

## Singular-value decomposition and the Grassberger-Procaccia algorithm

A. M. Albano, J. Muench,\* and C. Schwartz†

*Department of Physics, Bryn Mawr College, Bryn Mawr, Pennsylvania 19010*

A. I. Mees

*Department of Mathematics, University of Western Australia, Nedlands, Perth, Western Australia, Australia 6009*

P. E. Rapp

*Department of Physiology and Biochemistry, The Medical College of Pennsylvania, Philadelphia, Pennsylvania 19129*

(Received 14 January 1988)

A singular-value decomposition leads to a set of statistically independent variables which are used in the Grassberger-Procaccia algorithm to calculate the correlation dimension of an attractor from a scalar time series. This combination alleviates some of the difficulties associated with each technique when used alone, and can significantly reduce the computational cost of estimating correlation dimensions from a time series.

### I. INTRODUCTION

The correlation dimension<sup>1,2</sup> has become the most widely used measure of chaotic behavior.<sup>3</sup> It has been used in the analysis of hydrodynamic experiments,<sup>4,5</sup> laser systems,<sup>6</sup> neutron-star luminosities,<sup>7</sup> neuronal<sup>8</sup> and electroencephalographic<sup>9</sup> signals, business cycles,<sup>10</sup> etc.

One of the reasons for this popularity is the relative ease with which it can be calculated from a scalar time series. The calculation typically starts by reconstructing the system's trajectory in an "embedding space" using the method of "lags" or "time delays."<sup>11,12</sup> The dimension of the reconstructed trajectory is then calculated using, more often than not, an algorithm due to Grassberger and Procaccia,<sup>13</sup> although there are also widely used ones due to Termonia and Alexandrovitch<sup>14</sup> and Badii and Politi.<sup>15</sup> It has been known for some time, however, that while it may be easy to calculate an apparently precise value for a dimension, getting a reliably accurate one is often quite something else.<sup>16,17</sup>

Broomhead and King<sup>18</sup> have suggested that singular-value decompositions<sup>19</sup> may provide an alternative way of estimating dimensions. This procedure identifies orthogonal directions in the embedding space which may be ordered according to the magnitude of the variance of the trajectory's projection on them. The ordering is done using the singular values of the embedding. The number of these directions "visited" by the reconstructed trajectory, and indicated by large singular values, is an estimate of the dimension of the smallest space that contains the trajectory. Mees, Rapp, and Jennings,<sup>20</sup> however, point out that the number of large singular values may depend on the details of the embedding and the accuracy of the data as much as they do on the dynamics of the system. Mere counting of large singular values may, therefore, not give a reliable estimate of dimension. However, in subsequent papers Broomhead, Jones, and King<sup>21</sup> and Mees and Rapp<sup>22</sup> have shown that it is possible to estimate the dimension of a state-space manifold by using an appropri-

ately modified procedure.

In this paper, we show that combining singular-value decompositions and the Grassberger-Procaccia algorithm can alleviate some of the ambiguities associated with each technique when used alone. A singular-value decomposition leads to a statistically independent set of variables spanning the embedding space. While the numerical examples shown here are all done with the Grassberger-Procaccia algorithm, singular-value decomposition will apply equally well to other methods of estimating dimension.

In Sec. II we briefly review the embedding procedure and the Grassberger-Procaccia algorithm, incorporating details that will be needed in a subsequent analysis as well as some refinements in the implementation which have been introduced since the original formulation of the algorithm. In Sec. III we review singular-value decompositions and discuss the question of statistical independence. We also establish a connection between the covariance matrix of the embedding and the autocorrelation function of the time series which can serve as a basis for setting limits on the "window length" of the embedding—i.e., the time spanned by each embedding vector.

Section IV presents some computational evidence that for a given time series, the singular-value spectrum depends principally on the window length, corroborating results obtained earlier by Caputo, Malraison, and Atten,<sup>23</sup> in connection with correlation dimensions. We conclude with some suggestions for an interactive procedure combining singular-value decompositions and the Grassberger-Procaccia algorithm.

### II. EMBEDDING AND THE GRASSBERGER-PROCACCIA ALGORITHM

#### A. Embedding

Let a continuous signal  $v(t)$  be measured at equal "sampling intervals,"  $T_s$ , to yield a time series,

$$[v(k) | k=1, 2, \dots, N_T], \quad (2.1)$$

where  $v(k)$  is an abbreviation for  $v(kT_s)$ . We presume  $v(t)$  to be one of  $n$  state variables which completely describe a dynamical system, the trajectory of which lies on a  $d$ -dimensional ( $d \leq n$ ) attractor  $X$  in the system's phase space. Generalizations to multivariable time series are relatively straightforward. For simplicity of presentation they will not be described here. If the system's temporal

$$\begin{aligned} \mathbf{y}(1) &= (v(1), v(1+L), \dots, v(1+(M-1)L)), \\ \mathbf{y}(2) &= (v(1+J), v(1+J+L), \dots, v(1+J+(M-1)L)), \\ &\dots, \\ \mathbf{y}(p) &= (v(1+(p-1)J), v(1+(p-1)J+L), \dots, v(1+(p-1)J+(M-1)L)), \\ &\dots. \end{aligned} \quad (2.2)$$

Here,  $L$  is the "lag," or the number of sampling intervals between successive components of an embedding vector, and  $J$  is the number of sampling intervals between the first components of successive vectors.<sup>24</sup> The time  $(M-1)L$ , spanned by each embedding vector, is the "window length" of the embedding.<sup>18</sup>  $L$  is introduced to allow for the fact that, in an experiment, the sampling interval is often set without accurate prior knowledge of the time scales intrinsic to the system being studied. Determining an appropriate value of  $L$  then becomes part of the analysis. We return to this point in Secs. III and IV. The number  $J$  (or a set of  $J$ 's) describe how the time series is sampled to create a set of embedding vectors with a computationally manageable size from a possibly large data set.<sup>25</sup>

Theorems due to Takens<sup>12</sup> and Mañé<sup>26</sup> state that if the embedding dimension  $M$  and the dimension  $n$  of a manifold containing the attractor satisfy the inequality ("Takens criterion")

$$M \geq 2n + 1, \quad (2.3)$$

then given the assumption of an infinite amount of noise-free data one has a proper embedding except when the system has special symmetries. In particular, the dimension of the embedded or reconstructed attractor is the same as that of the system's phase space attractor.<sup>26</sup> In practice these requirements are not met. However, experimental<sup>6</sup> and computational<sup>17</sup> studies with progressively increased data sets suggest that good approximations to the dimension can often be obtained with fairly small amounts of data. For example, it is possible to estimate the dimension of the Hénon attractor to within 6% of its literature value with 500 data points.<sup>6,17</sup>

### B. Grassberger-Procaccia

In an  $M$ -dimensional embedding, the "correlation integral,"  $C_M(r)$  is defined as the fraction of the distances between embedding vectors that do not exceed  $r$ :

$$C_M(r) = (1/N_p) \sum_{\{i,j\}} \Theta(r - |\mathbf{y}(i) - \mathbf{y}(j)|), \quad (2.4)$$

where  $\Theta(\cdot)$  is the Heaviside unit-step function. To avoid

evolution is chaotic,  $X$  is a strange attractor characterized by a noninteger dimension. For the applications in which we are interested,  $n$  is unknown and  $v(t)$  is the only measured quantity.

Packard *et al.*<sup>11</sup> and Takens<sup>12</sup> have shown that starting from the time series (2.1), one may "embed" or reconstruct the trajectory in an  $M$ -dimensional "embedding space" by means of the vectors,

artificial correlates due to measurements being taken at nearly the same time,<sup>27,28</sup> the sum should be taken only over embedding vectors which are not too closely spaced in time. That is, a number  $K$  is chosen and the sum is taken over those  $i$ 's and  $j$ 's for which  $|i - j| > K$ . In the following, we will take  $|\mathbf{y}(i) - \mathbf{y}(j)|$  to be the Euclidean distance between  $\mathbf{y}(i)$  and  $\mathbf{y}(j)$ .<sup>29</sup>  $N_p$  is the number of distances used in the sum. Grassberger and Procaccia<sup>13</sup> show that the correlation dimension  $D_2$  is given by

$$D_2 = \lim_{M \rightarrow \infty} \lim_{r \rightarrow 0} D_2(M; r), \quad (2.5)$$

where  $D_2(M; r)$  is the slope of the log-log plot of  $C_M(r)$  versus  $r$ :

$$D_2(M; r) = d[\log_e C_M(r)] / d[\log_e(r)]. \quad (2.6)$$

In calculations involving actual experimental data, neither of the limits in Eq. (2.5) can be taken. Small values of  $r$  are blurred by noise and by limitations on experimental accuracy, while large values of  $M$  are precluded by practical limitations on data set sizes and computing times. In practice, for a given value of  $M$ , one looks for a "plateau" in the plot of  $D_2(M; r)$  versus  $\log_e(r)$ —that is, a range,  $r_L$  to  $r_U$  (the "scaling region"), over which  $D_2(M; r)$  has a constant value  $d(r_L, r_U)$ , say, to within some tolerance,  $\pm \Delta d$ . If this plateau is common to a number of embedding dimensions exceeding  $[2d(r_L, r_U) + 1]$ , then it is taken to be the correlation dimension for the range of lengths  $(r_L, r_U)$ .<sup>16</sup>

Although seemingly quite simple, this prescription needs to be applied with some care. Figure 1 shows the slope  $D_2(M; r)$  versus  $\log_e(r)$  for the Lorenz attractor,

$$\begin{aligned} (dx/dt, dy/dt, dz/dt) \\ = (-10(x-y), x(28-z)-y, xy - (\frac{8}{3})z). \end{aligned} \quad (2.7)$$

The calculations used 1000 embedding vectors formed from a data set of 10000 values of  $x$  with  $T_s = 0.01$  and  $L = 5$ . For each embedding dimension, a value of  $J$  was chosen so as to pick embedding vectors uniformly from the entire data set. The plots are for embedding dimensions of (a) 5, (b) 10, (c) 15, (d) 20, and (e) 25. The

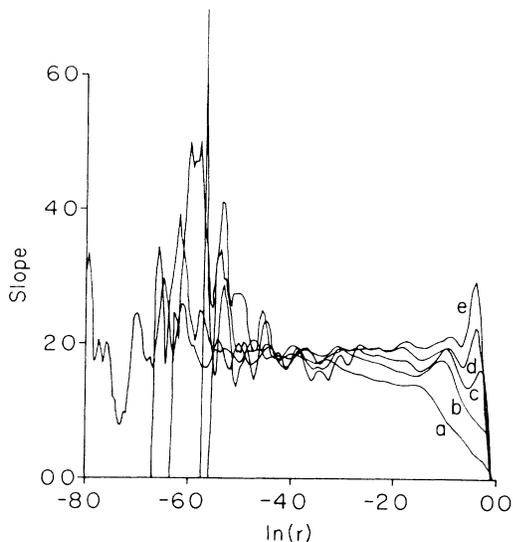


FIG. 1. Plot of slope  $= d[\log_e C_M(r)]/d \log_e(r)$  for data generated by the Lorenz equations with parameter values  $T_S=0.01$ ,  $L=5$ . The five curves correspond to (a)  $M=5$ , (b)  $M=10$ , (c)  $M=15$ , (d)  $M=20$ , (e)  $M=25$ .

length scale  $r$  is normalized to the largest interpoint distance in each embedding. This is done in order to use a length scale derived from the size of the attractor. With this normalization, the correlation integral  $C_M(r)$  [Eq. (2.4)] reaches its maximum value (“saturates”) at the same value of  $r$  making it easier to compare plateaus for different embedding dimensions. [This normalization, however, makes the plots of  $\log_e C_M(r)$  versus  $\log_e r$  inappropriate for calculating the entropy.] In previous work (see, e.g., Refs. 6 and 9), such comparisons were facilitated by plotting  $D_2(M; r)$  versus  $\log_e C_M(r)$ .

For small  $r$ 's [ $\log_e(r) < -5$ ] as well as for large  $r$ 's [ $\log_e(r) > -2$ ], the slopes generally increase with increasing embedding dimension. The behavior at small  $r$  is characteristic of noise which has infinite dimension.<sup>6,13,23,26</sup> At small distances, one has experimental noise, either inherent in the system or contributed by measuring instruments, as well as “computational noise,” arising from round-off errors. At large  $r$ 's, the increase in the slope is induced in part by using a relatively small number of embedding vectors in a high-dimensional space.<sup>23</sup> It becomes more pronounced and extends to smaller values of  $r$  as  $M$  is increased. Thus in the process of increasing the embedding dimension to check the convergence of the calculation, one causes the scaling region to shrink, making it difficult to ascertain where the convergence does occur. Any procedure that can legitimately reduce the required size of the embedding dimension will therefore be helpful. The singular value decomposition does this.

### III. SINGULAR-VALUE DECOMPOSITION AND AUTOCORRELATION

#### A. Singular-value decomposition

It is well known<sup>19</sup> that any  $N \times M$  matrix  $\underline{A}$ , with  $N \geq M$ , may be expressed as

$$\underline{A} = \underline{V} \underline{S} \underline{U}^T, \quad (3.1)$$

where  $\underline{U}^T$  means the transpose of  $\underline{U}$ .  $\underline{S}$  is an  $M \times M$  diagonal matrix,

$$S_{i,j} = \delta_{i,j} s(i), \quad i=1,2,\dots,M \quad (3.2)$$

$\underline{V}$  is an  $N \times M$  matrix with orthogonal columns,

$$(\underline{V}^T \underline{V})_{i,j} = \delta_{i,j}, \quad (3.3)$$

and  $\underline{U}$  is an  $M \times M$  orthogonal matrix,

$$(\underline{U}^T \underline{U})_{i,j} = (\underline{U} \underline{U}^T)_{i,j} = \delta_{i,j}. \quad (3.4)$$

The elements  $s(i)$  of the diagonal matrix  $\underline{S}$  are known as the “singular values” of  $\underline{A}$ .

#### B. Covariance and statistical independence

To relate the embedding process to singular-value decompositions, we follow Broomhead and King<sup>18</sup> and consider the case when  $\underline{A}$  is the “trajectory matrix”—that is, up to a normalization, the matrix whose rows are the  $M$ -dimensional embedding vectors formed from the time series (2.1):

$$\underline{A} = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{y}(1) \\ \mathbf{y}(2) \\ \vdots \\ \mathbf{y}(N) \end{pmatrix}. \quad (3.5)$$

The embedding defines a set of points in an  $M$ -dimensional space which may be described by a multivariate distribution whose variables are the  $M$  components of the embedding vectors, Eq. (2.2). We may rewrite these as

$$\mathbf{y}(k) = (y_1(k), y_2(k), \dots, y_i(k), \dots, y_M(k)).$$

From the definitions of the trajectory matrix  $\underline{A}$  [Eq. (3.5)] and of the covariance matrix  $\mu_{i,j}$  of this multivariate distribution<sup>30</sup> we have

$$\mu_{i,j} = (1/N) \sum_k y_i(k) y_j(k) = (\underline{A}^T \underline{A})_{i,j}. \quad (3.6)$$

If each column of  $\underline{A}$  has zero mean  $\sum_k y_i(k) = 0$  (which we presume hereafter), the off-diagonal elements of the covariance matrix are the (unnormalized) “correlation coefficients” of the distribution. The correlation coefficients measure the statistical dependence of the variables  $y_i$ ;  $i=1,2,\dots,M$ , on each other, or the redundancy of the information which they contain. Variables with vanishing correlation coefficients are, by definition, statistically independent.<sup>30</sup>

The transformation

$$\underline{A} \rightarrow \underline{A}' = \underline{A} \underline{U}, \quad (3.7)$$

or

$$y_i \rightarrow y'_i = \sum_j y_j U_{ji}, \quad (3.8)$$

diagonalizes the covariance matrix:<sup>18,20</sup>

$$\underline{U}^T (\underline{A}^T \underline{A}) \underline{U} = \underline{S}^2. \quad (3.9)$$

The squares of the singular values [ $s(i)$ ]<sup>2</sup> are the eigenvalues of  $\underline{A}^T \underline{A}$ , while the columns of  $\underline{U}$  are its eigenvectors.

Since  $\underline{U}$  is orthogonal, its eigenvectors form an orthonormal basis for the embedding space. The directions of the eigenvectors are called the “principal axes.” In addition, the Euclidean distances  $|\mathbf{y}(i) - \mathbf{y}(j)|$  used in the calculation of the correlation integral Eq. (2.4) are invariant under this transformation. The correlation dimension itself, therefore, is also invariant.  $\underline{A}'$  is called the “matrix of principal components” of  $\underline{A}$ ,  $y'_i$  is the  $i$ th “principal component” of  $y'$ , or its projection along the  $i$ th principal axis.

It also follows from the above that the eigenvalue  $[s(i)]^2$  is the variance of the  $i$ th principal component. If, for some  $j$ ,  $[s(j)]^2 = 0$ , then the reconstructed trajectory does not “visit” principal axis  $j$ . In the absence of noise, this means that the rank of the covariance matrix, which is equal to the number of its nonzero eigenvalues, is the dimension of the smallest subspace of the embedding space that contains the reconstructed trajectory.<sup>18</sup> Noise prevents any eigenvalue from vanishing, so that the dimension is estimated by counting the number of “large” eigenvalues.<sup>18,20</sup>

Although a singular-value decomposition is mathematically equivalent to a diagonalization of the covariance matrix, the former turns out to be more robust numerically. In this work, we use a version of the code developed by Mees, Rapp, and Jennings<sup>20</sup> implementing the Golub-Reinsch algorithm.<sup>31</sup>

### C. Autocorrelation

Using Eqs. (2.2) and (3.5), we may write the  $k, i$  element of the trajectory matrix as

$$A_{k,i} = (\mathbf{y}(k))_i = v(1 + (i-1)J + (k-1)L), \quad (3.10)$$

so that the  $(i, j)$ th element of the covariance matrix  $\underline{A}^T \underline{A}$  is

$$(\underline{A}^T \underline{A})_{i,j} = (1/N) \sum_{p_i} v(p_i) v(p_i + (j-i)), \quad (3.11)$$

where  $p_i = 1 + (k-1)J + (i-1)L$ . In the limit that the number  $N$  of embedding vectors is large,  $(\underline{A}^T \underline{A})_{i,j}$  becomes proportional to the value at  $t = (j-i)L$  of the signal's autocorrelation function,

$$g(t) = [\sum_k v(kL) v(kL + t)] / \{\sum_k [v(kL)]^2\}. \quad (3.12)$$

That is,

$$\lim_{N \rightarrow \infty} (\underline{A}^T \underline{A})_{i,j} = \sigma^2 g[(j-i)LT_S], \quad (3.13)$$

where  $\sigma^2$  is the variance of the time series.

This intimate relationship between the covariance matrix and the autocorrelation function helps to illuminate the critical importance of the window length  $(M-1)L$  in the embedding and in subsequent calculations. By definition,  $g(0) = 1$ , and for decorrelated signals,  $g(t) \rightarrow 0$  for  $t \gg t_c$ , where  $t_c$  is a time interval characteristic of the decay of the autocorrelation. Thus, for these signals, in the limit of large windows, where  $(M-1)L \gg t_c$ , the covariance matrix approaches the unit matrix, which has  $M$  degenerate unit eigenvalues. This means that regardless of the choice of  $M$ , the rank of the matrix is always  $M$ . If we took the rank of the covariance matrix to be an esti-

mate of the trajectory's dimension, we would be led to conclude that we are observing noise. On the other hand, in the limit of very small windows  $(M-1)L \ll t_c$ , all the elements of the covariance matrix approach 1. In the extreme case when all elements are equal to 1, the matrix has one nonzero eigenvalue equal to  $M$  and  $(M-1)$  zero eigenvalues corresponding to the case when the trajectory is projected onto a line segment on the space's main diagonal,  $y_1 = y_2 = \dots = y_M$ .

These effects are evident in Fig. 2 which shows the normalized eigenvalues for 16-dimensional embeddings of the Lorenz attractor with lags  $L = 1, 2, \dots, 9$ . The graphs with the larger eigenvalues correspond to larger windows. It does not seem possible to deduce from these graphs that these eigenvalues pertain to an attractor with a correlation dimension that is slightly greater than 2.

## IV. GRASSBERGER AND PROCACCIA IN ROTATED EMBEDDING SPACE

### A. Testing convergence

The process of checking the convergence of a Grassberger-Procaccia calculation in embedding spaces of relatively high dimensions can obscure the very convergence that is being tested. Counting the number of “large” singular values of a trajectory matrix does not give a reliable estimate of the dimension of the embedded trajectory. However, since the principal components of the embedding vectors form a statistically independent set of variables and since the relative contributions of these variables to the distances used in the calculation of the correlation dimension are directly measured by the eigenvalues, there are obvious advantages to combining these two techniques.

The combination proceeds as follows: For a given embedding, a singular-value decomposition is performed, yielding the matrix of singular values  $S$ , and the orthogonal matrix  $U$  of Eq. (3.1). The trajectory matrix is rotated [Eq. (3.7)] to get the matrix of principal components which is then used in a Grassberger-Procaccia calculation.

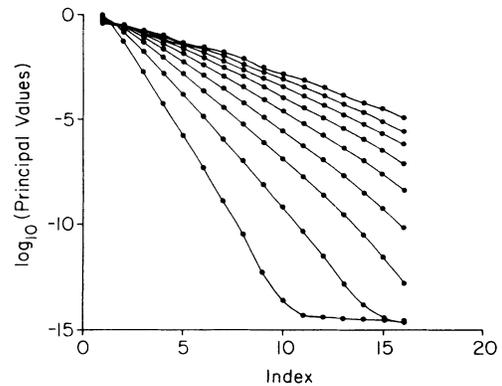


FIG. 2. Normalized principal-value spectrum for data generated from the Lorenz equations. The diagram shows normalized principal values  $\log_{10}(s_j^2 / \sum s_i^2)$  as a function of index  $j$ . The nine spectra correspond, from bottom to top, to lag = 1, 2, 3, ..., 9. The embedding dimension is 16 in each case. One thousand vectors were used in each calculation.

The convergence of the calculation may be checked by performing the calculations in a series of subspaces of the embedding space spanned by principal axes with the largest singular values. The procedure is considered to have converged if the calculations yield similar results for a few of the largest subspaces. Actually, this check need not always be done explicitly. A few calculations are usually enough to establish how large a singular value must be in order for its principal components to introduce a measurable difference in the dimension calculation.

To illustrate this we show in Fig. 3 some results for the Lorenz attractor Eq. (2.7). To compare the effects of principal components characterized by singular values that span several orders of magnitude, we have chosen a nine-dimensional embedding with lag,  $L = 1$ . This gives a window length that would be too small for a "good" dimension calculation (see Sec. V) but does give a range of eigenvalues spanning 15 orders of magnitude. The figure consists of four superimposed graphs of slope versus  $\log_e(r)$  obtained by using principal components corresponding to the largest 2, 3, 4, and 9 eigenvalues, that is, 2, 3, 4, and 9 columns of the principal component matrix  $AU$  were used in the calculations. The *normalized eigenvalues*  $[s(i)]^2 / \{\sum_{j=1, M} [s(j)]^2\}$  are  $0.98$ ,  $1.62 \times 10^{-2}$ ,  $1.69 \times 10^{-4}$ ,  $1.48 \times 10^{-6}$ ,  $1.65 \times 10^{-8}$ ,  $7.46 \times 10^{-11}$ ,  $4.35 \times 10^{-13}$ ,  $9.25 \times 10^{-15}$ , and  $4.37 \times 10^{-15}$ . In this case, inclusion of principal components with normalized singular values less than  $10^{-4}$  has a negligible effect on the results of the calculation. This was found to be true for other embeddings of the Lorenz attractor as well as for the Rössler attractor [Eq. (4.1)] and a 3-torus [Eq. (4.2)]. In the case of Fig. 3, the calculations clearly agree for  $\log_e(r) > -3$  when only the two largest principal components are used.

Again we explicitly note that this diagram does not demonstrate a successful calculation of the dimension of

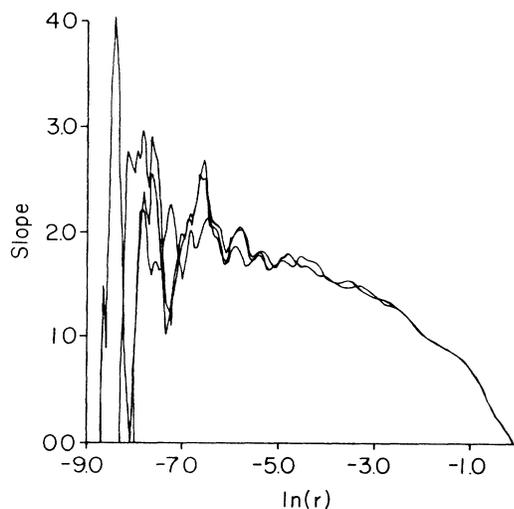


FIG. 3. Plot of slope vs  $\log_e(r)$  for data generated by the Lorenz equations with parameter values  $T_s = 0.01$  and  $L = 1$ . Matrix  $A$  was formed by 1000 vectors embedded in a nine-dimensional space. Correlation integrals and the corresponding derivatives shown here were calculated using the first 2, 3, 4, and 9 columns of the rotated principal component matrix  $A \cdot U$ .

the Lorenz attractor. Because  $L = 1$ , there is no plateau in the derivative of the correlation integral that can be used to estimate the dimension. The issue of an appropriate lag is addressed presently. The purpose of this calculation is to demonstrate that, given an appropriate singular-value spectrum, only a few columns of the principal component matrix are required to estimate accurately the correlation integral. Using the Golub-Reinsch algorithm, singular-value decompositions can be rapidly computed. For example, the double precision decomposition of a  $2000 \times 10$  matrix can be performed in approximately 84 sec on an 8-MHz personal computer with a floating point coprocessor. Singular-value decomposition is not a major computational cost because the computational costs of dimension-estimation procedures are mainly in the calculation of the correlation integral. A correlation integral calculation using 1000 ten-dimensional vectors takes some 15 min on the same computer. Because the time required to calculate these integrals increases rapidly with embedding dimension, the results summarized in Fig. 3 are of considerable practical significance.

A comparison of Figs. 1 and 3 displays quite vividly the advantage of implementing the algorithm in terms of the principal components. One can check for convergence without introducing noiselike behavior at large distances.

#### B. Window length

Results obtained by Broomhead and King<sup>18</sup> and by Mees, Rapp, and Jennings<sup>20</sup> suggest that for a given time series, the normalized singular values may depend only on the window length  $(M - 1)L$ , and not separately on the embedding dimension  $M$ , or the lag  $L$ . Figure 4 shows a semilog plot of the normalized eigenvalues versus the index for a number of embeddings with window lengths that are nearly equal. The values of embedding

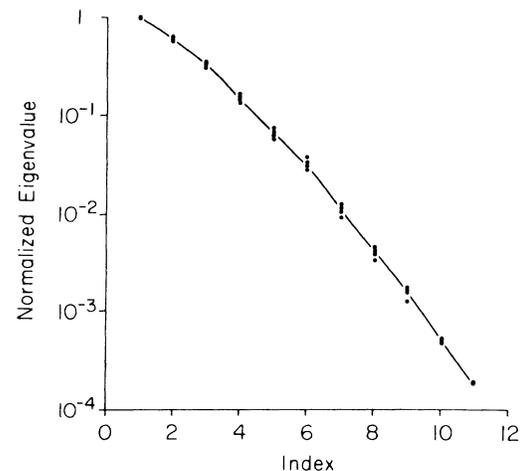


FIG. 4. Normalized principal values  $\log_{10}(s_j^2 / \sum s_j^2)$  plotted as a function of index  $j$  for data generated by the Lorenz equations. In each case, matrix  $A$  contained 1000 rows. The values of embedding dimension  $M$  and lag  $L$  are  $(M, L) = (5, 24)$ ,  $(6, 19)$ ,  $(7, 16)$ ,  $(8, 14)$ ,  $(9, 12)$ ,  $(10, 11)$ , and  $(11, 10)$ . The solid line connects the eigenvalues for the 11-dimensional embedding.

dimension and lag  $(M, L)$  used were (5,24), (6,19), (7,16), (8,14), (9,12), (10,11), and (11,10). The eigenvalues are seen to fall essentially on a single curve. The solid lines connect eigenvalues for the 11-dimensional embedding with a lag of 10.

This dependence on window length rather than separately on embedding dimension and lag is also displayed by results of correlation dimension calculations. On the basis of extensive calculations on the Lorenz attractor, Caputo *et al.* suggest that the local slope obeys a “universal law” that depends only on the window length.<sup>23</sup> Figure 5 shows plots of the slope of  $\log_e[C_M(r)]$  versus  $\log_e(r)$  for the Lorenz attractor, Eq. (2.7), for  $(M, L) = (5, 24)$ , (8,14), (9,12), and (11,10). The thick, dark curve is a superposition of results from the three largest subspaces of the 8-, 9-, and 11-dimensional embeddings, while the light curve is from the 5-dimensional embedding using all five components. Note that embedding dimension 5 does not satisfy Taken’s criterion (2.3) for a proper embedding. Nevertheless, in the scaling region  $-3 < \log_e(r) < -1$ , the results for five dimensions do not differ from the rest by more than 10%. Of course many  $n$ -dimensional manifolds (and attractors contained in them) can be embedded in fewer than  $2n + 1$  dimensions, and indeed what we have here is an embedding of  $\mathbb{R}^3$  in  $\mathbb{R}^5$ . Figures 1 and 5 show the relative effects of varying the embedding dimension for fixed lag and of varying both lag and embedding dimension, while keeping the window length approximately constant. It could be argued that the results in Fig. 5 might reflect a behav-

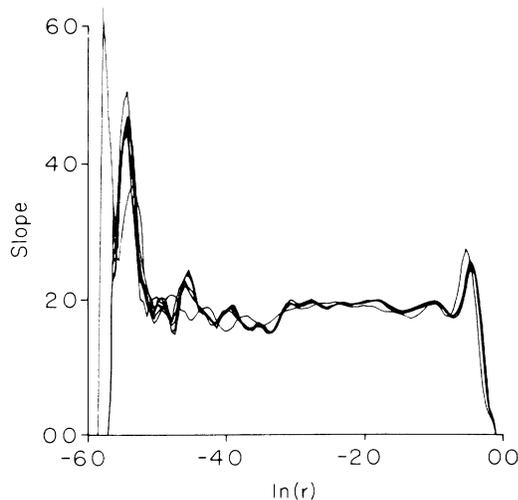


FIG. 5. Plot of slope vs  $\log_e(r)$  for data generated by the Lorenz equations. In each case, the trajectory matrix  $A$  had 1000 rows. The curves correspond to embedding dimension  $M$  and lag  $L$  of  $(M, L) = (5, 24)$ , (8,14), (9,12), and (11,10). The thick dark curve results from the superposition of three calculations each for  $M = 8, 9$ , and 11. In each embedding dimension  $M$ , the first  $M - 2$ ,  $M - 1$ , and  $M$  columns, respectively, of the rotated matrix  $A \cdot U$  were used in the dimension calculation. The thin curve corresponds to the five-dimensional embedding in which all five columns were used in the dimension calculation.

ior unique to the Lorenz equations. Additional computations on other systems suggest that this is not the case.

Figure 6 is similar to Fig. 5, but for the Rössler attractor,

$$(dx/dt, dy/dt, dz/dt)$$

$$= (-y - z, x + 0.2y, 0.4 + xz - 5.7z), \quad (4.1)$$

for  $(M, L) = (6, 13)$ , (7,11), (9,8). Figure 7 shows results for the 3-torus,

$$x(k) = \sin(\omega_0 k) + \sin(2^{3/2} \omega_0 k) + \sin(3^{3/2} \omega_0 k), \quad (4.2)$$

with  $\omega_0 = 6.0 \times 10^{-3}$  and for  $(M, L) = (7, 20)$ , (9,15), (10,13), and (11,12). All cases show remarkable stability as both lag and embedding dimension are changed while the window length remains approximately constant.

Upon reflection, the dependence of the correlation integral on window length rather than on lag or embedding dimension should have been expected. Similar window lengths compare similar segments of a trajectory. Changes of lag within the window simply give different discrete approximations of the correlation integral.

## V. WHAT IS A GOOD WINDOW?

In Sec. III B we saw that the window length used in the embedding crucially affects the singular-value spectrum. It also affects the outcome of attempts to estimate the attractor’s dimension. The importance of choosing an appropriate window was pointed out by Froehling *et al.*<sup>32</sup> who proposed using 10% of the “folding time” which was defined as “the average time between “folding” of adjacent sheets of the attractor.”<sup>32</sup> Unfortunately this criterion is difficult to implement in a systematic way for all attractors. Aspects of the importance of window

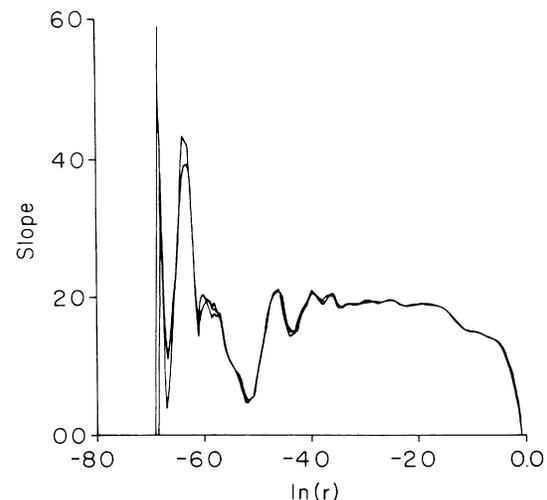


FIG. 6. Plot of slope vs  $\log_e(r)$  for data generated by the Rössler equations ( $T_s = 0.05$ ). In each case matrix  $A$  consisted of 1000 rows. The curves correspond to embedding dimension  $M$  and lag  $L$  of  $(M, L) = (6, 13)$ , (7,11), and (9,8).

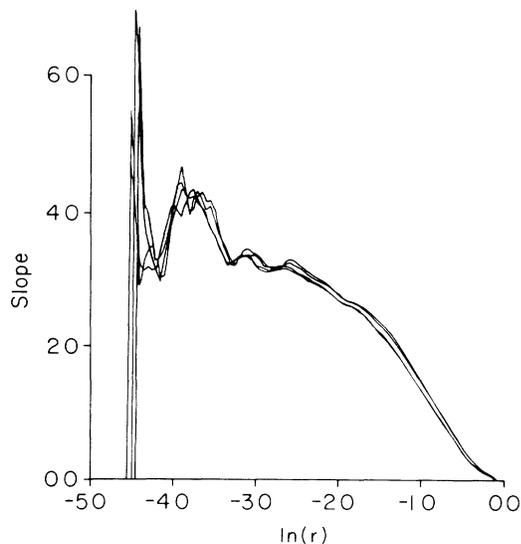


FIG. 7. Plot of slope vs  $\log_e(r)$  for the 3-torus Eq. (4.2). In each case matrix  $A$  consisted of 1000 rows. The curves correspond to embedding dimension  $M$  and lag  $L$  of  $(M,L)=(7,20)$ ,  $(9,15)$ ,  $(10,13)$ , and  $(11,12)$ .

value to dimension estimates using the Grassberger-Procaccia algorithm are shown in Figs. 1 and 3. The various graphs in Fig. 1 are results of calculations using different window lengths. There is so much variability between graphs that one cannot tell if the calculations are converging to a limit. Figure 3, on the other hand, shows that the calculations agree, but it does not display a scaling region, and so does not yield a value for the correlation dimension. Problems such as these have been known for some time and there have been a number of efforts to define an optimum window.<sup>17</sup>

The first issue to address is why there should be an optimal window. It is clear that there is a lower bound to the window length. If the window is too short, noise dominates. Up to a point, we want the largest window possible. The question is, why should we have to stop? Why not use giant windows? The answer to this question is not to be found in the Grassberger-Procaccia algorithm, but rather in the Takens's theorem<sup>12</sup> that produces the theoretical foundations for the calculation. Takens's theorem states that

$$(v(J), v(J+L), v(J+2L), \dots, v(J+(M-1)L))$$

is an embedding. However, if the system decorrelates, as it will in chaotic cases, we must ensure that, as required by the theorem, all of these points are on the same trajectory by ensuring that  $(M-1)L$  is not too big. Data generated by numerical integration or in noisy experiments do not present a trajectory but a perturbation of a trajectory. This puts an upper bound on an acceptable value of  $(M-1)L$ . Effective window length is thus bounded from below and from above. The present object of our investigation is to locate the optimal window within this range.

The relationship between the covariance matrix and the autocorrelation function of the signal [Eq. (3.13)] suggests that a time-scale characteristic of the autocorrela-

tion function is an appropriate time scale to use. In previous work we have used the first zero of the autocorrelation function,<sup>33</sup> but subsequent calculations on data from simulations as well as from experiments show that the correlation time, which is defined as the time required for the autocorrelation function to decrease to  $1/e$  of its original value, is more robust. For band-limited data, Broomhead and King use the inverse of the band-limiting frequency.<sup>18</sup> While straightforward in principle, numerical estimation of the band-limiting frequency poses some practical problems.

In slightly different terms but in similar context, Caputo, Malraison, and Atten<sup>23</sup> use a number of the order of the first return time, while Fraser and Swinney<sup>34</sup> use the first minimum of the mutual information function. We have performed a series of calculations of mutual information for these three example systems that are analogous to those presented here of the autocorrelation function. In summary, the mutual information function did not prove to be a reliable indicator of optimal window length. Additionally, significant technical problems are associated with numerical estimation of mutual information that are very nearly as complex as those encountered in dimension calculations themselves. A systematic account of these results is given in Martinerie *et al.*<sup>35</sup>

To test if an "optimum" window length does exist and, if it exists, whether it is related to a time scale characteristic of the correlation function, we performed a series of calculations using a variety of windows. The analysis was performed on the data used for Figs. 5–7. These were the Lorenz attractor [Eq. (2.7)], the Rössler attractor [Eq. (4.1)], and a 3-torus [Eq. (4.2)]. The results of the calculations are shown in Figs. 8–10, respectively, and are summarized in Table I.

For our present purposes, the optimum window is

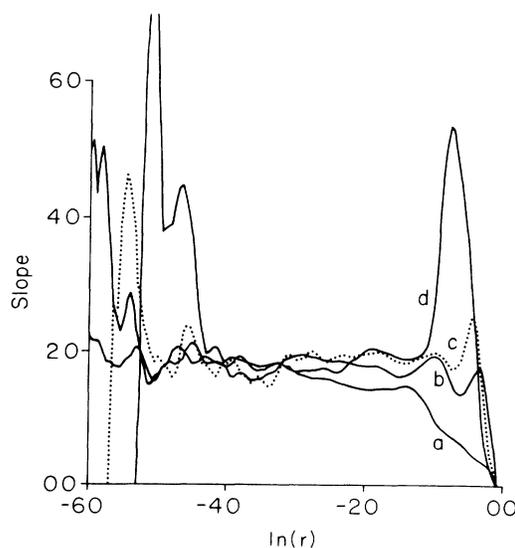


FIG. 8. Plot of slope vs  $\log_e(r)$  for data generated by the Lorenz equations. In each case matrix  $A$  consisted of 1000 rows. The curves correspond to embedding dimension  $M$  and lag  $L$  of (a)  $(M,L)=(5,6)$ , (b)  $(8,10)$ , (c)  $(11,10)$ , and (d)  $(9,30)$ .

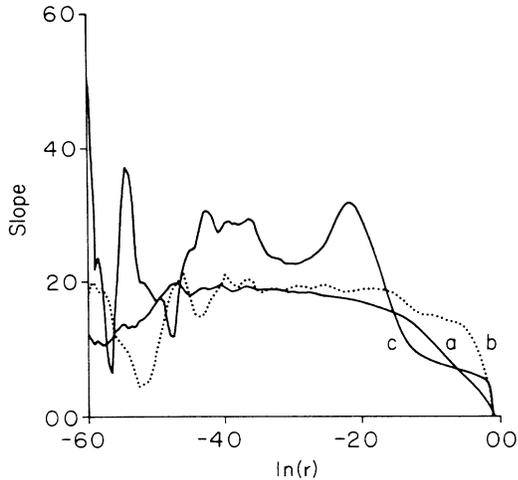


FIG. 9. Plot of slope vs  $\log_e(r)$  for data generated by the Rössler equations. In each case matrix  $A$  consisted of 1000 rows. The curves correspond to embedding dimension  $M$  and lag  $L$  of (a)  $(M,L)=(9,3)$ , (b)  $(7,11)$ , (c)  $(9,12)$ .

merely that which gives the broadest plateau at the attractor's known correlation dimension. For each window, the time series is embedded in an  $M$ -dimensional space that satisfies Takens's criterion (2.3). A singular-value decomposition is performed on the trajectory matrix, and the transformation Eq. (3.7) is done. This is followed by a Grassberger-Procaccia calculation in a subspace of the embedding space spanned by principal axes with normalized eigenvalues exceeding  $1.0 \times 10^{-4}$ .

Figure 8 shows results for the Lorenz attractor using a data set with a correlation time of  $30T_S$ . The four graphs are labeled according to the window length used in the calculation: (a)  $24T_S$ , (b)  $70T_S$ , (c)  $100T_S$ , (d)  $240T_S$ . Curves (a) and (b) corresponding to the smaller window lengths show low values of the slope for large values of  $r$  [ $\log_e(r) > -3.5$ ] and oscillate as the small, noise-dominated length scales are reached. On the other hand, curve (d), with a window length that is eight times the correlation time, shows an anomalously large peak at large values of  $r$  [ $\log_e(r) > -1.0$ ], while becoming as erratic as the curves for small windows at the small length scales. Curve (c), with a window length three times the correlation time, displays a plateau in the region  $-3.0 < \log_e(r) < -0.6$ , at a slope of  $2.0 \pm 0.1$ .

Figure 9 shows similar results for the Rössler attractor,

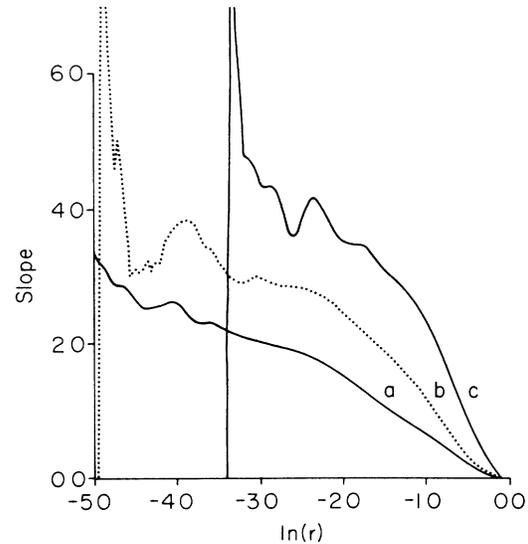


FIG. 10. Plot of slope vs  $\log_e(r)$  for the 3-torus. In each case matrix  $A$  consisted of 1000 rows. The curves correspond to embedding dimension  $M$  and lag  $L$  of (a)  $(M,L)=(9,5)$ , (b)  $(11,12)$ , (c)  $(10,20)$ .

for a data set with a correlation time of  $23T_S$ . Curve (a) is for a window of  $24T_S$ , (b) for  $66T_S$ , and (c) for  $96T_S$ . Figure 10 is for a 3-torus with a correlation time of  $62\tau_S$ , curves (a), (b), and (c) corresponding, respectively, to window lengths of  $40T_S$ ,  $120T_S$ , and  $180T_S$ .

The results of comparisons of window length and the autocorrelation function are summarized in Table I. In all three cases, windows of the order of the correlation time or smaller result in small or nonexistent plateaus with values of the slope that are smaller than the attractor's dimension except in noise-dominated regions at small distance scales. On the other hand, windows of the order of four times the correlation time or greater result in similarly small or nonexistent plateaus with values of the slope that are larger than the attractor's dimension. In all three cases, the broadest plateaus, at values within 10% of the attractor's dimension, are obtained for windows that are between 1.6 and 3.5 times the correlation time. In these numerical experiments,  $T_{\min}$  tracks  $T_{\text{window}}$  slightly better than  $T_{AC}$ . However, as previously mentioned, numerical difficulties are frequently encountered in estimating  $T_{\min}$ .  $T_{AC}$  estimates are much more robust against variations in size and quality of the data

TABLE I. Comparison of window length with the autocorrelation function.  $T_{AC}$  is the autocorrelation time,  $T_{\text{zero}}$  is the time of the first zero crossing of the autocorrelation function, and  $T_{\min}$  is the time of the first local minimum of the autocorrelation function.  $T_{\text{window}}$  is the window length of those tested that results in the broadest plateau at the attractor's known dimension. All times are reported in units of sample interval  $T_S$ .

	$T_{AC}$	$T_{\text{zero}}$	$T_{\min}$	$T_{\text{window}}$	$T_{\text{window}}/T_{AC}$
Lorenz	30	122	72	100	3.33
Rössler	23	32	60	66	2.87
3-torus	62	90	122	100	1.61

set. For this reason, the present discussion focuses on  $T_{AC}$ .

These results indicate that though the correlation time is an imperfect predictor of the optimal window, the rate at which a signal decorrelates is certainly a determinant of optimal window. That this should be so was anticipated by our previously outlined examination of Takens's theorem. The Lyapunov exponents, particularly the system's largest Lyapunov exponent, provide a quantitative measure of the average exponential divergence of nearby trajectories in the phase space of a dynamical system.<sup>36</sup> We hypothesize that if the product (window  $\cdot L_{\max}$ ), where  $L_{\max}$  is the largest Lyapunov exponent, is too large, the signal decorrelates and the scaling region is lost. The smaller the maximum Lyapunov exponent, the larger we can take the window. Tests of this hypothesis using a numerical procedure for estimating Lyapunov exponents constructed by Wolf *et al.*<sup>37</sup> are now underway.

While these admittedly anecdotal results confirm the overall dependence of the results of the calculations on the window length, they do not give a precise prescription that would allow determination of an "optimum" in terms of the signal's correlation time. They do, however, suggest a relatively narrow range over which the window length may be varied to maximize the plateau.

We propose the following interactive procedure for combining a singular-value decomposition and the Grassberger-Procaccia algorithm.

- Choose an embedding dimension  $M$  and a lag  $L$  so that the window  $(M - 1)L$  is a few times larger than the correlation time.
- Perform a singular-value decomposition and rotate the embedding space using the matrix  $\underline{U}$  of eigenvectors of the covariance matrix.
- Perform Grassberger-Procaccia calculations in subspaces of the rotated space spanned by principal axes with eigenvalues exceeding  $10^{-4}$ . If this results in a dimension that does not satisfy Takens' criterion, increase  $M$  until it does. (In practice,  $10^{-4}$  is not invariably an appropriate criterion. However, it cuts off dimensions that contribute no more than 0.01% to the variance of the multivariate data. If the quality of the data used and the precision of the calculations warrant it, more or less stringent criteria may be used.)
- Vary the window length to maximize the plateau.

## VI. FURTHER PROBLEMS AND POSSIBILITIES

Establishing  $r_L$  and  $r_U$ , the lower and upper bounds of the derivative plateau, presents computational problems that are not readily resolved. In calculations using noisy

experimental data, the plateau boundaries may not be obvious. Numerical experimentation demonstrates that even small variations in  $r_L$  and  $r_U$  can have significant effects on the dimension estimate. Caswell and Yorke<sup>16</sup> have suggested including explicit values of  $r_L$  and  $r_U$  in dimension calculations;  $D_c(r_L, r_U)$  would be reported. In many experimental investigations, the absolute value of dimension is not of interest, rather the change in dimension in response to changes in experimentally controlled parameters (Reynolds number, temperature, cognitive activity) may be of interest. In this case, the importance of systematic errors associated with plateau boundary estimates can be reduced by reporting the ratio of dimensions estimated with common values of  $r_L$  and  $r_U$ . In some cases no single number can accurately approximate the derivative of the correlation dimension. In these instances, producing the Slope versus  $\log_e(r)$  plot is perhaps the only legitimate means of presenting the results of these calculations.

The results presented here have addressed the question of window length. The question of epoch length has not been considered. If  $N_T$  data points are collected at a uniform sample interval  $T_s$ , then the epoch length, the duration of the record, is given by  $T_E = T_s(N_T - 1)$ . Selection of an appropriate epoch length is a much more difficult problem than selecting a window because its resolution turns on questions of signal stationarity. This problem has already been encountered in classical signal analysis, and the generalization of a classical technique used in the characterization of nonstationary signals may be applicable here. The Wigner transform<sup>38</sup> estimates the instantaneous energy for a given time and frequency. Others have shown<sup>7,39</sup> that it is possible to generalize the concept of dimension to construct a continuous dimension spectrum. We have proposed<sup>17</sup> that it should be possible to construct an analog to the Wigner transform that can be applied successfully to the dimension spectra of nonstationary signals.

## ACKNOWLEDGMENTS

We thank N. B. Abraham for his numerous illuminating comments and suggestions. The implementation of the Golub-Reinsch singular-value decomposition algorithm we used is due to L. S. Jennings, to whom our profuse thanks are extended. This work was supported in part by Contract No. 22919-PH with the U.S. Army Research Office (A.M.A.), Grant No. NS19716 from the U.S. National Institutes of Health (P.E.R.), and a Helena Rubinstein Foundation Grant (C.S.). A.I.M. thanks King's College, University of Cambridge for support and hospitality.

\*Present address: Department of Oceanography, University of Washington, Seattle, WA 98109.

†Present address: Bell Communications Research, 331 Newman Springs Road, Red Bank, NJ 07701-7020.

<sup>1</sup>H. G. E. Hentschel and I. Procaccia, *Phys. Rev. Lett.* **50**, 346

(1983).

<sup>2</sup>J. D. Farmer, E. Ott, and J. A. Yorke, *Physica* **7D**, 153 (1983).

<sup>3</sup>See, e.g., *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress (Springer, Berlin, 1986).

<sup>4</sup>B. Malraison, P. Atten, P. Bergé, and M. Dubois, *J. Phys.* **50**,

- 897 (1983).
- <sup>5</sup>A. Brandstätter, J. Swift, H. L. Swinney, A. Wolf, J. D. Farmer, E. Jen, and P. J. Crutchfield, *Phys. Rev. Lett.* **51**, 1442 (1983).
- <sup>6</sup>See, e.g., A. M. Albano, J. Aboundadi, T. H. Chyba, C. E. Searle, S. Yong, R. S. Gioggia, and N. B. Abraham, *J. Opt. Soc. Am. B* **2**, 102 (1985); G. P. Puccioni, A. Poggi, W. Gadowski, F. T. Arecchi, and J. R. Tredicce, *Phys. Rev. A* **34**, 2073 (1986); N. B. Abraham, A. M. Albano, T. H. Chyba, L. M. Hoffer, M. F. H. Tarroja, S. P. Adams, and R. S. Gioggia, in *Instabilities and Chaos in Quantum Optics*, edited by F. T. Arecchi and R. G. Harrison (Springer, Berlin, 1987).
- <sup>7</sup>W. Voges, H. Atmanspacher, and H. Scheingraber, *Astrophys. J.* **320**, 794 (1987); H. Atmanspacher, H. Scheingraber, and W. Voges, *Phys. Rev. A* **37**, 1314 (1988).
- <sup>8</sup>P. E. Rapp, I. D. Zimmerman, A. M. Albano, G. C. de Guzman, and N. N. Greenbaun, *Phys. Lett.* **110A**, 335 (1985); G. Mpitsos, in *Dynamic Patterns of Complex Systems*, edited by J. A. S. Kelso, A. J. Mandell, and M. F. Schlesinger (World Scientific, Singapore, 1988).
- <sup>9</sup>A. M. Albano, N. B. Abraham, G. C. de Guzman, M. F. H. Tarroja, D. K. Bandy, R. S. Gioggia, P. E. Rapp, I. D. Zimmerman, N. N. Greenbaun, and T. R. Bashore, in Ref. 3, p. 231; P. E. Rapp, I. D. Zimmerman, A. M. Albano, G. C. de Guzman, N. N. Greenbaun, and T. R. Bashore, in *Nonlinear Oscillations in Chemistry and Biology*, edited by H. G. Othmer (Springer, Berlin, 1987).
- <sup>10</sup>W. A. Brock and C. L. Sayres (unpublished).
- <sup>11</sup>N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
- <sup>12</sup>F. Takens, in *Dynamical Systems of Turbulence*, edited by D. A. Rand and L.-S. Young (Springer, Berlin, 1981).
- <sup>13</sup>P. Grassberger and I. Procaccia, *Physica* **9D**, 189 (1983); *Phys. Rev. Lett.* **50**, 349 (1983); *Phys. Rev. A* **28**, 2591 (1983); *Physica* **13D**, 34 (1984).
- <sup>14</sup>Y. Termonia and Z. Alexandrovitch, *Phys. Rev. Lett.* **45**, 1265 (1983).
- <sup>15</sup>R. Badii and A. Politi, *J. Stat. Phys.* **40**, 725 (1985).
- <sup>16</sup>W. E. Caswell and J. A. Yorke, in Ref. 3, p. 123.
- <sup>17</sup>A. M. Albano, A. I. Mees, G. C. de Guzman, and P. E. Rapp, in *Chaotic Biological Systems*, edited by A. V. Holden (Pergamon, Oxford, 1987).
- <sup>18</sup>D. S. Broomhead and G. P. King, *Physica* **20D**, 217 (1986); in *Nonlinear Phenomena and Chaos*, edited by S. Sarkar (Hilger, Bristol, 1986).
- <sup>19</sup>G. H. Golub and C. F. van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1983).
- <sup>20</sup>A. I. Mees, P. E. Rapp, and L. S. Jennings, *Phys. Rev. A* **36**, 340 (1987).
- <sup>21</sup>D. S. Broomhead, R. Jones, and G. P. King (unpublished).
- <sup>22</sup>A. I. Mees and P. E. Rapp (unpublished).
- <sup>23</sup>J. G. Caputo, B. Malraison, and P. Atten, in Ref. 3, p. 180.
- <sup>24</sup>The terminology used here is in the spirit of a system of nomenclature discussed at the Conference on the Brain as a Dynamical System, Pecos River Ranch, New Mexico, 1987 (unpublished). It was hoped that a more or less uniform terminology would make it possible, at least for those who use nonlinear dynamics in the study of electrical signals from the brain, to communicate with each other the details of their work with minimal ambiguity. This nomenclature is still being discussed.
- <sup>25</sup>In most of the calculations presented here,  $J$  is chosen so as to get  $10^3$  equally spaced embedding vectors from data sets consisting of between  $5 \times 10^3$  and  $10 \times 10^3$  points, the limit on the number of embedding vectors being dictated mainly by the fact that most of the calculations were done on microcomputers.
- <sup>26</sup>J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
- <sup>27</sup>J. Theiler, *Phys. Rev. A* **34**, 2427 (1986).
- <sup>28</sup>P. Grassberger, *Nature* **323**, 609 (1986).
- <sup>29</sup>Subsequent analysis will require an orthogonal transformation of the embedding space. The Euclidean distance is chosen here to make the dimension calculation invariant under this transformation.
- <sup>30</sup>See, e.g., M. Schwartz, *Information, Transmission, Modulation, and Noise* (McGraw-Hill, New York, 1970), p. 375.
- <sup>31</sup>G. B. Golub and C. Reinsch, *Numer. Math.* **14**, 403 (1970).
- <sup>32</sup>H. Froehling, J. P. Crutchfield, D. Farmer, and N. H. Packard, *Physica* **3D**, 605 (1981).
- <sup>33</sup>A. M. Albano, L. Smilowitz, P. E. Rapp, G. C. de Guzman, and T. R. Bashore, in *The Physics of Phase Space*, edited by Y. S. Kim and W. W. Zachary (Springer, Berlin, 1987).
- <sup>34</sup>A. M. Frazer and H. L. Swinney, *Phys. Rev. A* **33**, 1135 (1986).
- <sup>35</sup>J. Martinerie, A. M. Albano, A. I. Mees, T. R. Bashore, and P. E. Rapp (unpublished).
- <sup>36</sup>V. I. Osledec, *Trans. Moscow Math. Soc.* **19**, 197 (1968); *Ya. B. Pesin, Russ. Math. Surveys* **2**, 55 (1977).
- <sup>37</sup>A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, *Physica* **16D**, 285 (1985); related investigations have also been presented by I. Shimada and T. Nagashima, *Prog. Theor. Phys.* **61**, 1605 (1979); G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, *Meccanica* **15**, 9 (1980); **15**, 21 (1980).
- <sup>38</sup>E. Wigner, *Phys. Rev.* **40**, 749 (1932); T. A. C. M. Claassen and W. F. G. Mecklenbrauker, *Phillips J. Res.* **35**, 217 (1980); **35**, 276 (1980); **35**, 372 (1980).
- <sup>39</sup>T. C. Halsey *et al.*, *Phys. Rev. A* **33**, 1141 (1986).