Attractor dimension of nonstationary dynamical systems from small data sets

James W. Havstad
2111 Mill Creek Drive, Prospect, Oregon 97536

Cindy L. Ehlers
Research Institute of Scripps Clinic, La Jolla, California 92037
(Received 20 May 1987; revised manuscript received 20 June 1988)

Several sources of error in the calculation of attractor dimension from time series by the Grassberger and Procaccia or similar algorithms can be avoided or minimized by appropriate choice of algorithm and of criteria for selection of parameter values. Some problems and means of avoiding them are demonstrated, and statistical results are presented for dimension calculated from many independent sets of each of a wide range of data set size, to show that sets of a few hundred vectors or even less are useful for dimension calculation. This method with small data sets is shown to be effective in following changes in attractor dimension of a nonstationary dynamical system.

INTRODUCTION

Widely used and efficient methods for calculation of attractor dimension from experimental time series[1-3] may give results which depend strongly on details of the way in which the method is used.[4-6] We have particularly noted this problem in calculation of dimension of the attractor of the dynamical system producing the electroencephalogram (EEG).[7-9] Much of the uncertainty of results for this system may be due to nonstationarity of the system, so that consistent results may require use of small data sets, covering only short time intervals during which the system may be approximately stationary.

Several sources of error in dimension calculations will be described, together with methods for reduction or elimination of these errors, by suitable choice of algorithm and of values of parameters needed for the algorithm. Results are illustrated for a stationary mathematical dynamical system having an attractor information dimension of 7.5, which is sufficient to show some problems which may not be evident with systems of lower dimension. Statistical results from many independent data sets of each of many sizes show the variability to be expected from small data sets. These results show that useful estimates of dimension can be obtained from small data sets of a few hundred vectors or even less. This is consistent with other reports using very small sets.[10,11] Dimension calculations from many small successive but overlapping sets allow dimension to be studied as a function of time, for nonstationary dynamical systems in which the system is approximately stationary over the time interval of small sets.

METHODS

Problems and solutions are illustrated with a time series generated by the Mackey-Glass equation[12] with delay parameter τ,

$$\frac{dx(t)}{dt}=\frac{0.2x(t-\tau)}{1+x(t-\tau)^{10}}-0.1x(t) \tag{1}$$

which we integrate by a 1000-dimensional set of difference equations[1] of the form

$$x(t+T)=\frac{2-0.1T}{2+0.1T}x(t)+\frac{0.2T}{2+0.1T}\left[\frac{x(t-T)}{1+x(t-T)^{10}}+\frac{x(t+T-\tau)}{1+x(t+T-\tau)^{10}}\right], \tag{2}$$

where period T is the delay parameter τ divided by the number of dimensions of the difference equation set. Except where noted, we use $\tau=100$ and $T=0.1$. Random numbers are used to initialize the system, and Eq. (2) is iterated 200 000 times to allow the system to settle onto the attractor. The time series from continued iteration is normalized to zero mean and standard deviation of 512, and rounded to the nearest integer for consistency with experimental data digitized to 12 bits.

The phase-space trajectory for this Mackey-Glass system is formed from vectors X_i constructed from the scalar time series x_{jT} by the method of lags,[13,14]

$$X_i=\{x_{iaT},x_{iaT+bT},\ldots,x_{iaT+(n-1)bT}\}, \tag{3}$$

where n is the embedding dimension. The time interval aT between the first elements of successive vectors and the lag bT between vector elements are independent, and neither is necessarily equal to the sample period.

Each data set for dimension calculation consists of N vectors and M reference vectors[5] formed from the time series, with M not greater than N and less than N when N is large.

Statistical results from many small sets of the same size, from the same time series, and for a range of set sizes, will be shown. Differences in results may have two

sources: small sets may show differences which are directly due to the small number of vectors, regardless of how the system is sampled, but small sets may also show differences which are the result of sampling over a short time interval which does not adequately characterize the attractor. Two types of sampling are used in order to estimate the contribution of each of these sources of variability. Sequential sets use the same interval between vectors, regardless of the size of the set, so that the time interval sampled by a set is proportional to the number of vectors in the set. Interleaved sets of various sizes use greater intervals between vectors for small sets, so that each set, regardless of number of vectors, spans substantially the same time interval.

The major results of this paper will use an algorithm which calculates information dimension, but, in order to illustrate problems and their solutions, the correlation dimension of Grassberger and Procaccia will serve as the starting point. For each value of embedding dimension n from 1 to 32 and for each of a predetermined set of values of distance r, the distance from a reference vector to each of N vectors is calculated, and the number $j(r,n)$ of vectors is counted for which this distance is not greater than r. The maximum norm is conventionally used for distance so that the vectors counted for a particular reference vector are those within an n-dimensional cube of radius r centered on the reference vector.

For data with finite resolution and limited numbers of vectors, the effective correlation dimension at embedding dimension n is

$$\hat{D}_C(n)=\frac{\Delta\ln[\langle j(r,n)\rangle]}{\Delta\ln(r)}\ , \tag{4}$$

where the angle brackets indicate averaging over all reference vectors. Here radius r is the independent variable, and the dependent variable $j(r,n)$ is calculated at predetermined values of r for each value of embedding dimension n. The Δ symbols indicate the slope of a log-log plot of $\langle j(r,n)\rangle$ and r, estimated by a least-squared error fit of a straight line over a selected range of r. The form of this equation is chosen for consistency with equations introduced below, rather than in terms of the integral correlation coefficient as originally given by Grassberger and Procaccia.[1] The caret indicates an "effective" correlation dimension, which may be taken to be an estimate of correlation dimension D_C if scaling and saturation are satisfactory at embedding dimension n.

The ideal log-log plot of $\langle j(r,n)\rangle$ and r is a straight line over a wide range of $\langle j(r,n)r\rangle$, reflecting the scaling property of the attractor, so that the slope of the curve for each value of embedding dimension is well defined. This approximately straight region will be referred to as the scaling region. In addition, ideally, the effective dimension increases with increasing embedding dimension until it saturates at a constant value with further increase in embedding dimension. This saturation value of effective dimension is taken to be the dimension of the attractor, and the region of saturation will be referred to as the saturation region. The success of variations in methods may be judged by the quality of the scaling and saturation regions.

Calculation of correlation dimension with Eq. (4), and using values of aT and bT of Eq. (3) such as are commonly used, may result in several problems in approaching the ideal of long scaling regions and saturation of effective dimension, as shown in Fig. 1(a). The scaling regions are of poor quality, apparent saturation results only from an arbitrary choice of scaling region, and dimension is about 40% less than the correct value of 7.5.[1] The following paragraphs describe variations of dimension definition, method of calculation, and choice of parameter values, which improve the accuracy and stability of results.

(1) *Increased time interval between reference vectors and neighboring vectors.* The prominent distortion of slope of the curves of Fig. 1(a) at large n and near $j=2$ is moved relatively downward by increasing the size of the data set, leaving a longer scaling region and contributing perhaps to the general belief that large N is essential. A better solution is to remove the cause of the distortion, without increasing N, by decreasing the correlation between reference vectors and their neighbor vectors. This can be done by increasing aT of Eq. (1),[5] but this may reduce unnecessarily the effective size of the data set. We use a method also described by Theiler[4] which maintains a minimum temporal spacing greater than aT between reference vectors and their nearest neighbors, by deleting a number W of vectors on each side of the reference vector in use. The results of skipping four vectors on either side of the reference vector in use are shown in Fig. 1(b). Scaling regions are improved though not yet satisfactory, and the effective correlation dimension is increased but is still highly dependent on the choice of region for slope measurement. The apparent excellent saturation of effective dimension in Fig. 1(a) is seen to result from the distortion which is absent in Fig. 1(b).

(2) *Radius as dependent variable.* In the method of Eq. (4), each of the curves for a single data set is the average of M curves, each of which results from distances from a single reference vector to N vectors. The points averaged together are the values of $j(r,n)$ at each of the predetermined values of the independent variable r. This can be thought of as vertical averaging, because the points averaged are the intersections of the unaveraged curves with a vertical line through the specified value of r. Holzfuss and Mayer-Kress[5] have shown that unaveraged curves may vary greatly in position along the $\ln(r)$ axis so that points lying in the scaling region of some curves may be averaged with points outside the scaling region of other curves, resulting generally in reduction of slope of the averaged curve. Short scaling regions, from small data sets, are particularly distorted by this problem.

To solve this problem we previously used an alignment method in which each unaveraged curve is shifted along the radius axis, without change of slope, so that all unaveraged curves pass through a common point in their scaling regions. The curves are then averaged vertically as before. We now prefer horizontal averaging along lines of constant j, which has equivalent benefits in improved averaging, and leads to other benefits as well. This is accomplished by exchanging variables so that j is the independent variable and $r(j,n)$ is the dependent

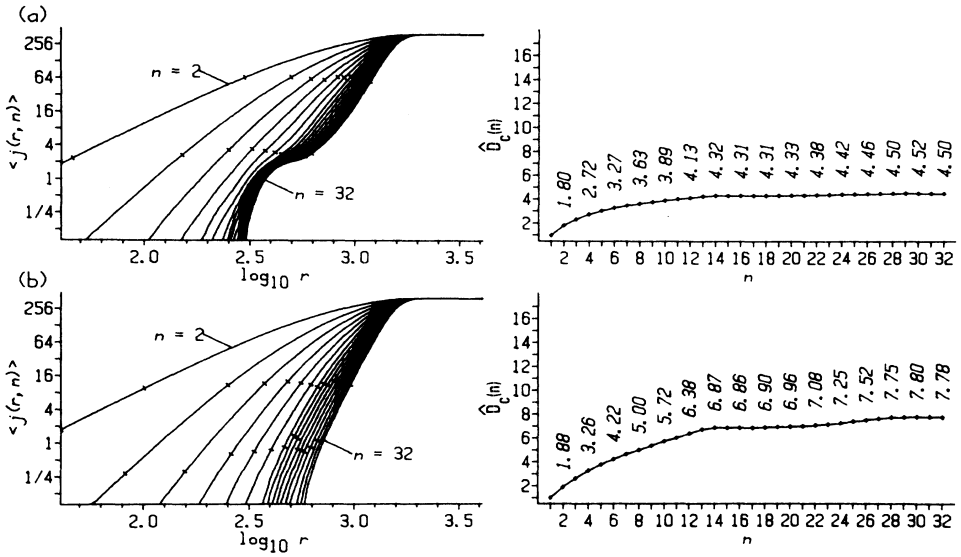

FIG. 1. Problems with calculation of correlation dimension of attractor of Mackey-Glass equation with delay of 100. Means of 32 sets with $N=400$, $M=200$, $aT=3.75$, $bT=7.5$, $W=0$ in (a), $W=4$ in (b). Effective correlation dimension $\hat{D}_C(n)$ is the slope of the $\ln\langle j(r,n)\rangle$ vs $\ln(r)$ curve for the corresponding value of embedding dimension n, calculated by least-squares fit of a straight line to the section of the curve between the short bars. The major fault in (a) is the distortion of curves at high embedding dimension near $j=2$, due to highly correlated neighbor vectors about each reference vector, from $W=0$. This fault is absent in (b) because $W=4$, but other faults remain.

variable. The problem is then to find the radius of the smallest n-dimensional sphere, about each reference vector, which contains exactly j vectors, for each of a predetermined set of values of j. In this respect this method is that of Termonia and Alexandrowicz,[2] but we follow Grassberger[3] in using different definitions of dimension.

Effective correlation dimension with radius as the dependent variable is

$$\hat{D}_C(n)=\frac{\Delta\ln(j)}{\Delta\ln[\langle r(j,n)\rangle]}\ , \tag{5}$$

where the angle brackets indicate averaging over all reference vectors. Euclidean norm is used for this and subsequent calculations, for reasons discussed below.

(3) *Averaging of log of radius.* Averaging $\ln[r(j,n)]$ rather than $r(j,n)$, that is, using $\langle\ln[r(j,n)]\rangle$ rather than $\ln\langle r(j,n)\rangle$, results in effective information dimension[3]

$$\hat{D}_I(n)=\frac{\Delta\ln(j)}{\Delta\langle\ln[r(j,n)]\rangle}\ . \tag{6}$$

Information dimension may be more interesting than other similar dimensions but our main reason for adopting it is that this change, in combination with others, leads to some computational convenience.

(4) *Independence of vectors.* Figures 2(a) and 2(b) show

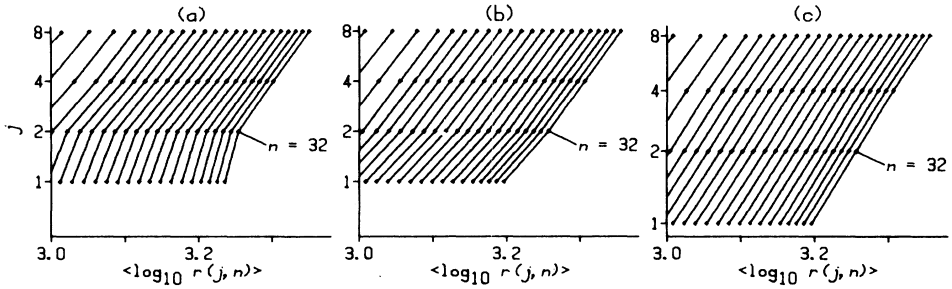

FIG. 2. Two problems with slopes at small j, for Mackey-Glass equation with delay of 100, and with parameter values as in Fig. 1(b) except as noted. Scales are expanded to show problems at small j and high embedding dimension. (a) Equation (6) and Euclidean norm are used with $aT=3.75$. Small value of aT results in excess correlation of vectors and excess slope of curves, partially reduced by lack of Poisson correction. (b) As in (a) but with $aT=15$, reducing correlation of vectors, but slopes are now too small because of lack of Poisson correction. (c) Equation (7) with Poisson correction and Euclidean norm are used. Compared to $\ln(j)$ in (b), $\psi(j)$ in (c) stretches the j axis at small j, extending the scaling region to $j=1$.

that varying the time interval between vectors changes the slope of dimension curves at small scales. The difference in mean $\ln[r(j,n)]$ at $j=1$ and $j=2$ is abnormally small in Fig. 2(a) because the closeness of vectors in time results in an increased probability that they are close in phase space. Increasing the time interval between vectors from half of the lag in Fig. 2(a) to twice the lag in Fig. 2(b) avoids the excess slope at small scales. Vectors must be sufficiently independent, by being sufficiently separated in time, if the scaling region is to extend to the smallest scales. This is distinct from the problem of independence of reference vectors with respect to their neighbor vectors, which has the opposite effect on slope.

The need for independence of vectors, and for even greater independence between reference vectors and neighbor vectors, is in contrast with the common practice of using sample rates greater than the Nyquist frequency, such as would be needed for spectral analysis or for phase portraits, and using the sample period as the interval between vectors. We believe that such close spacing of vectors in time is not only unnecessary for dimension calculation, but is a source of error.

(5) *Correction for Poisson distribution.* Grassberger[3] has shown that numbers of vectors j within a specific volume have a Poisson distribution which requires a modification of Eq. (6),

$$\hat{D}_I(n)=\frac{\Delta\psi(j)}{\Delta\langle \ln[r(j,n)]\rangle}\ , \tag{7}$$

where the digamma function $\psi(x)=d\ln\Gamma(x)/dx$ has the effect of stretching the j axis at small j, thereby increasing slope, as shown in Fig. 2(c). The scaling region extends to the smallest values of j when vectors are independent and the Poisson correction is used, resulting in an increase in effective information dimension.

(6) *Optimum lag.* Lag can be chosen almost arbitrarily if infinite amounts of noise-free data are used[14] but in practice the choice of lag, bT in Eq. (1), is found to be critical.

The effect on effective information dimension saturation of various values of lag is shown in Figs. 3(a), 3(b), and 3(c) for data sets of moderate size, for small interleaved sets, and for small sequential sets, respectively. For small values of lag the effective dimension increases slowly with increasing embedding dimension and reaches saturation only at high embedding dimension, if at all. At large values of lag, the effective dimension increases rapidly with increasing embedding dimension and does not saturate at the attractor dimension. Saturation is found only for lags near 7.5.

Figure 3 indicates that 12 800 vectors does not approach the number at which choice of lag is not critical. Lag is even more critical for small data sets, and more so for sequential than for interleaved sets. The most evident problem for small sets is an inability of a small number of vectors, drawn from a short time series, to represent the fractal structure of the attractor at high embedding dimension, as seen by the failure of saturation to extend to high embedding dimension for sequential sets.

We estimate lag from inspection of the power spectrum or phase portraits, and refine this estimate by testing larger and smaller lags. Familiarity with the shapes of saturation curves for lags greater than or less than optimum leads rapidly to optimum lag. We have not used mutual information[15] for estimation of optimum lag because it is said to require very large data sets unless dimension is very low, with noisy data producing additional difficulties.[16]

(7) *Euclidean norm.* Maximum norm for distance between vectors and reference vectors results in counting vectors within cubes, whereas Euclidean norm results in counting vectors within spheres. Representative differences in dimension curves for maximum and Eu-

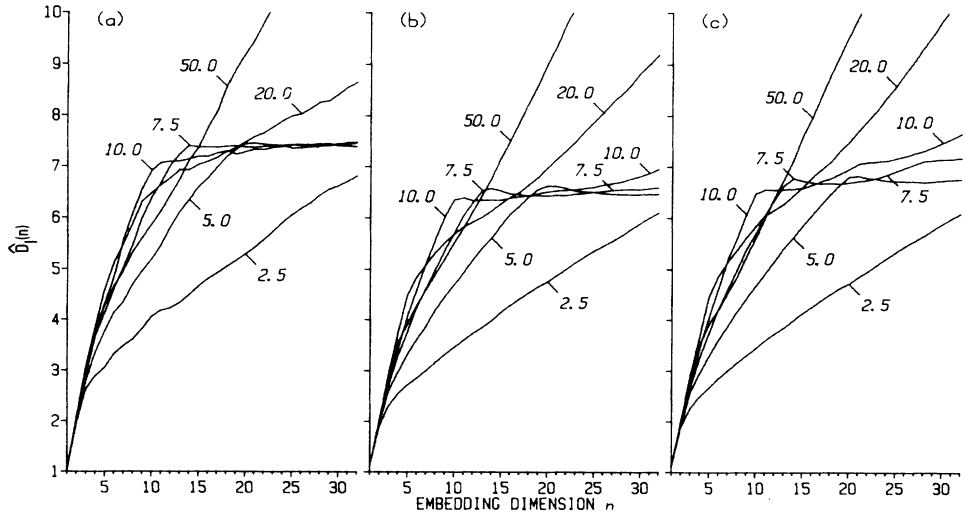

FIG. 3. Effect of value of lag on effective information dimension as a function of embedding dimension, for the attractor of the Mackey-Glass equation with delay of 100, using Eq. (7) and Euclidean norm, with $aT=15$ and lag bT as shown, $W=1$, $M=200$. (a) One set with $N=12\,800$. (b) Means of 64 interleaved sets with $N=200$. (c) Means of 64 sequential sets with $N=200$. Choice of lag is critical for all set sizes, but especially so for small sequential sets.

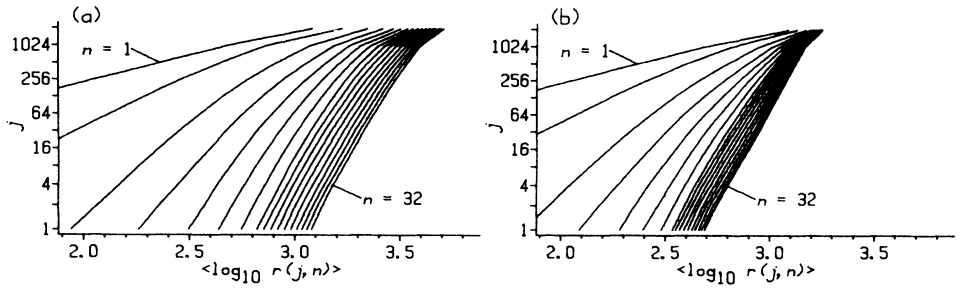

FIG. 4. Comparison of Euclidean norm (a) and maximum norm (b) for calculation of information dimension, for the attractor of the Mackey-Glass equation with delay of 100, using Eq. (7), $aT=15$, $bT=7.5$, $N=2200$, $M=200$, and $W=1$, for embedding dimension of one and even values to 32.

clidean norms are shown in Fig. 4. Maximum norm tends to underestimate dimension.

Successive points in the reconstruction of the trajectory in phase space are significantly correlated when optimum values of lag are used. The attractor tends therefore to be clustered about the positive diagonal of the phase space. We believe that the distortions of dimension curves with maximum norm result from the tendency of the diagonals of cubes to be aligned with surfaces of the attractor. The spheres of Euclidean norm show no such selectivity.

EFFECT OF DATA SET SIZE ON DIMENSION

Information dimension of the attractor of the Mackey-Glass equation with delay of 100, using Eq. (7) and Euclidean norm, is shown in Fig. 5(a) for a moderately large set of $N=12\,800$, and in Fig. 5(b) for 64 small sequential sets of $N=400$. Mean effective dimension for the small sets is about 5% less than dimension for the large set. The scaling regions used for dimension calculation are shown by the dots on the curves, and are taken to be the best region for each case. The region for small

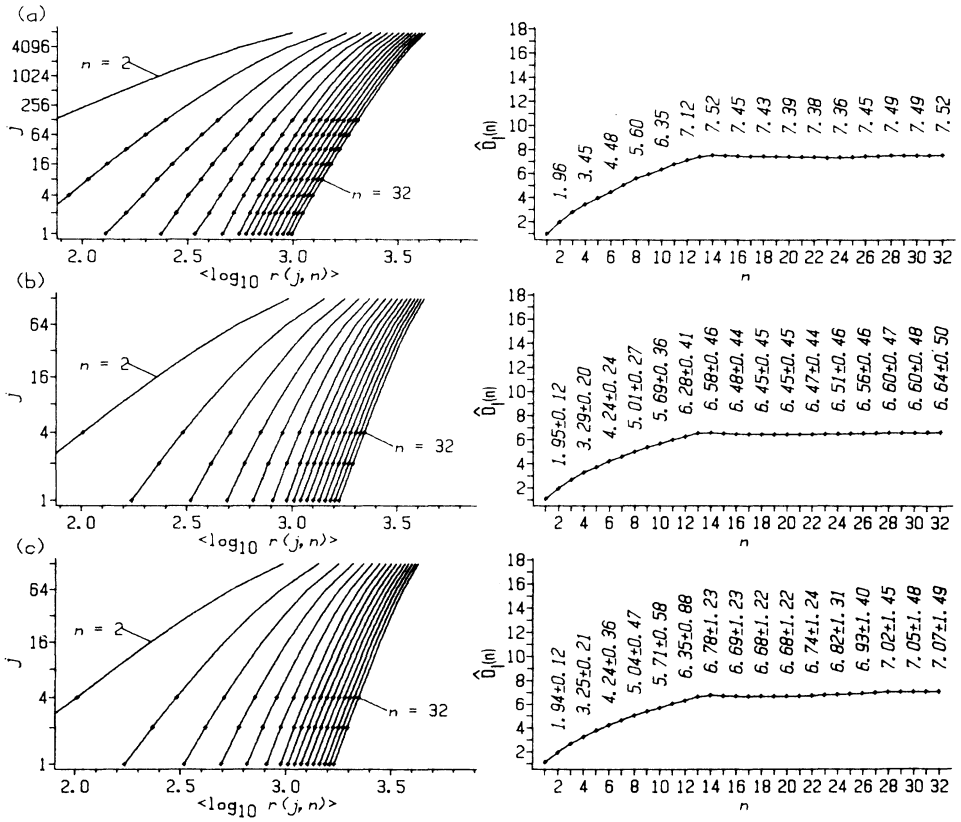

FIG. 5. Information dimension for large and small data sets, for the attractor of the Mackey-Glass equation with delay of 100, using Eq. (7) and Euclidean norm. (a) One set with $N=12\,800$, $M=200$, $aT=15$, $bT=7.5$, $W=1$. (b) As in (a) but means of 64 interleaved sets with $N=200$, $aT=960$. (c) As in (b) but with sequential sets, $aT=15$. Standard deviations of information dimension for sequential small sets are much greater than for interleaved sets of the same size.

sets has smaller slope, resulting in lower dimension, because, at the smallest scale available with small data sets, the scale is not small enough to be free of effects of the finite size of the attractor.

The result of about 7.5 for the large set agrees with the correlation dimension value of Grassberger and Procaccia[1] but it is unclear how their assertion that the effective dimension saturated at embedding dimension of 16 could result from lags of 50 and 100. Termonia and Alexandrowicz[2] report dimension of 12.6 at embedding dimension of 30, with lag of 100. This is consistent with our effective dimension values at this lag, but because no saturation is found at that lag this value cannot be taken as an estimate of attractor dimension. We find the Mackey-Glass equation attractor to be sufficiently uniform so there is little difference between correlation and information dimensions.

Table I summarizes results for various data set sizes, from $N=50$ to $N=12\,800$, using sequential and interleaved sets. The scaling regions stated in the caption of the table are not chosen to yield the best estimates of dimension, but use the same range of j/N, so far as possible, in order to determine if differences with respect to set size shown in Fig. 5 are due directly to set size or to differences in choice of scaling regions. It is seen that there is little difference in mean dimension of large and small sets when the same range of scale is used for slope calculation, although this choice of scaling region, to be useful for the smallest set sizes, results in underestimation of dimension by about 10%. There is also little difference in mean dimension for sequential and interleaved sets, but the standard deviation of dimension for sequential sets is about 3.2 times that of the interleaved sets.

Some of the variance of dimension is assumed to be due directly to the number of vectors, independent of how they are distributed in time, and the rest of the variance is assumed to result from the way the vectors are distributed in time, independent of their number. Interleaved sets of various sizes sample the same length of time series so that only the former component of variance contributes significantly to the variance of these sets. The variance of sequential sets is the sum of both components. Applying this linear model to the data of Table I indicates that about 90% of the total variance for small sequential sets is due to variations in short sequential time series, and only about 10% of the variance is due directly to the use of small numbers of vectors. This supports the view that a few well-selected vectors may be superior to a large number of poorly selected vectors.

Slope of mean effective information dimension as a function of embedding dimension is not greater than 0.01 or less than -0.02 for all sets of Table I over the range $n=14$ to 18, indicating good saturation, and the mean effective information dimension over this range is taken to be a good estimate of information dimension. For very small sequential sets, effective information dimension increases beyond the saturation region, as in Fig. 3(c) for $bT=7.5$. This is not seen in Fig. 3(b) for interleaved sets, suggesting that the short saturation region for small sequential sets is not due to the small number of vectors but to the short time interval from which they are taken.

The standard deviations in Table I are calculated from the distributions of dimension for many sets of the same size, using the same method. They are not necessarily estimates of error because they do not include any bias resulting from the method. Error, including bias, can be estimated by comparison of means of Table I with the best estimates from large data sets, as described above. Mean dimension estimates in Table I are less than the best estimates from large sets because of the decision to use as nearly as possible the same scaling regions for all set sizes, in order to isolate sources of variability. Bias in estimates of dimension for small data sets is thereby shown to be due primarily to short scaling regions, where slopes are decreased by effects of the finite size of the attractor.

Standard deviation of slope estimates from least-squared error fitting of a straight line are much less than the standard deviations of Table I and very much less than error estimates. The error estimates of Holzfuss and Mayer-Kress[5] are much larger than our experimentally determined values.

TABLE I. Effect of size of data sets on mean information dimension of the attractor of the Mackey-Glass equation with delay of 100, for sequential and interleaved sets. Means and standard deviations are calculated for many sets of each size, except for the largest set sizes. $M=200$ except where N is less than 200, $aT=15$, $bT=7.5$, $W=2$, using Eq. (7) with Euclidean norm. For comparison over corresponding radius scales, slope is calculated over the same range of j/N, from 0.02 to 0.001 25, except that the lower limit is $1/N$ for N less than 800 and the upper limit is $4/N$ for N less than 200.

Data set Size N	Number of sets	Information dimension D_I Sequential sets	Interleaved sets
50	256	6.44±1.74	6.17±0.58
100	128	6.68±1.49	6.48±0.46
200	64	6.71±1.22	6.53±0.41
400	32	6.77±0.50	6.78±0.23
800	16	6.77±0.45	6.98±0.11
1600	8	6.82±0.27	6.72±0.05
3200	4	6.71±0.10	6.70±0.04
6400	2	6.66	6.73
12 800	1	6.69	

MINIMUM DATA SET SIZE

Mayer-Kress[17] has derived a formula for the minimum number of vectors needed for dimension calculation, in support of the assertion that large data sets are required, though no numerical results are provided in support of the formula. The central idea is that many loops of the attractor must be sampled to adequately represent its structure. Because this idea seems reasonable, but the conclusions contradict our results for small sets, this problem is examined in some detail.

The formula, Eq. (1) of Ref. 17 with some rearrangement and change of symbols, is

$$N = S^D s^{1-D} x \ , \tag{8}$$

where N is the number of vectors, S is the diameter of the attractor, s is the mean distance between loops of the attractor, D is the dimension of the attractor, and x is the mean distance between successive vectors. The derivation requires that vectors be very closely spaced in time so that the length of the trajectory is Nx. The equation is also said to relate data set size to sample rate, but sample period is proportional to vector distance x only if vectors are closely spaced in time. It is asserted that "optimum sampling" results from $s = x$, and N is then the minimum number of vectors required to meet this condition,

$$N_{\min} = (S/x)^D \ . \tag{9}$$

We have devised a method for calculation of the quantities of Eqs. (8) and (9) from experimental data, and have applied it to a Mackey-Glass time series with delay of 100, normalized to zero mean and standard deviation of 512. Mean distance x between successive vectors is calculated by summing the Euclidean distances between N successive vectors in n dimensions and dividing by $N-1$. Mean distance between trajectory loops s is calculated by counting the intersections between the attractor and a hyperplane through the origin and perpendicular to the nth coordinate axis. The plane of $n-1$ dimensions is partitioned into annular segments to allow determination of the diameter of the smallest hypersphere containing all of the intersections. Because the attractor is not in general spherically symmetrical, the diameter is not in general the difference between extreme values of the time series. The mean volume per intersection is the volume of the sphere divided by the number of intersections contained in it. The mean distance between loops of the attractor is taken to be the diameter of a sphere of $n-1$ dimensions whose volume is equal to the mean volume per intersection.

The quantities calculated from Eqs. (8) and (9) are first compared with those calculated from the Mackey-Glass time series when the data set is formed in accordance with the assumption of Eq. (8). Using a large data set of $N = 32\,000$ with lag $bT = 8$, and using $aT = 1$ so that vectors are closely spaced in time and therefore not independent, it is found at embedding dimension $n = 24$ that $x = 243$, $s = 2956$, and $S = 7962$. These values and Eq. (8) result in set size $N = 21\,000$. This is of the same order as the correct value $N = 32\,000$, but x is much smaller than s so the conditions for "optimum sampling" are not met. Since Eq. (8) requires short intervals between vectors, x cannot be increased by increasing these intervals, and s must be decreased by increasing N if the requirement of optimum sampling is to be met. Setting s equal to the calculated values of x, Eq. (9) results in a minimum data set size in excess of 10^{10} vectors. This unreasonable result is not due to differences in the way quantities are defined in the derivation and the way they are calculated from experimental data. The number of intersections with the hyperplane is 945, for example, compared to 977 calculated as in the derivation from Nx/S.

To show that the assumption of closely spaced vectors is essential to Eq. (8), $aT = 16$ is chosen so that vectors are independent. The results at $n = 24$ are $x = 2463$, $s = 2973$, and $S = 8934$, from which Eq. (8) yields a set size $N = 4600$. Agreement with $N = 32\,000$ is less satisfactory, indicating that the assumption of very short time intervals between vectors is essential to Eq. (8).

Similar results are found for other data set sizes. We conclude that the ideas behind these equations of Mayer-Kress provide valuable insight into the problem of minimum data set size, but the equations assume vectors very closely spaced in time, which conflicts with the need for independent vectors, and also conflicts with the questionable assumption regarding optimum sampling so that estimates of minimum data set size are unreasonably large by many orders of magnitude. The arguments of Mayer-Kress against small sets do not appear to overcome our results supporting their use.

Results from the preceding sections indicate three consequences of decreasing data set size: Standard deviation increases, underestimation of dimension increases, and the length of the saturation region decreases. The number of sets available for analysis, the stationarity of the system, the characteristics of the particular attractor, and the use to be made of the results will determine which of these is the limiting factor in determining optimum set size.

DIMENSION AS A FUNCTION OF TIME FOR NONSTATIONARY SYSTEMS

A time series constructed of a repeated sequence of equal length segments of Gaussian white noise, of Mackey-Glass equation with delay of 100, and of Mackey-Glass equation with delay of 30, is used to demonstrate the ability of small sets to follow dimension changes in a nonstationary dynamical system. The information dimensions of these systems are infinite, 7.5, and 3.0, respectively. Each segment is taken from a stationary time series so that transitions are artificially abrupt. Figure 6 shows this time series, and, as functions of time, its power spectra, effective information dimension, and information dimension. Dimension calculations use 200 vectors for each sequential set, and the Mackey-Glass delay of 100 arbitrarily equals 97 msec, and delay of 30 equals 58 msec. The analysis window for spectra and dimension is 3.14 sec, and the window advances 0.5 sec between analysis epochs. Effective information dimension

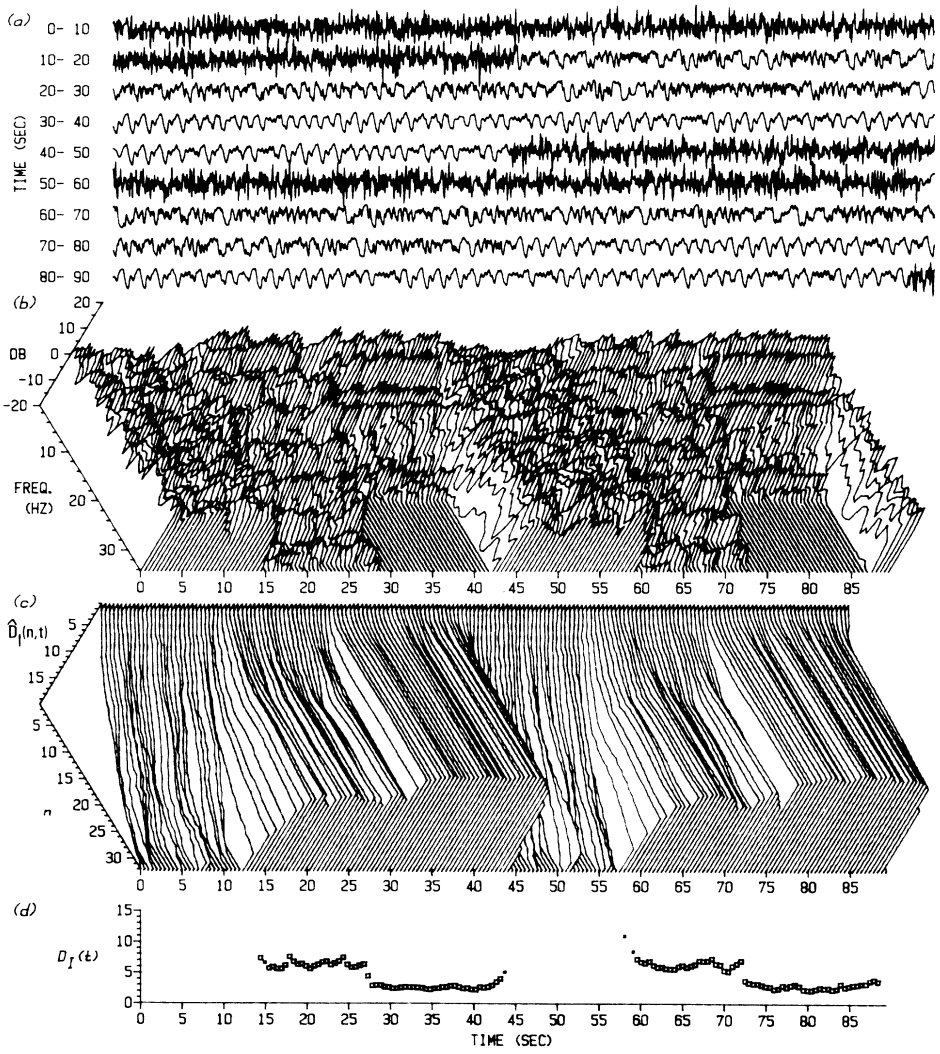

FIG. 6. (a) Nonstationary time series consisting of a repeated sequence of 15 sec each of Gaussian white noise, Mackey-Glass equation with delay of 100, and Mackey-Glass equation with delay of 30. (b) Power spectra with decibel reference equal to mean power spectral density for all spectra. (c) Effective information dimension $\hat{D}_I(n)$ calculated with Eq. (7) and Euclidean norm with $N=200$, $M=200$, $aT=14.6$ msec, $bT=7.3$ msec, $W=2$. (d) Information dimension from saturation curves of (c).

curves are inverted, with increasing dimension downward, so that curves with good saturation are unlikely to be masked by those without saturation. The slope of each saturation curve is the slope of the segment spanning an embedding dimension range of eight whose slope is nearest zero, and the mean effective information dimension over this segment is taken as an estimate of information dimension, provided, in this case, that the absolute value of the slope is not greater than 0.05.

Figure 6 shows that dimension of a nonstationary system can be calculated from small overlapping sets with reasonable accuracy, when the dimension is not greater than at least 10, and abrupt changes of dimension produce corresponding changes in calculated estimates of dimension with a delay and rise time of about half the time duration of each set.

CONCLUSIONS

This study indicates that significant sources of error in estimation of dimension can be avoided by (1) careful choice of intervals between vectors and between reference vectors and neighbor vectors so that vectors are adequately independent, (2) by use of optimum lag, (3) by use of $\langle \ln[r(j,n)] \rangle$ as the averaged dependent variable and $\psi(j)$ as the Poisson corrected form of the independent variable, and (4) by use of Euclidean norm. Mean dimension is then found to be independent of set size when the same range of j/N is used for slope calculation, and sets of as few as 50 vectors, for a 7.5-dimensional attractor, may be useful. The largest source of variability of dimension from small sets sampled from short time series segments is not the small number of vectors, but the short

time interval spanned by these vectors. While dimension calculated from large data sets of well-chosen vectors is more accurate for stationary dynamical systems, results from small data sets may be acceptable when only small data sets are available, and such sets may be preferred for dynamical systems which are not stationary. For such systems, small sets can follow changes in dimension with reasonable accuracy.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Gottfried Mayer-Kress, Peter Grassberger, Joachim Holzfuss, and Alan Garfinkel. This work has been supported in part by grant Nos. 06059 and 00098 from the National Institute of Alcoholism and Alcohol Abuse, GCRC Grant No. RR00833, and a grant from the John and Catherine Macarthur Foundation.

[1]Peter Grassberger and Itamar Procaccia, Physica D **9**, 189 (1983).

[2]Yves Termonia and Zeev Alexandrowicz, Phys. Rev. Lett. **51**, 1265 (1983).

[3]Peter Grassberger, Phys. Lett. **107A**, 101 (1985).

[4]James Theiler, Phys. Rev. A **34**, 2427 (1986).

[5]J. Holzfuss and G. Mayer-Kress, in *Dimensions and Entropies in Chaotic Systems* Vol. 32 of *Springer Series in Synergetics,* edited by G. Mayer-Kress (Springer-Verlag, New York, 1986).

[6]J. G. Caputo, D. Malraison, and P. Atten, in Ref. 5, p. 180.

[7]A. Babloyantz, in Ref. 5, p. 241.

[8]S. P. Layne, G. Mayer-Kress, and J. Holzfuss, in Ref. 5, p. 246.

[9]Gottfried Mayer-Kress and Scott P. Layne, Ann. N.Y. Acad. Sci. **504**, 62 (1987).

[10]N. B. Abraham, A. M. Albano, R. Das, G. DeGuzman, S. Yong, R. S. Gioggia, G. P. Puccioni, and J. R. Tredicce, Phys. Lett. **114A**, 217 (1986).

[11]Peter Grassberger, Nature **323**, 609 (1986).

[12]M. C. Mackey and L. Glass, Science **197**, 287 (1977).

[13]Harold Froehling, J. P. Crutchfield, Doyne Farmer, N. H. Packard, and Rob Shaw, Physica **3D**, 605 (1981).

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

[15]Andrew M. Fraser and Henry L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[16]A. M. Fraser, in Ref. 5, p. 82.

[17]Gottfried Mayer-Kress, in *Drections in Chaos,* Vol. 3 of *World Scientific Series on Directions in Condensed Matter Physics,* edited by Hao Bai-lin (World Scientific, Singapore, 1987).