, then for a typical point on the attractor the first m coordinates are sufficient to predict the last coordinate. They caution that T should be small with respect to A/h_μ where A is the accuracy of the measurements and h_μ is the measure theoretic entropy.

2.3. *Redundancy analysis*

The qualitative analysis of the previous subsection is instructive, but a quantitative procedure would be more helpful for experimenters. In response to this need Fraser and Swinney [4] developed a procedure for choosing the best delay to use for a two-dimensional reconstruction. The criterion which was suggested by Shaw [10] is that the delay should be set at the first local minimum of *mutual information* between the coordinates.

The mutual information between two random variables Y and Z is a functional defined on their joint probability density [11],

$$I(Y;Z) = \left\langle \log\left[\frac{p_{Y,Z}}{p_Y p_Z}\right]\right\rangle,$$

and is always positive. Note that $I(Y;Z)$ is a functional of the ensemble Y, Z, not a function that has a value at a particular location (y,z). If the base of the logarithms is 2, the information is measured in bits. For a joint density $p_{Y,Z}$ the number of bits of information about Y that a specification of Z yields is given by $I(Y;Z)$. Thus when variables are more independent, their mutual information is smaller. As the symmetry in the definition indicates, $I(Y;Z) = I(Z;Y)$. Notice that if the joint density factors ($p_{Y,Z} = p_Y p_Z$), then $I(Y;Z) = 0$, which is the smallest value possible, and if z determines y exactly, i.e., $y = f(z)$ then $I(Y;Z) = \infty$.

Fraser and Swinney state that the best value of delay T is the smallest T for which $I(X(t); X(t+T))$ is a local minimum. Thus $I(X(t); X(t+T))$ must be calculated from joint densities $p_{X(t),X(t+T)}$ for many values of T. Mutual information can detect nonlinear correlations. Thus while plots of I vs. T are reminiscent of simple autocorrelation plots, they are sensitive to dependencies of any kind.

In a second paper [5] we generalized the procedure to reconstructions of arbitrary embedding dimensions and added an analysis that determined the required embedding dimension. We defined the *redundancy* of a multidimensional distribution

$$R_m^T = \left\langle \log\left[\frac{p_{X_m^T(t)}}{p_{X(t)}p_{X(t+T)}\cdots p_{X(t+(m-1)T)}}\right]\right\rangle.$$

We also defined the marginal redundancy

$$R_m'^T = I\left(X(t+mT); X_m^T(t)\right) = R_{m+1}^T - R_m^T$$
$$= \left\langle \log\left[\frac{p_{X(t+mT)|X_m^T(t)}}{p_{X(t+mT)}}\right]\right\rangle.$$

We developed a recursive algorithm for estimating redundancy that produces a much more accurate estimate than could be obtained by naively applying a uniform partition to phase space. The execution time for the algorithm is proportional to $N \log N$, where N is the number of data points.

Consideration of redundancy provides a quantitative foundation for the ideas of "the degree of independence" and "sharply peaked distributions" discussed by Packard et al. [1]. The variables $X(t)$, $X(t+T)$, $X(t+2T)$ are "more independent" when R_3^T is smaller, and the conditional distribution $p_{X(t+mT)|X_m^T}$ is more "sharply peaked"* when $R_m'^T$ is larger.

To find a good reconstruction using these tools one should first choose the embedding dimension m so that the conditional densities $p_{X(t+mT)|X_m^T(t)}$ are "sharply peaked", i.e., m is chosen to make $R_m'^T$ large. Next T must be selected so that on the average a point on the attractor in the reconstructed phase space $v(t)$ provides as much *useful information* about the original phase space point $s(t)$ as possible. In [5] we argued that the best estimate for this useful information is

$$Q(m,T) = (m-1)(A - h_\mu T) - R_m^T. \qquad (2)$$

Here the accuracy A is defined as the average information that an isolated scalar measurement x provides about the system state S, i.e., $A \equiv I(X;S)$. In maximizing $Q(m,T)$ (eq. (2)), the term $((m-1)(A - h_\mu T))$ assures that T is small with respect to A/h_μ and the term R_m^T enforces independence of the coordinates.

2.4. *The singular systems approach*

Broomhead and King [6, 7] have introduced a sophisticated linear analysis of time series which was intended to provide good phase space reconstructions. They call their analysis the singular system approach. It is based on the Karhunen–Loeve theorem [12, vol. 2, p. 144], which is discussed in the framework of information theory in

*There are subtleties in these quantifications of "independent" and "sharp" because they are measure theoretic and strictly construed are devoid of topological properties.

Gallager [9, p. 402]. The basic idea is that the mean-squared distance between points on the reconstructed attractor should be *maximized*.

In order to describe their approach, we introduce the *n-window* $\boldsymbol{w}(t) \equiv \boldsymbol{x}_n^1(t)$. The idea is that n should be larger than the embedding dimension m and that the reconstruction should consist of a linear projection from $\boldsymbol{W}$ to an m-dimensional subspace $\boldsymbol{V}$. The basis of $\boldsymbol{V}$ is chosen to maximize $\langle |\boldsymbol{v}|^2 \rangle$. The first basis vector $\boldsymbol{e}_0$ is chosen to maximize $\langle |\boldsymbol{w} \cdot \boldsymbol{e}_0|^2 \rangle$, and the second is chosen to maximize $\langle |\boldsymbol{w} \cdot \boldsymbol{e}_1|^2 \rangle$ subject to $\boldsymbol{e}_0 \cdot \boldsymbol{e}_1 = 0$. The procedure is continued in the obvious fashion and produces an orthonormal basis such that $\langle (\boldsymbol{e}_i \cdot \boldsymbol{w})(\boldsymbol{e}_j \cdot \boldsymbol{w}) \rangle = \delta_{i,j}\sigma_i^2$, i.e., the random variables $v_i = (\boldsymbol{e}_i \cdot \boldsymbol{w})$ are *linearly* independent. It is important to note that this is not statistical or general independence but only linear independence. Given an experimental data set this analysis can be done using a singular value decomposition routine available in many subroutine libraries [13]. The numbers $\sigma_i = \sqrt{\langle |\boldsymbol{w} \cdot \boldsymbol{e}_i|^2 \rangle}$ are called singular values (where $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \sigma_3 \cdots$).

Broomhead and King explain that if the data are noisy and the variance of the noise is σ_η^2, then every singular value will be augmented by σ_η. Thus noise dominates any v_i whose corresponding singular value σ_i is comparable to σ_η, and such components should be discarded. This process results in a reduction of the noise power in the reconstruction $\boldsymbol{V}$ by a factor m/n where $(n-m)$ is the dimension of the discarded subspace and m is the dimension of the retained subspace. They say that the significant singular values $\sigma_0 \cdots \sigma_{m-1}$ and the corresponding singular vectors $\boldsymbol{e}_0 \cdots \boldsymbol{e}_{m-1}$ represent the deterministic aspects of the time series and that the remaining singular values and singular vectors represent the noise.

They comment that the number of significant singular vectors m is not invariant with respect to diffeomorphisms of the scalar observable ψ. In fact m increases if the $\boldsymbol{n}$-windows are made longer. In order to control the size of m they suggest selecting an $\boldsymbol{n}$-window so that $nt_s = \tau^*$, where τ^* is the time corresponding to the first zero crossing of the second derivative of the autocorrelation function $C(T) = \langle x(t)x(t+T) \rangle$, and t_s is the sample time.

Broomhead and King's approach to phase space reconstruction has two striking advantages over the previous approaches. First, the procedure has a built-in filter for reducing noise. Second, the coordinate functions used to obtain the reconstruction are more general, i.e., all linear projections are considered, not just delta functions $\delta_{iT,j}$. There are also practical advantages in that singular value decomposition subroutines are available in many subroutine libraries and these subroutines are faster and require fewer data than a redundancy analysis. These strengths can be overwhelmed by the weaknesses that we describe below. (The weaknesses of the singular system approach have been examined in previous work by Brandstater et al. [14] and Meese et al. [15].)

3. Correlation and independence

The singular system analysis suggested by Broomhead and King and the redundancy analysis that we suggested are best distinguished by their different underlying notions of independence. For the singular system analysis, the interdependence of random variables is measured by their correlation, and for the redundancy analysis the connection is measured by mutual information. To illustrate these two measures of dependence, suppose that there are two random variables y and z whose joint probability density is described by the Gaussian

$$p(y,z) = \frac{\sqrt{ab-c^2}}{2\pi} \times \exp\left[-\frac{ay^2+bz^2+2cyz}{2}\right].$$

Here

$$\Sigma^{-1} = \begin{pmatrix} a & c \\ c & b \end{pmatrix},$$

where

$$\Sigma = \begin{pmatrix} \sigma_{YY} & \sigma_{YZ} \\ \sigma_{YZ} & \sigma_{ZZ} \end{pmatrix}$$

is the covariance matrix. Given such a distribution, the correlation $C_{Y,Z}$ between Y and Z is

$$C_{Y,Z} = \frac{\sigma_{YZ}}{\sqrt{\sigma_{YY}\sigma_{ZZ}}},$$

and the mutual information is

$$I(Y;Z) = \left\langle \log \frac{p_{Y,Z}}{p_Y p_Z} \right\rangle.$$

In fact if the distribution is Gaussian, mutual information and correlation are related by

$$I(Y;Z) = -\tfrac{1}{2}\log\left[1 - C_{Y,Z}^2\right].$$

While the difference between mutual information and correlation for Gaussian random variables is trivial, the difference becomes significant for non-Gaussian densities. If the probability density $p_{Y,Z}$ is Gaussian then the following three statements are equivalent:

1. $C_{Y,Z} = 0$,
2. $I(Y;Z) = 0$,
3. y and z are *generally independent*, i.e., $p_{Y,Z} = p_Y p_Z$.

If $p_{Y,Z}$ were not Gaussian, then general independence would still be implied by $I(Y;Z) = 0$, but $C_{Y,Z} = 0$ would only imply *linear independence*, i.e., $\langle yz \rangle = \langle y \rangle \langle z \rangle$. Since the probability densities generated by dynamical systems are not Gaussian, correlation functions can only indicate linear independence. Unfortunately, linear independence is not a very important characteristic.

Broomhead and King [6] point out that the autocorrelation function

$$C(T) = \frac{\langle x(t)x(t+T)\rangle - \langle x(t)\rangle^2}{\langle x^2(t)\rangle - \langle x(t)\rangle^2}$$

is the only input required for their singular system analysis. Hence the notion of independence that underlies their analysis is *linear* independence while the notion of *general* independence underlies our redundancy analysis.

Let us illustrate the inadequacy of characterizations in terms of autocorrelation functions by considering two time series obtained by integrating the Lorenz system [16] of differential equations

$$\dot{x} = 10(-x + y),$$
$$\dot{y} = 28x - y - xz,$$
$$\dot{z} = -2\tfrac{2}{3}z + xy,$$

and recording 65536 y values at intervals of $t_s = 0.02$. We added an independent Gaussian noise term to each y value, and then did a discrete Fourier transform. A low pass filter was accomplished by setting all the terms above 0.5 times the Nyquist frequency to zero and transforming back to the time domain. We generated a second filtered time series by setting the same high frequency terms to zero and using a random number generator to adjust the phases of the remaining low frequency terms. Thus we obtained two time series with identical power spectra and consequently identical autocorrelation functions (fig. 1). The first time series (fig. 1(a)) represents noisy measurements of a deterministic process of three degrees of freedom, while the second (fig. 1(b)) is essentially filtered noise with 16387 degrees of freedom. The different characteristics of the two time series can be distinguished by a redundancy analysis but not by their autocorrelation functions.

4. Comparing reconstructions

We need a quantitative measure of the quality of a reconstruction to compare reconstructions suggested by the singular system analysis suggested by Broomhead and King with reconstruc-

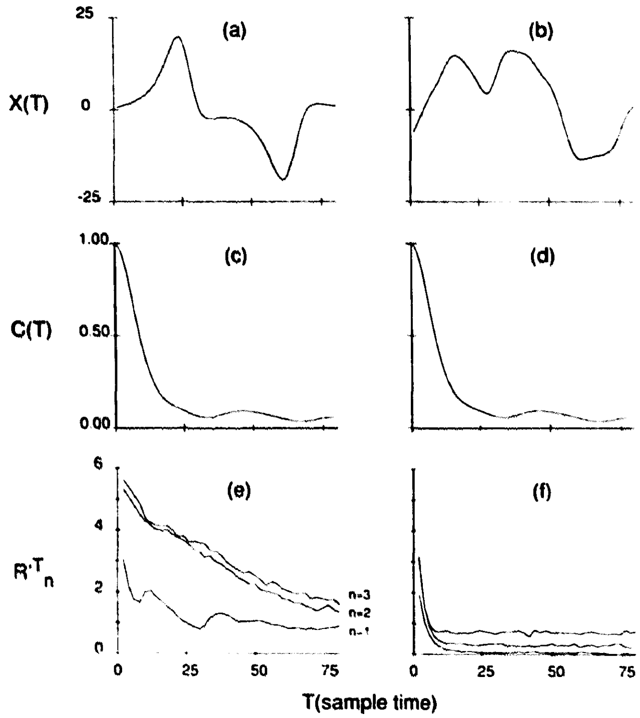

Fig. 1. Two time series. Plot (a) is part of a filtered time series from the Lorenz system. The time series in (b) has the same power spectrum as (a) but in (b) the phases have been randomized. The autocorrelation functions of (a) and (b) appear in (c) and (d) and the redundancy analyses appear in (e) and (f), respectively. Notice that the differences between (a) and (b) affect the redundancy analyses but not the autocorrelation functions.

tions suggested by our redundancy analysis. We develop such a quantitative measure by mimicking Shannon's discussion of fidelity evaluation [11, p. 108]. The reconstruction V should be related in a particular way to the original phase space S. We want to measure the degree to which such a relationship fails to hold with a *distortion functional** $\mathscr{D}(S, V)$.

The measure of quality we use is based on the idea that a reconstruction should be diffeomorphic to the *true* dynamics. We require that it be possible to approximate the diffeomorphism by a piecewise linear map. We fit a piecewise linear map from the reconstruction to our original phase space and consider the residual of the fit to be a measure of the inadequacy or distortion $\mathscr{D}$ of the reconstruction. Thus we have the distortion measure

$$\mathscr{D} = \langle |s - L(\Phi(s))|^2 \rangle, \tag{3}$$

where Φ is the map that describes the reconstruction process and L is the best piecewise linear inverse of Φ that we can fit using a particular partition.

We have developed an algorithm to calculate $\mathscr{D}(S, V)$ which operates on two files, one of which contains points in S the original phase space, while the second contains the reconstructions V of these points obtained via the noisy scalar time series. First the set of reconstructed points is bisected recursively into a partition, whose elements each contain fewer than 32 points. Then linear maps are fit between the points in each partition element and the corresponding points in the original phase space S. Finally the residuals $|s(t) - L(v(t))|^2$ are added up.

Beyond using distortion functionals to simply compare two specific reconstructions generated by different methods, we suggest that the correct way to formally approach the question of how to best reconstruct phase space from scalar time series is in terms of distortion functionals. One should start by writing down a distortion functional $\mathscr{D}(S, V)$ that describes how one would like the reconstruction V to be related to the original phase space S. We have coarsely defined such a functional in eq. (3). One can then consider a set of reconstructions $\tilde{V}$, say all those with linear coordinate functions (eq. (1)), and order them from the least distorted to the most. One should then choose the least distorted reconstruction.

Since the experimental problem is to choose a good reconstruction when the original phase space is not available, a way to determine the ordering implied by $\mathscr{D}(S, V)$ without using S directly is needed. Following Packard et al., we suggest that a reconstruction should allow one to predict measurements that are nearby in time. The idea is that if there is *any* map from a set of intervals of the time series $\{x(t), \dots, x(t+\tau): t \in Z\}$ to the original phase space, then in order to be confident that

*For a good discussion of distortion measures see Shannon [11] or Berger [17].

there is a map from a given reconstruction V to the original phase space, it is sufficient to demonstrate a map from V to $\{x(t),\ldots,x(t+\tau): t \in \mathbf{Z}\}$. Thus what one wants is a new distortion functional of the reconstruction and the time series $\mathscr{D}'(X,V)$ that gives the same ordering of reconstructions as the old $\mathscr{D}(S,V)$.

While we have not implemented the formal approach suggested above, we can use the ideas to examine the differences between techniques of selecting reconstructions. Both the singular system approach of Broomhead and King and our redundancy analysis can be thought of as having a family of candidate reconstructions $\tilde{V}$ and a distortion functional $\mathscr{D}'(X,V)$ that is used to determine which $V \in \tilde{V}$ is best. For the singular system approach suggested by Broomhead and King, $\tilde{V}$ is the set of projections (eq. (1)) with a constraint on the window length, and $\mathscr{D}'(X,V) = -\langle |v|^2 \rangle$. The redundancy analysis that we suggested constrains $\tilde{V}$ to equally spaced delta functions, i.e.,

$$v_k(t) = \sum_j \delta_{kT,j} x(t+j) = x(t+kT),$$

and uses $\mathscr{D}'(X,V) = -Q(m,T)$ (eq. (2)).

Broomhead and King draw from a richer set of candidate reconstructions $\tilde{V}$, but their distortion functional emphasizes irrelevant features. The idea behind the distortion functional that we use for our redundancy analysis is that the number of distinguishable predictions about the state of the time series should be maximized, while Broomhead and King simply try to maximize the separation between reconstructed points. In other words, while we emphasize distinguishability Broomhead and King have simply assumed that "bigger is better".

5. An example

In this section we use an example to illustrate how seriously flawed Broomhead and King's choice of distortion functional is.

5.1. The time series

The example we have chosen is a quasiperiodic attractor in the flow generated by the following system of first-order differential equations:

$$\begin{aligned} \dot{s}_0 &= s_1 + a + s_0(1 - s_0^4), \\ \dot{s}_1 &= 1 - (s_0 + 1)^3, \\ \dot{s}_2 &= \left[s_3 + a + s_2(1 - s_2^4)\right]/b, \\ \dot{s}_3 &= \left[1 - (s_2 + 1)^3\right]/b. \end{aligned} \tag{4}$$

Here $a = -0.3811735$ is chosen to make $\langle s_1 \rangle = 0$, and $b = \sqrt{5} + 1$ is chosen to give our attractor an irrational winding number. The observable is taken to be

$$\psi(s(t)) = s_1(t) + s_3(t)/5,$$

and the measurements are modeled as having Gaussian noise $\eta(t)$ with a standard deviation of 0.03

$$x(t) = \psi(s(t)) + \eta(t) = s_1(t) + s_3(t)/5 + \eta(t). \tag{5}$$

We selected this system and observable because the dynamics are very simple, yet reconstruction is

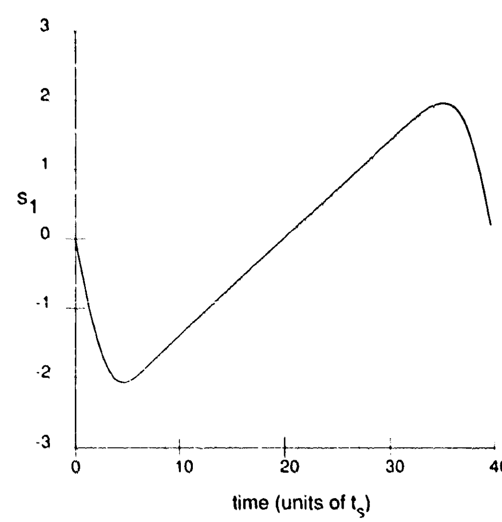

Fig. 2. A relaxation oscillation. This component of the solution to eq. (4) is a straight line for 70% of a cycle to within ±0.03. The sample time is $t_s = 0.02817$ or about 1/40 orbit.

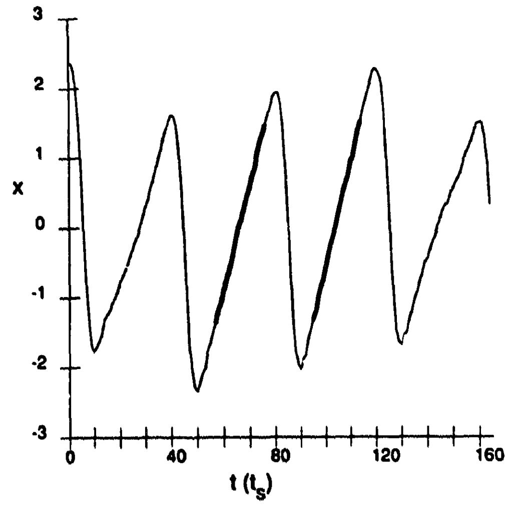

Fig. 3. A time series of the model measurements (eq. (5)). The bold sections (length $= 20t_s$) are nearly identical, but they correspond to very different parts of the attractor. Thus no reconstruction that uses such a short time base can resolve the attractor.

difficult. The system is completely uncoupled between s_0, s_1 and s_2, s_3, and the uncoupled parts are identical except that they have incommensurate time scales. Thus we get quasiperiodicity by design. The nonlinearities have been selected so that the solution $s_1(t)$ is a rounded sawtooth as shown in fig. 2. Such waveforms are typical of the relaxation oscillations that appear in many physical situations. Since the measurements consist of the sum of two sawtooth functions, they have the same slope almost everywhere. Thus by examining a small segment of the time series one cannot determine the phase of either of the sawtooth oscillations (see fig. 3). The power spectrum and autocorrelation function shown in fig. 4 reflect the simple harmonic structure of the signal. Henceforth times will be given in units of the sampling interval which is $t_s = 0.02817$ or about forty times as frequent as the more rapid oscillation, and we will treat the system as a discrete map.

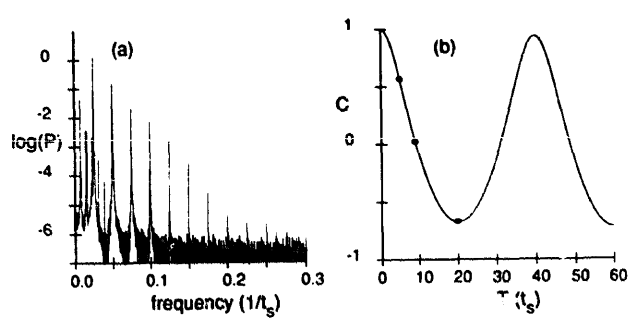

Fig. 4. (a) Power spectrum and (b) autocorrelation function of the model measurements X. In (b) the first inflection point, first zero, and first local minimum are marked at $T = 5$, $T = 9$, and $T = 20$, respectively.

5.2. *Two analyses*

Fig. 5 shows the singular values and basis vectors produced by the analysis that Broomhead and King suggest for this signal. The length of the window vectors was chosen to be 5, as given by the Broomhead–King criterion (the first inflection point in the autocorrelation function). Fig. 6 shows the reconstruction these basis vectors produce. The dense flat region at the top of fig. 6(a) arises from the long straight regions of constant slope in the time series. As indicated by the Poincaré sections of figs. 5(c) and 5(d), these segments of the time series are so straight that information about the second phase cannot be extracted from them.

Fig. 7 shows the type of redundancy analysis we suggested in [5]. Rather than simply invoking eq. (2) to select the best m and T values to use for a reconstruction, let us derive the values by an intuitive discussion of fig. 7. Notice that the $R_n'^T$ curves in fig. 7(a) accumulate along a horizontal line at about 5 bits. The slope of this accumulation line is zero because the measure theoretic entropy is zero ($h_\mu = 0$). The fact that the curves $R_1'^T$ and $R_2'^T$ fall significantly below the accumulation line indicates that neither a single measurement nor a pair of measurements are sufficient to predict a subsequent measurement. In other words, neither $p_{X(T+t)|X(t)}$ nor $p_{X(2T+t)|X_2^T(t)}$ is as "sharp" as $p_{X(3T+t)|X_3^T(t)}$. Thus a 3-d measurement is required to specify the state of the system, and we set $m = 3$. Next we turn to fig. 7(b) to determine a value for T. Because $h_\mu = 0$, the best choice of T is simply the one which minimizes R_3^T. Since the R_3^T curve is almost flat, the choice of T is not critical. R_3^{34} is the actual minimum of R_3^T and we

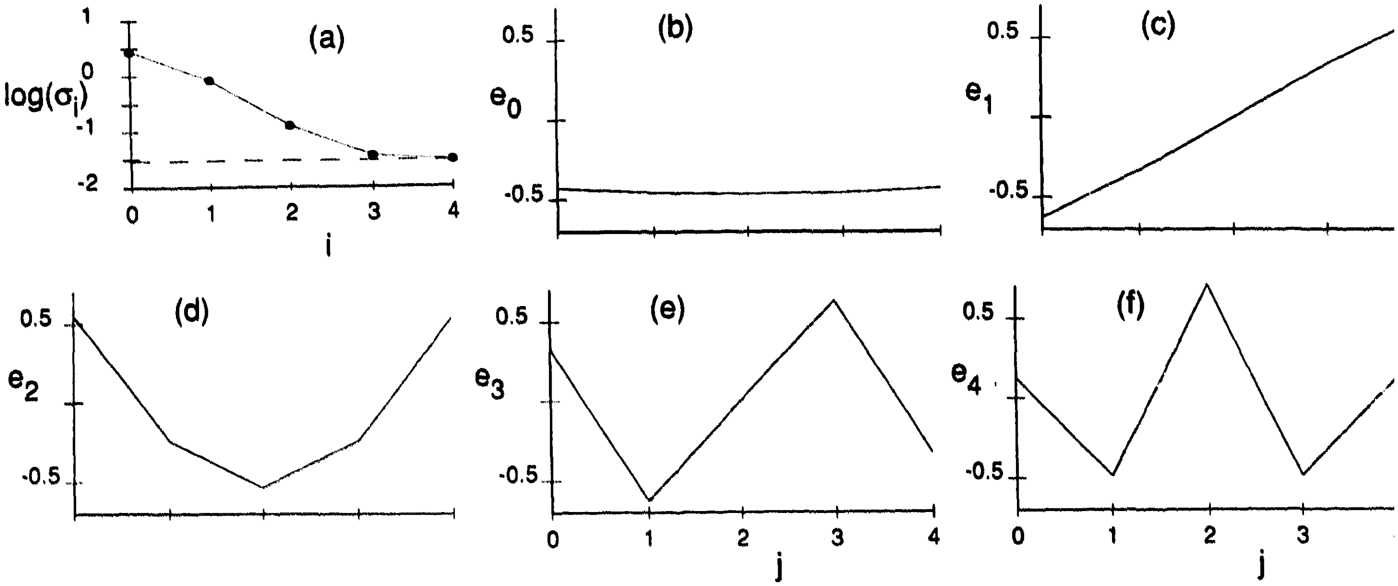

Fig. 5. A singular system analysis with a window length of $5t_s$. Logarithms of the singular values appear in (a) and the singular vectors appear in (b–f). The dashed line in (a) indicates the noise floor σ_η.

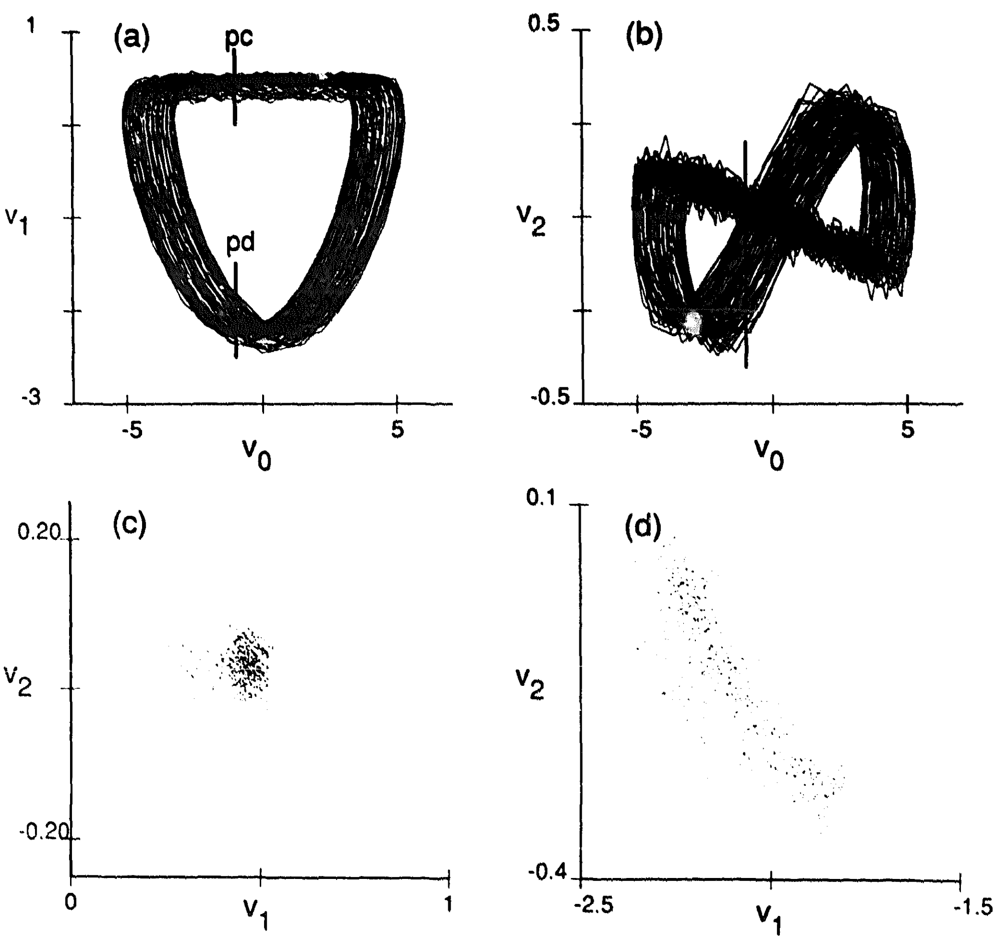

Fig. 6. The reconstruction generated by the first three singular vectors from the analysis of 5-windows (fig. 5). Plots (a) and (b) show projections of the reconstruction. Plots (c) and (d) show the Poincaré sections whose locations are indicated in (a) by pc and pd respectively. Plot (c) confirms that this reconstruction cannot resolve two degrees of freedom in the rising portion of the sawtooth (fig. 3)

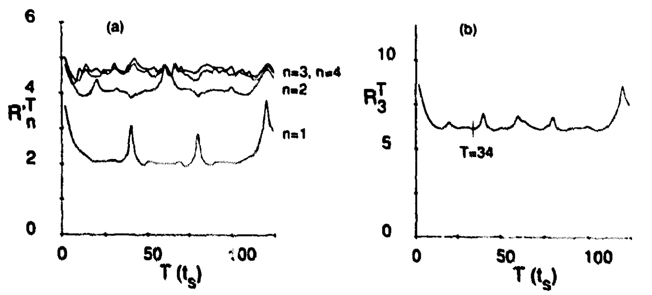

Fig. 7. A redundancy analysis of the model measurements. The marginal redundancy curves in (a) indicate that a three-dimensional reconstruction is appropriate, and the total redundancy curve (b) indicates that the best delay is $T = 34$.

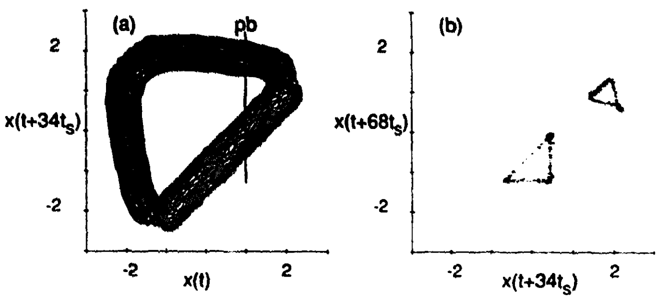

Fig. 8. The reconstruction suggested by the redundancy analysis (fig. 7). A projection is shown in (a). The location of the Poincaré section shown in (b) is indicated by pb in (a). The openings in the Poincaré section indicate the quasiperiodic character of the dynamical system (there are no such openings in fig. 6).

choose $T = 34$. Fig. 8 indicates that the choice of $m = 3$ and $T = 34$ produces a good reconstruction.

We have estimated the distortion (eq. (3)) of the reconstructions represented in figs. 6 and 8 using the algorithm described in the previous section. For the reconstruction suggested by the singular system approach (fig. 6) $\mathscr{D} = 1.01$, and for the reconstruction suggested by a redundancy analysis (fig. 8) $\mathscr{D} = 0.16$. *Thus the reconstruction suggested by redundancy analysis is six times better than the reconstruction suggested by the singular system approach*. To give a sense of scale to these numbers note that if the reconstruction were simply a single point the distortion would be the variance in the original phase space $\sigma_S^2 = \langle(\langle s \rangle - s)^2\rangle = 4.4$.

5.3. *An even better reconstruction*

It could be argued that we have not been fair, in that the reconstruction of fig. 6 reflects a time base of $5t_s$ while the reconstruction of fig. 8 reflects a time base of $68t_s$. It is true that with a longer time base it is possible to get a better reconstruction, but even with a longer window the singular value decomposition technique cannot determine what the good reconstructions are. Fig. 9 shows the first 8 basis vectors and singular values associated with a 520-window, and fig. 10 shows the reconstruction obtained from the first three basis vectors. The reconstruction is terrible because even though the random variables v_0, v_1, and v_2 are *linearly* independent, they are very dependent by any other measure. The first component suggested by fig. 9 that is not very dependent on v_0 is v_4.

Fig. 11 shows that a *nonlinear* reconstruction obtained by

$$v_0(t) = (12 + e_4 \cdot w(t))e_0 \cdot w(t),$$
$$v_1(t) = (12 + e_4 \cdot w(t))e_1 \cdot w(t), \quad (6)$$
$$v_2(t) = e_5 \cdot w(t)$$

is very good. *The point is that the singular value decomposition says very little about general independence, and the first statistically independent basis vector could have been made to fall arbitrarily far down in the sequence of basis vectors by adjusting the amplitude with which we mixed the second frequency into the observable.* We have checked reconstructions based on other window lengths, and as the data in table I indicate, all of these reconstructions have higher distortion $\mathscr{D}$ than the one obtained from a redundancy analysis. The window lengths of 9 and 20 were chosen because in the autocorrelation function they correspond to the first zero and first local minimum, respectively.

The filtering provided by discarding noise-dominated basis vectors is equivalent to doing a weighted average that is local in time. Kostelich and Yorke [18] note that it is better to use a filter that is local in phase space. Using their filter and

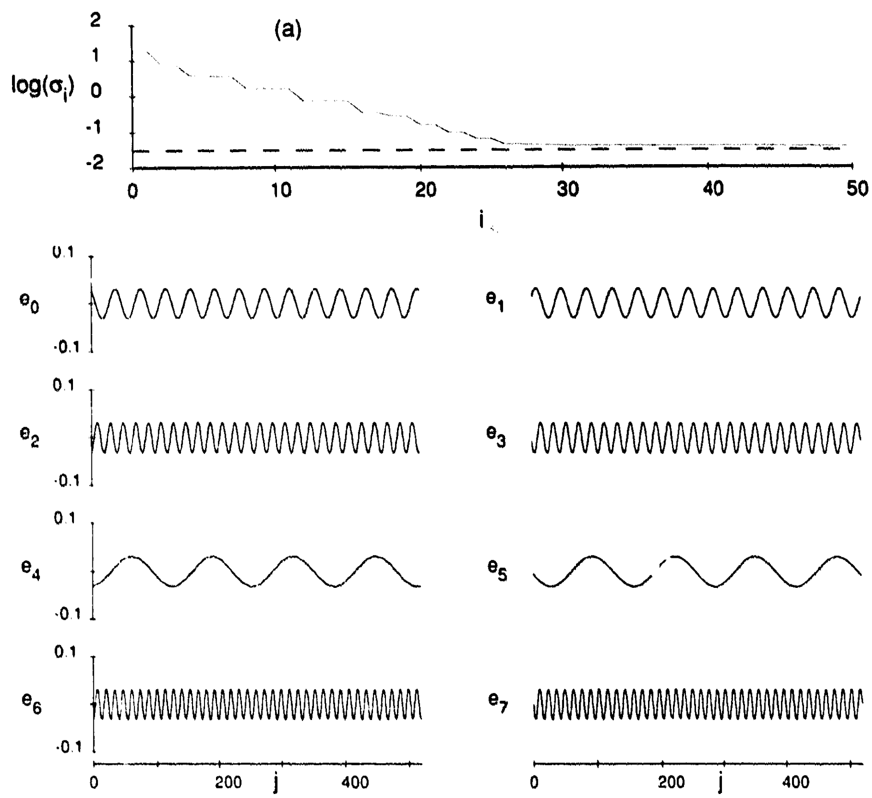

Fig. 9. A singular system analysis of the model measurements using a 520-window. The first 50 singular values are plotted in (a), followed by the first eight basis vectors e_0–e_7 as labeled. The dashed line in (a) indicates the noise floor σ_η.

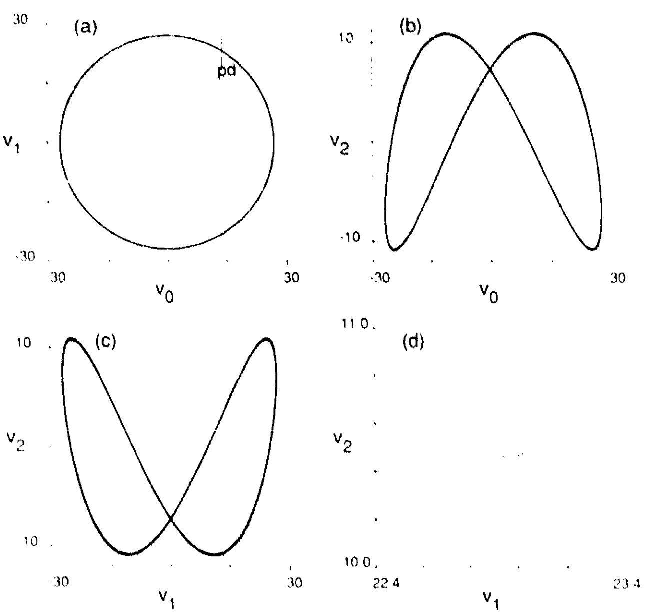

Fig. 10. The reconstruction generated by the first three singular vectors from the analysis of 520-windows (fig. 9). Plots (a), (b), and (c) are projections while plot (d) is the Poincaré section whose location is marked by pd in (a). This reconstruction is a limit cycle to within the noise, while the original attractor is quasiperiodic.

the delay suggested by a redundancy analysis produces a reconstruction with $\mathcal{D} = 0.046$ which is better than any of the other reconstructions using linear coordinate functions in table I.

We selected this example because the dynamics are very simple and reconstruction is difficult. The simplicity of the dynamics is due to quasiperiodicity, while the difficulty in reconstruction is due to the sawtooth character of the relaxation oscillations. Both quasiperiodicity and relaxation oscillations are frequently seen in physical systems.

6. Conclusion

In this paper we have explored the strengths and weaknesses of two approaches to reconstructing attractors from scalar time series. While each approach has areas of relative strength and neither approach is best for all time series, we have argued that on the whole our redundancy analysis [5] is

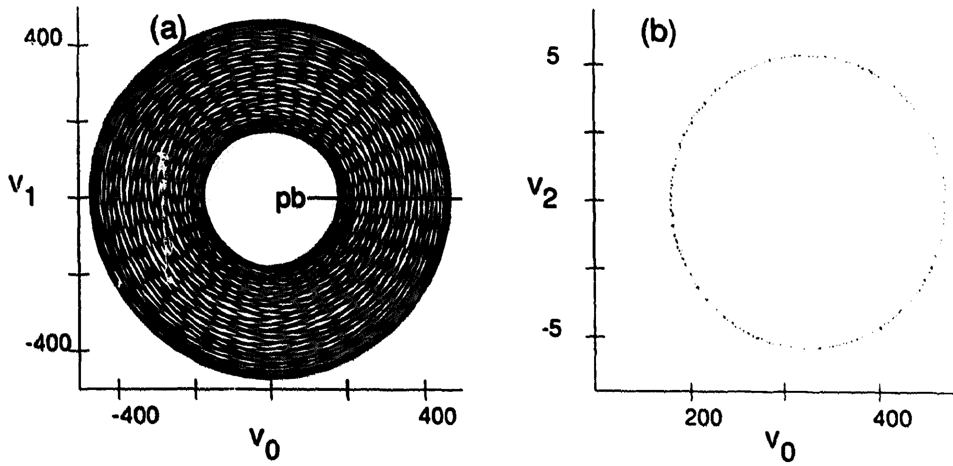

Fig. 11. A reconstruction using a 520-window and the *nonlinear* coordinate functions of eq. (6). Plot (a) is a projection, and plot (b) is the Poincaré section whose location is marked by pb in (a).

Table I
Distortion of various reconstructions. All reconstructions are 3-d. The abbreviation SSA stands for the singular system approach suggested by Broomhead and King [6, 7], and RA stands for the redundancy analysis we suggested in [5]. The distortion measures the degree to which a reconstruction is non-invertible (eq. (3)). The K–Y filter used for the penultimate entry is described in Kostelich and Yorke [18].

Description	Distortion	Average samples/element
SSA 520-window	1.579	15.0
SSA 5-window	0.999	19.8
SSA 9-window	0.647	18.0
SSA 80-window	0.306	17.2
SSA 20-window	0.241	17.3
RA unfiltered	0.156	17.7
RA K–Y filtered	0.045	18.1
Nonlinear eq. (6)	0.0001	18.3

superior to Broomhead and King's singular system approach [6].

We have used an operational measure of distortion, i.e., $\mathscr{D}(S, V)$ in eq. (3), to compare a reconstruction produced by the singular system analysis of Broomhead and King with a reconstruction produced by our redundancy analysis. For the particular time series we considered, our redundancy analysis produced a reconstruction that was 6.4 times less distorted than the reconstruction produced by Broomhead and King's singular system analysis.

Each of the two approaches has a family of possible reconstructions $\tilde{V}$ from which it selects a single reconstruction V. The selection procedures are intended to ensure a particular relationship between the original phase space S and the selected reconstruction V. Since only the time series measurements X and not the phase space S is accessible to an experimenter, each procedure selects a reconstruction V on the basis of a derived distortion measure $\mathscr{D}'(X, V)$. For the singular system approach suggested by Broomhead and King, $\tilde{V}$ is the set of linear projections (eq. (1)) with a constraint on the window length, and $\mathscr{D}'(X, V) = -\langle |v|^2 \rangle$. The redundancy analysis that we suggested constrains $\tilde{V}$ to equally spaced delta functions and uses $\mathscr{D}'(X, V) = -Q(m, T)$ (eq. (2)). The singular system approach selects from a broader class of candidate reconstructions, while the redundancy analysis has a better distortion measure. For the example we considered, the advantage of a good distortion measure proved to be more important than the advantage of selecting from a broad class of reconstructions.

We hope that better reconstruction techniques will be developed. A technique that would draw from a family of candidate reconstructions as large as that considered by Broomhead and King and would use a distortion functional that was as good as ours would be an improvement. Until such an improved distortion measure can be integrated into Broomhead and King's singular system approach, we suggest that any reconstruction produced by their technique be regarded with caution.

Acknowledgements

This work was supported by the Department of Energy Office of Basic Energy Sciences. We thank Harry Swinney, Eric Kostelich and Laurette Tuckerman for helpful suggestions.

References

[1] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, Phys. Rev. Lett. 45 (1980) 712.
[2] F. Takens, in: Dynamical Systems and Turbulence, Warwick (1980), Lecture Notes in Math. 898, D.A. Rand and L.S. Young, eds. (Springer, Berlin, 1981), p. 366.
[3] H. Haucke, R.E. Ecke and J.C. Wheatley, in: Dimensions and Entropies in Chaotic Systems, G. Mayer-Kress, ed. (Springer, Berlin, 1986), p. 198.
[4] A.M. Fraser and H.L. Swinney, Phys. Rev. A 33 (1986) 1134.
[5] A.M. Fraser, I.E.E.E. Trans. Inf. Theory IT-35 (1989), to appear.
[6] D.S. Broomhead and G.P. King, Physica D 20 (1986) 217.
[7] G.P. King, R. Jones and D.S. Broomhead, in: Chaos 87: Int. Conf. Physics of Chaos, and Systems Far from Equilibrium, M. Duong-Van, ed. (North-Holland, Amsterdam, 1987), p. 379.
[8] J.C. Roux, R.H. Simoyi and H.L. Swinney, Physica D 8 (1983) 257.
[9] R.G. Gallager, Information Theory and Reliable Communication (Wiley, New York, 1968).
[10] R.S. Shaw, The Dripping Faucet as a Model Chaotic System (Aerial, Santa Cruz, CA, 1985).
[11] C.E. Shannon and W. Weaver, The Mathematical Theory of Communication (University of Illinois Press, Urbana, 1949).
[12] M. Loeve, Probability Theory (Springer, New York, 1977), 4th ed.
[13] J.J. Dongarra, C.B. Moler, J.R. Bunch and G.W. Stewart, Linpack Users' Guide (SIAM, Philadelphia, 1979).
[14] A. Brandstater, H.L. Swinney and G.T. Chapman, in: Dimensions and Entropies in Chaotic Systems, G. Mayer-Kress, ed. (Springer, Berlin, 1986), p. 150.
[15] A.I. Meese, P.E. Rapp and L.S. Jennings, Phys. Rev. A 36 (1987) 340.
[16] E.N. Lorenz, J. Atmos. Sci. 20 (1963) 130.
[17] T. Berger, Rate Distortion Theory (Prentice-Hall, Englewood Cliffs, NJ, 1971).
[18] E.J. Kostelich and J.A. Yorke, Phys. Rev. A 38 (1988) 1649.