Spurious dimension from correlation algorithms applied to limited time-series data

James Theiler
Department of Physics, California Institute of Technology, Pasadena, California 91125
(Received 7 May 1986)

An algorithm devised for measuring the dimension of a strange attractor from a time series is applied both to autocorrelated Gaussian noise and to a dynamical system. It is analytically shown that a finite sequence of stochastic data—where by "finite" it is meant that $N < 2\tau^{m/2}$, where N is the number of points in the sequence, τ is the autocorrelation time (in units of sampling period), and m is the embedding dimension—exhibits anomalous structure in its correlation integral. The anomaly is seen numerically in both stochastic and dynamical data. Unrecognized, it can lead to unnecessarily inaccurate and possibly spurious estimates of dimension. We propose a slight modification of the standard algorithm which eliminates this difficulty.

I. INTRODUCTION

Since the work of Lorenz,[1] it has become widely known that simple deterministic systems with only a few degrees of freedom can display complicated aperiodic behavior. The experimentalist, confronted with complex motion, seeks to determine whether that motion is "chaotic," deriving from just such a simple system, or whether it is "stochastic," deriving from a system which itself is complicated and which possesses many degrees of freedom.

Methods proposed[2] for the determination of the dimension of a system's trajectory through phase space provide a means of distinguishing these two kinds of motion. An algorithm of Grassberger and Procaccia[3] involves the computation of a "correlation integral" whose power-law behavior is used to estimate the dimension of the attractor. This dimension is thought to be a measure of the number of "active modes" in the system, or of the "effective number of degrees of freedom."

We examine the structure of the correlation integral in two representative cases: for stochastic and for dynamical data. We find in both cases that for limited data sets with high autocorrelation, the correlation integral displays an anomalous "shoulder" which inhibits good estimates of dimension. For stochastic data, we can model this effect analytically. The model provides a criterion from which one can predict whether the effect will be important in a given system; also, by pointing to the source of the anomaly, the model provides a quick and easily implemented fix. We demonstrate the effect and its fix numerically for both stochastic and dynamical data.

II. CORRELATION INTEGRAL

The experimentalist measures at regular and discrete intervals of time the value of some state variable (such as temperature or density or voltage) and records: $x_1, x_2, \ldots, x_N$; where $x_i \in \mathbb{R}$ is the measurement taken at time $t_i = t_0 + i\,\Delta t$.

Takens[4] suggests creating out of this sequence of one-dimensional variables a set of m-dimensional vectors whose components are just the time-delayed values of the variables:

$$\mathbf{v}_i = (x_i, x_{i+T}, x_{i+2T}, \ldots, x_{i+(m-1)T}), \quad \mathbf{v}_i \in \mathbb{R}^m . \tag{1}$$

Note that (for $T, m \ll N$ which is typical) there are (almost) as many vectors $\mathbf{v}_i$ as there are data points x_i. The dynamical information in the one-dimensional data has been "converted" to spatial information in the m-dimensional set. Takens has shown that for sufficiently large m and with "certain generic conditions" on the measurement, a system which has a ν-dimensional attractor in its phase space will have its Takens vectors lying on a ν-dimensional subset of the embedding space $\mathbb{R}^m$.

To find this ν, Grassberger and Procaccia[3] introduce the following correlation integral:

$$C(r,N) \equiv \frac{1}{N^2} \times [\text{the number of pairs } (i,j) \text{ for which } |\mathbf{v}_i - \mathbf{v}_j| \le r] . \tag{2}$$

In anticipation of the modification that will be proposed, we write an equivalent form:

$$C(r,N) \equiv \frac{2}{N^2} \sum_{n=1}^{N} \sum_{i=1}^{N-n} H(r - |\mathbf{v}_{i+n} - \mathbf{v}_i|) , \tag{3}$$

where $H(x)$ is the Heaviside step function. The correlation dimension ν is then defined by the limit:

$$\nu = \lim_{r \to 0} \lim_{N \to \infty} \frac{\log C(r,N)}{\log r} ; \tag{4}$$

or when it exists[5]

$$\nu = \lim_{r \to 0} \lim_{N \to \infty} \frac{d[\log C(r,N)]/dr}{d(\log r)/dr} . \tag{5}$$

With finite data, however, one must compromise. Neither

limit can actually be taken; N can be only as large as is available and r can be only as small as the least of the N^2 distances. What is usually (though not always[6]) done is this: A log-log plot of the $C(r,N)$ curve is constructed, and a slope is sought in the small r regime. One hopes for a broad range over which the slope is constant so that the value of that slope (which estimates ν) can be accurately determined.

It is the purpose of this paper to show that for autocorrelated data, the range of constant slope on the log-log plot may be *unduly* restricted if the number of data points, N, is limited. The restriction is not a fundamental one, however, and can be fixed without increasing N. This can be done, as we shall demonstrate, with a slightly generalized version of the correlation integral. Take

$$C(r,N,W)=\frac{2}{N^2}\sum_{n=W}^{N}\sum_{i=1}^{N-n}H(r-|\mathbf{v}_{i+n}-\mathbf{v}_i|) \tag{6}$$

keeping in mind that $W=1$ gives back the standard Grassberger-Procaccia formula defined in Eq. (3).

III. ANALYTICAL RESULTS: STOCHASTIC DATA

Consider a limited (finite N) data set of autocorrelated Gaussian noise.[7] Specify the mean $\mu=0$, the variance σ^2, and the autocorrelation α. That is, $\langle x_i\rangle=\mu=0$, $\langle x_i x_i\rangle=\sigma^2$, and $\langle x_{i+n}x_i\rangle/\sigma^2=\alpha^n$. Take $\alpha<1$ and note that the autocorrelation time (in units with $\Delta t=1$) is given by $\tau\equiv 1/\ln(1/\alpha)$ or $\tau\approx 1/(1-\alpha)$ for α near unity.

Because the data is stochastic we can use statistical methods to obtain an analytical expression for the correlation integral $C(r,N,W)$. We will see that the behavior of the correlation integral can be classified into two qualitatively distinct regimes. If N is large enough (or α small enough), then the effect of autocorrelation is negligible, the trajectory "fills out" the phase space, and the slope of the $\log C(r,N,W)$ versus $\log r$ curve approaches the embedding dimension m for $r\ll\sigma$. On the other hand, for N not sufficiently large (or for α too near unity), the effect of autocorrelation becomes noticeable and although the trajectory still fills out phase space the correlation integral is not so well behaved; a structure is induced in the correlation curve which inhibits good dimension estimates. And as we will show, the "sufficiently large" N which separates these two regimes can be quite unreasonably large.

It is not essential, but it makes the analysis easier[8] if we impose that the delay time T be much larger than the autocorrelation time τ. This allows us to treat x_i and x_{i+T} as independent and simplifies the expression for the probability that two vectors are separated by a distance less than r. We use the $L\infty$ metric (again for convenience[9]) which specifies that the distance between two vectors is the maximum of the differences between those vectors' components.

Rewrite Eq. (6) for the correlation integral by replacing the inner sum over Heaviside functions with its expectation value:

$$C(r,N,W)=\frac{2}{N^2}\sum_{n=W}^{N}(N-n)P(|\mathbf{v}_{i+n}-\mathbf{v}_i|\le r)\,, \tag{7}$$

where $P(X)$ denotes the probability that statement X is true. Now

$$P(|\mathbf{v}_{i+n}-\mathbf{v}_i|\le r)=P(|x_{i+n+kT}-x_{i+kT}|\le r \text{ for all } 0\le k<m) \tag{8}$$

$$=\prod_{0\le k<m}P(|x_{i+n+kT}-x_{i+kT}|\le r) \tag{9}$$

$$=[P(|x_{i+n}-x_i|\le r)]^m\,. \tag{10}$$

For correlated Gaussian variables,

$$P(x_{i+n}=x,\ x_i=y)=Ae^{-(x^2-2\alpha^n xy+y^2)/B} \tag{11}$$

with $B=2\sigma^2(1-\alpha^{2n})$ and normalization $A=1/[2\pi\sigma^2(1-\alpha^{2n})^{1/2}]$. Then

$$P(|x_{i+n}-x_i|\le r)$$

$$=A\int_{-\infty}^{\infty}dx\int_{x-r}^{x+r}dy\,e^{-(x^2-2\alpha^n xy+y^2)/B} \tag{12}$$

$$=\operatorname{erf}\left[\frac{r}{2\sigma(1-\alpha^n)^{1/2}}\right], \tag{13}$$

where erf is the error function. Substitution back into Eq. (7) yields our main analytical result:

$$C(r,N,W)=\frac{2}{N^2}\sum_{n=W}^{N}(N-n)\left[\operatorname{erf}\left[\frac{r}{2\sigma(1-\alpha^n)^{1/2}}\right]\right]^m. \tag{14}$$

A. Uncorrelated limit

In the case of zero autocorrelation, Eq. (14) reduces to a simple form:

$$C(r,N,W)=[\operatorname{erf}(r/2\sigma)]^m,\quad \text{for } \alpha=0\,. \tag{15}$$

Recall that $\operatorname{erf}(x)\propto x$ for $x\ll 1$ and saturates at unity for $x\gg 1$. The correlation integral looks like $C(r,N,W)\approx(\sigma\sqrt{\pi})^{-m}r^m$ for $r\ll 2\sigma$. And the exponent m is just the value that we want our $\log C(r,N,W)$ versus $\log r$ algorithm to pick out.

In fact, this is the same limit that is approached for $N\to\infty$, independent of α. Since $\alpha^n\to 0$ as $n\to\infty$, *most* of the terms in the sum will be error functions with argument very near $r/2\sigma$; thus,

$$\lim_{N\to\infty}C(r,N,W)=[\operatorname{erf}(r/2\sigma)]^m\,, \tag{16}$$

and again the embedding dimension m will be approached.

B. Effect of autocorrelation

On the other hand, if α is very nearly one, or if N is *not* sufficiently large, the sum in Eq. (14) cannot be so simply expressed. For small n, $1-\alpha^n$ is noticeably less than unity, and the argument of the error function will be noticeably larger than $r/2\sigma$. (Also, since the erf is raised to the mth power, this effect is magnified with greater embedding dimension.) Although $r/2\sigma$ is a good approximation to the argument of the erf for most of the terms (those with large n), those few for which this is not the case can actually dominate the sum for small r.

In particular, for $r \ll 2\sigma(1-\alpha^W)^{1/2}$ the first term ($n=W$) of the right-hand side of Eq. (14) is

$$\frac{2}{N^2}(N-W)\left[\operatorname{erf}\left[\frac{r}{2\sigma(1-\alpha^W)^{1/2}}\right]\right]^m \approx \frac{2}{N}[\sigma^2\pi(1-\alpha^W)]^{-m/2}r^m . \quad (17)$$

As $N\to\infty$ this vanishes. But if $N \ll 2(1-\alpha^W)^{-m/2}$ then this first term will be much larger than

$$\operatorname{erf}(r/2\sigma)^m \approx (\sigma\sqrt{\pi})^{-m}r^m$$

and indeed dominate the entire sum. In this case,

$$C(r,N,W)=\frac{2}{N}\left[\operatorname{erf}\left[\frac{r}{2\sigma(1-\alpha^W)^{1/2}}\right]\right]^m \quad (18)$$

for $r \ll 2\sigma(1-\alpha^{W+1})^{1/2}$. For $r \approx 2\sigma(1-\alpha^W)^{1/2}$, the error function saturates at unity and $C(r,N,W)$ displays a plateau at $2/N$. Finally, for $r > \sigma\sqrt{\pi}(2/N)^{1/m}$, the first term loses its significance, and the correlation integral $C(r,N,W)$ begins to look like its corresponding $\alpha\to 0$ (or $N\to\infty$) limit.

Now, the least nonzero value that $C(r,N,W)$ may have is $2/N^2$; this is because the sum of Heaviside functions must be integral. The *usable* range of $C(r,N,W)$, i.e., the range over which $C(r,N,W)\propto r^m$, will be between $2/N^2$ and of order 1 for the uncorrelated limit, and between $2/N^2$ and of order $2/N$ for the case where autocorrelation is important. These two cases are *qualitatively* different; and the first is better by a factor of $2/N$. On logarithmic axes, the first has almost double the range of the second.

The uncorrelated limit may be achieved by taking N sufficiently large, but "sufficiently large" can be very large indeed. With $W=1$, one needs $N \gg 2\tau^{m/2}$. For example: $\tau=10$ and $m=20$ demands $N \gg 2\times 10^{10}$. This is $N^2/2 \gg 10^{20}$ distances to compute. One possibility[10] is to decrease the sampling rate in the original data, thus decreasing τ. Another recommendation, which our notation has probably made obvious by now, is to take $W>1$.

C. Recommendations for W

As a minimum recommendation, we point out that if $W > \tau(2/N)^{2/m}$ then there will be "sufficiently many" data points N that the range of linearity (in the log-log plot) will not be compromised. We note that this W is typically much less than N, so the modification we are proposing is actually quite minor.

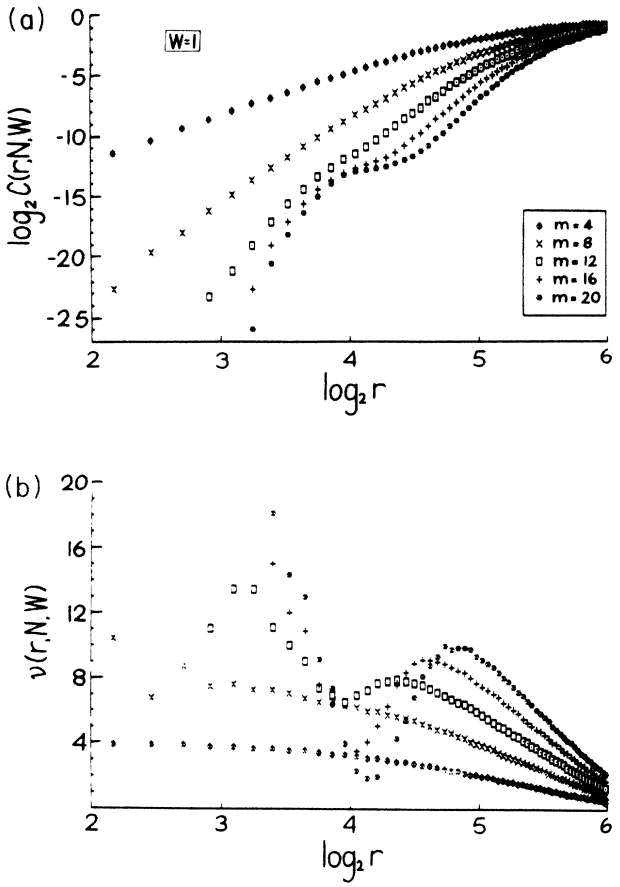

FIG. 1. (a) log-log plot of the standard ($W=1$) Grassberger-Procaccia (see Ref. 3) correlation integrals for stochastic data with $N=10000$ points, standard deviation $\sigma=20$, and autocorrelation $\alpha=0.9$ for a range of embedding dimensions m. Notice the horizontal plateau at $C(r,N,W)=2/N$; n.b. $\log_2(2/N)\approx -12.29$. (b) Slope of curves in (a). Here, the derivative $\nu(r,N,W)\equiv d[\log_2 C(r,N,W)]/d(\log_2 r)$ is approximated by $\nu(r,N,W)\approx\Delta[\log_2 C(r,N,W)]/\Delta(\log_2 r)$ where operator Δ is defined by $\Delta f(r)\equiv f(r+1)-f(r)$.

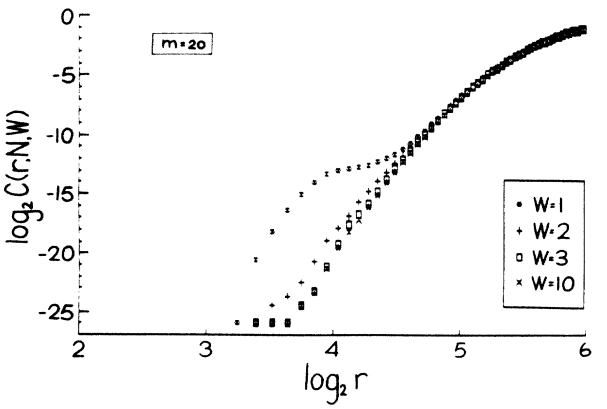

FIG. 2. log-log plot of the generalized correlation integrals for stochastic data described in caption of Fig. 1(a). Embedding dimension is fixed at $m=20$ and the cutoff parameter W is varied.

Up to now, the problem has been discussed as one that "goes away" when $N \to \infty$. As has been seen, though, it is the first few terms which cause all the trouble: and they do not go away, they are merely overwhelmed. A better algorithm, we argue, *even if N is sufficiently large,* is to toss out those overcontributing early terms right from the start. In the example case of autocorrelated stochastic data, this is achieved with $W > \tau \ln(m/2)$.

From a more intuitive point of view, the taking of $W > \tau$ ensures that the small r behavior of the correlation function counts only the "accidentally" close pairs of vectors; it is not biased by those pairs whose vectors are close in space only because they are close in time.

IV. NUMERICAL RESULTS

A. Stochastic data

With the initial goal of mimicking real data from our own physical system (the California Institute of Technology research tokamak[11]), we created a stochastic data set with $\sigma = 20$, $a = 0.9$ (so $\tau \approx 10$), and $N = 10000$. With $T = 5$ for the delay time[8] in our Takens embedding, and $m = 4$, 8, 12, 16, and 20 for the embedding dimensions,

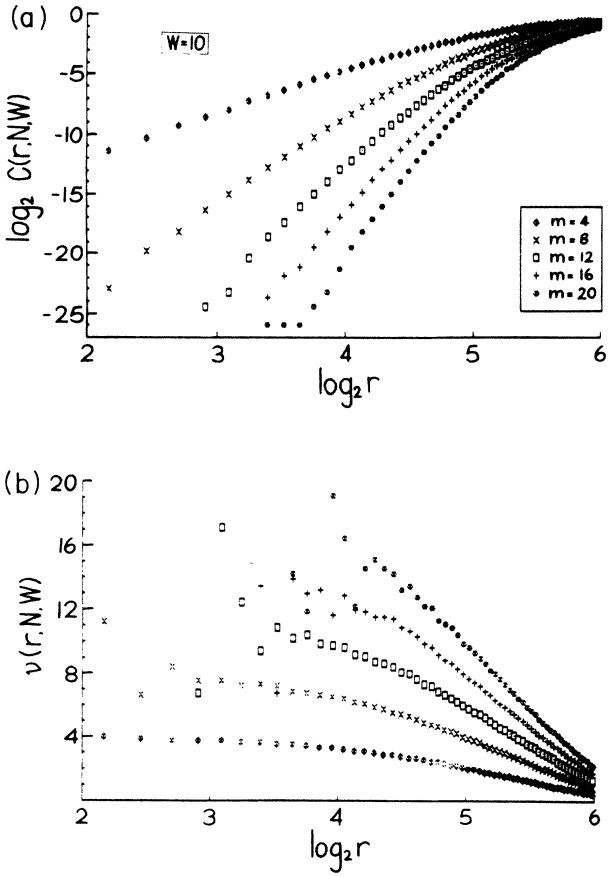

FIG. 3. (a) and (b) same as Figs. 1(a) and 1(b) except that the modified ($W = 10$) correlation integrals are plotted.

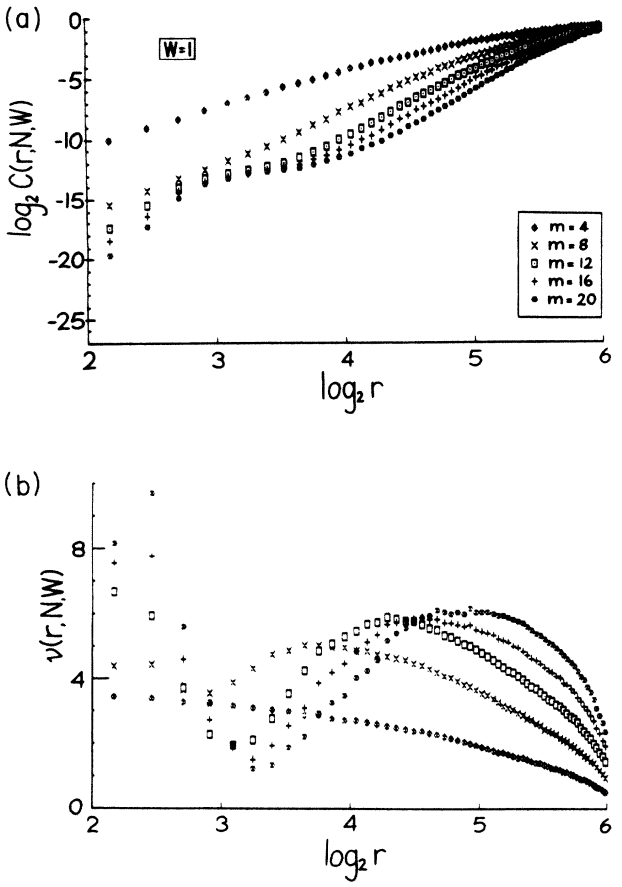

FIG. 4. (a) log-log plot of standard correlation integrals for Mackey-Glass differential delay equation (see Ref. 13) with $N = 10000$ points, $a = 0.2$, $b = 0.1$, $s = 100$, renormalized so that $\sigma = 20$ and sampled at a rate $\Delta t = 2$, for a range of embedding dimensions m. (b) Slope of curves in (a).

correlation integrals were computed.[12] A quick and uncareful look at these curves [Fig. 1(a)] might suggest a slope ν that saturates with increasing m. A closer look, however, reveals a more complicated structure. There is an extra shoulder, due almost entirely to the (anomalously large) $n = 1$ term. When a $W = 2$ correlation curve is plotted (Fig. 2) the shoulder disappears; though as the $W > 2$ curves demonstrate, anomalous contributions come also from the $n = 2$ term. Notice, however, that for $W \geq 3$ the correlation integrals are essentially unaffected by further increases in W. Figure 3(a) shows that the spurious saturation with m that was seen in the $W = 1$ curves is not present for $W = 10$; as m increases so does the slope of the $C(r,N,W = 10)$ curve. These effects are more dramatically apparent in plots [Figs. 1(b) and 3(b)] of the slopes of the $C(r,N,W)$ curves as a function of r.

B. Dynamical data

Although the effect we describe is best modeled with stochastic data, it is in fact a general feature of autocorrelated input and can be seen in dynamical data as well. The Mackey-Glass[13] differential delay equation

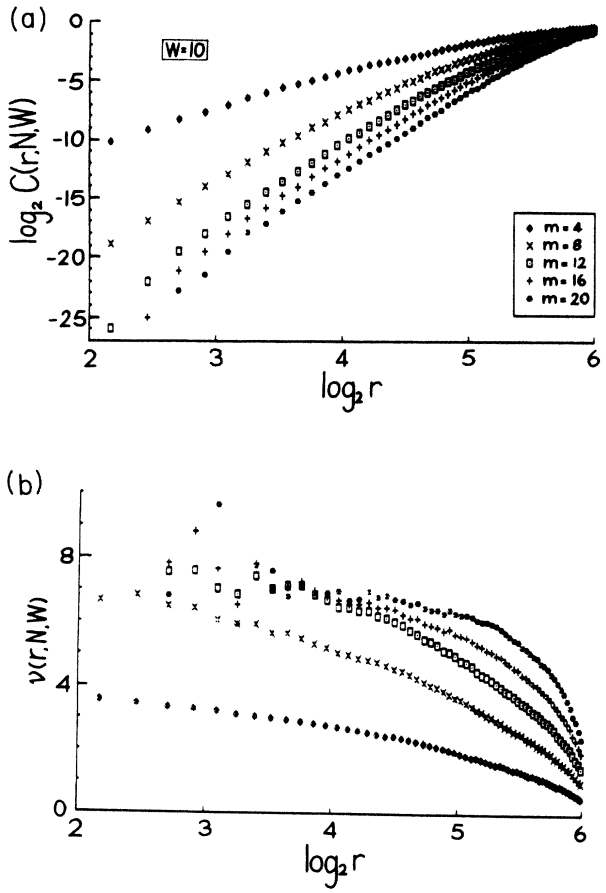

FIG. 5. (a) and (b) same as Figs. 4(a) and 4(b) except that the modified ($W=10$) correlation integrals are plotted.

$$\frac{dx}{dt}=\frac{ax(t-s)}{1+[x(t-s)]^{10}}-bx(t) \tag{19}$$

(with $a=0.2$, $b=0.1$) models a dynamical system of arbitrary complexity. Strictly, there are an infinite number of degrees of freedom—note that the initial condition is the function $x(t)$ specified over a range $t\in[t_0-s,t_0]$. Farmer[14] however has found that the *effective* number of degrees of freedom (i.e., the dimension) is finite and increases with delay time s. Grassberger and Procaccia[3] give a dimension of about 7.5 for the $s=100$ case. Using $s=100$ and the same numerical algorithm that Grassberger and Procaccia use (but increasing the sampling frequency so that the autocorrelation time is $\tau=10$, and renormalizing so that the standard deviation is $\sigma=20$—all this so that comparisons can be made with our random data), we compute correlation integrals of the Mackey-Glass data for various W and m. As in the random case, the standard ($W=1$) curves [Fig. 4(a)] display the unwanted shoulders and the modified ($W=10$) curves [Fig. 5(a)] do not. Again the derivative curves are especially compelling. Convergence of the slope to the attractor dimension (~ 7.5) is readily apparent for the modified correlation [Fig. 5(b)] but no convergence is seen in the curves [Fig. 4(b)] obtained by the standard algorithm.

V. CONCLUSION

We find that the introduction of a cutoff parameter $W>1$ improves the convergence of the standard correlation algorithm toward its $N\to\infty$ limit. Although we recommend $W\approx\tau$, where τ is the autocorrelation time of the input time series; we point out that as long as $W>\tau(2/N)^{2/m}$, where N is the number of points in the time series and m is the embedding dimension, the exact choice of W is not important.

As a final comment, we remark that these $W>1$ curves, once the $W=1$ curves had already been calculated, were very easy to obtain. We merely computed the $n=1,2,\ldots,W-1$ terms separately (each of which required only $2/N$, or 0.02% in our examples, of the work required to compute the whole curve) and subtracted them from the $W=1$ curve.

ACKNOWLEDGMENTS

I am grateful to Dr. H. S. Greenside for a conversation which led eventually to this work, also to Mark Muldoon and to Chi-Bin Chien for assorted discussions about the theoretical and computational aspects of fractional dimensions, to Dr. Roy Gould and to Dr. Stewart Zweben whose data it was I originally analyzed, and especially to Dr. Noel Corngold for many constructive hours of criticism, encouragement, and advice.

[1]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[2]For a comprehensive review of these and related methods, see J.-P. Eckman and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[3]P. Grassberger and I. Procaccia, Physica **9D**, 189 (1983).

[4]F. Takens, *Dynamical Systems and Turbulence, Warwick, 1980,* in Vol. 898 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1981).

[5]If the limit exists, then l'Hospital's rule assures us that the limit is the correct value; furthermore, the approach as $r\to 0$ is much quicker. There are attractors however for which the limit does not exist. One example is given by G. M. Zaslavskii, Phys. Lett. **69A**, 145 (1978). (This example is also quoted in Refs. 3 and 4.)

[6]F. Takens, *Dynamical Systems and Bifurcations, Groningen, 1984,* Vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1985).

[7]Formally, this is an Ornstein-Uhlenbeck process; see J. Doob, Ann. Math. **43**, 351 (1942).

[8]An excellent method for choosing T in the case of a strange attractor is given by A. Fraser and H. Swinney, Phys. Rev. A **33**, 1134 (1986). We note that although the method is not directly applicable to stochastic data, a value of T on the order of τ (or even τ/m, see Ref. 10) is probably most appropriate. The algebra is more complicated in this case, however,

and no more enlightening. Note that in the numerical experiment, we take $T=\tau/2$ to demonstrate that the assumption $T \gg \tau$ is not essential.

[9]Numerical experiments with other metrics, namely the Euclidean ($L2$) and the "taxicab" ($L1$) metric, yielded similar results.

[10]P. Atten, J. G. Caputo, B. Malraison, and Y. Gagne, J. Mec. Theor. Appl., Numero special, 133, 1984. Attention is directed to Fig. 14 of this paper, where an effect similar to the one we describe is discussed.

[11]S. Zweben and R. Gould, Nucl. Fusion **23**, 1625 (1983).

[12]Initial computations were performed on the Hypercube Mark II, a 32-node array of (Intel 8086) parallel processors. We are indebted to Dr. Geoffrey Fox and the Caltech Concurrent Computation Project for the use of this machine.

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

[14]J. D. Farmer, Physica **4D**, 366 (1982).