Estimation of the dimension of a noisy attractor

Jaap C. Schouten
Delft University of Technology, Department of Chemical Process Technology, Julianalaan 136, 2628 BL Delft, The Netherlands

Floris Takens
University of Groningen, Department of Mathematics and Computer Science, P.O. Box 800, 9700 AV Groningen, The Netherlands

Cor M. van den Bleek
Delft University of Technology, Department of Chemical Process Technology, Julianalaan 136, 2628 BL Delft, The Netherlands
(Received 27 October 1993)

A simple method is proposed to estimate the correlation dimension of a noisy chaotic attractor. The method is based on the observation that the noise induces a bias in the observed distances of trajectories, which tend to appear farther apart than they are. Under the assumption of noise being strictly bounded in amplitude, this leads to a rescaling of interpoint distances on the attractor. A correlation integral function is obtained that accounts for this effect of noise. The applicability of the method is illustrated with two examples, viz., the Lorenz attractor with additive noise and experimental time series of pressure fluctuation data measured in gas-solid fluidized beds.

PACS number(s): 05.45.+b

INTRODUCTION

One of the main problems in dimension estimation from chaotic time series has to do with the fact that time signals from natural phenomena are corrupted by noise. Although one can do one's best to limit the amount of noise in the time signal as much as possible by carrying out a very-well-controlled measurement, a complete separation between the deterministic component in the signal, if present, and the noise is practically impossible. Thus, the measured time signal will always contain some noise due to random influences and inaccuracies that can never by ruled out completely. These inaccuracies may be caused by, for example, the measuring device (relative and absolute accuracies) and analog-to-digital conversion (discretization errors). These possible sources of inaccuracies are generally referred to as *measurement noise* or *additive noise.* Principally, this type of noise is considered to be independent with no or only very short term correlations. In most practical cases, measurement noise is considered to be Gaussian, especially when it is due to a combination of various sources of independent random fluctuations. However, this does not necessarily have to be the case in every situation: experimental signals may also be corrupted by measurement noise that is strictly bounded in magnitude.

Furthermore, minor random fluctuations in settings of main system parameters (e.g., flow, pressure, or temperature in physical experiments) may propagate in the dynamic system causing randomlike fluctuations that are not specific to the system. It might also be the case that dynamic phenomena or systems are (partly) influenced by *intrinsic* events taking place at random. These types of noise are generally defined as *dynamic noise;* they directly influence the evolution of the dynamic system in time. A practical example of the latter is that of gas-solid flow phenomena in some apparatus or chemical reactor where local solid-solid or solid-gas interactions might be of a more random nature, while global phenomena like the solid and gas circulation patterns are determined by deterministic laws. This means that "observed" invariants that characterize the dynamics, like dimension and entropy, may depend on the specific scale of observation. For example, pressure fluctuation measurements in gas-solid flow systems may reveal randomlike behavior at small scales, while at larger scales, deterministic chaotic characteristics can be found. A complicating feature is that scale dependence of observed invariants may also occur in noise-free systems; however, specifically at small length scales, it is expected that this is of a different nature.

NOISE REDUCTION

It does not seem to be an easy task to separate *dynamic noise* from the rest of a (chaotic) time signal. Especially when the source of the dynamic noise is an intrinsic feature of the system, one can even raise the question whether such a separation would be of great use in trying to understand the system. The point then is, of course, whether the dynamic noise has completely destroyed the deterministic chaos or whether the chaotic dynamics are still present in some neighborhood of the noisy time signal (shadowing property). This latter case is exactly the basis of recently proposed noise reduction schemes [1]. Shadowing properties, if they are present, can be interpreted as a kind of rough equivalence between dynamical and additive (measurement) noise in the case of times that are not too long.

Dealing with the problem of *measurement noise* seems, therefore, to be a little bit easier. In fact, one can consider one or more of the following approaches: first of all,

before the experiment is performed, one may decide to try to measure the time series as accurately as possible. This means that one is required to use an accurate measuring device (e.g. inaccuracy typically less than 1% or even less than 0.1% of the full scale). In addition, it is advisable to optimize the measurement resolution by adapting the full scale of the measuring device as close as possible to the maximum difference between expected data points. Furthermore, analog-to-digital conversion should be done with a reasonably high resolution to limit discretization errors [typically at least 12-bit precision (1:4096) or even 16-bit precision (1:65 536) may be required].

During the acquiring of the data with a data-acquisition system, one should of course avoid aliasing. Moreover, it might be useful to remove high frequency noisy components in the signal using a low-pass analog filter. In doing so, one should be sure that no relevant and characteristic time scales in the signal are removed as well. In general, it will be possible to recognize these time scales from the power spectrum. A problem may arise when the characteristic time scales of the noise and of the data are similar.

After the experiment has been carried out, it may be useful to try to clean the measured data set by application of a noise reduction method [1] and then to pass the cleaned time signal through a dimension estimation routine. However, after such a suppression of dynamic or measurement noise, a certain (generally unknown) level of noise may still be present. This remaining noise will inevitably influence the dimension estimation. Furthermore, the specific choices of the parameters in these noise reduction algorithms may not always be straightforward and unambiguous as these are dependent on the noise level, the sampling rate, and the length of the time series [1]. Moreover, it is not always clear in what way nonlinear filtering changes the characteristics of the underlying chaotic time signal. It may be possible that after filtering the remaining "skeleton" contains properties that were not present in the original data. For these reasons, in the case of dimension estimation, it might be very useful to evaluate first in what way the noise in the (unfiltered) time series influences the *correlation integral* from which the dimension is computed.

In this paper we will specifically focus upon this latter issue. We will derive a simple analytical expression for the rescaled correlation integral that is a function of the noise strength as well as of the dimension of the underlying (uncorrupted) chaotic attractor. This correlation integral function differs from that given recently by Schreiber [2] which is derived under the assumption that the noise is Gaussian. Our method is derived under the assumption that the data are samples from a low dimensional attractor contaminated by noise which is strictly bounded in amplitude.

We have also tried an approach similar to that of Schreiber [2] in which we have determined the influence of Gaussian noise on the correlation integral in a somewhat different way (see Ref. [3]). Assuming that the noise added to each data point is independent and normally distributed, we have been able to derive an exact description of the correlation integral function of the perturbed distances. However, for our data it did not lead to significantly better results in comparison to the (much simpler) method presented in this paper.

CORRELATION INTEGRAL

Our method is based on the correlation integral that is defined as [4,5]

$$C(l)=\frac{1}{N(N-1)}\sum_{i\neq j}\Theta(l-\|\underline{X}_i-\underline{X}_j\|)\ , \tag{1}$$

where Θ is the Heaviside function. The correlation integral counts the number of pairs of points $(\underline{X}_i,\underline{X}_j)$ on the attractor whose distance is smaller than l. It scales according to a power law with $C(l)\approx l^D$ in the limit as $l\rightarrow 0$ and the number of points $N\rightarrow\infty$. D is defined as the correlation dimension of the attractor.

To be able to compute the correlation integral, first we have to reconstruct the attractor in the state space. Hereto we consider a scalar time series of a single variable of the dynamic system

$$X(t)=(x(t_1),x(t_2),\ldots,x(t_N))=(x_1,x_2,\ldots,x_N)\ .$$

An m-dimensional state space is reconstructed using the delay time method [6] by defining reconstructed vectors as

$$\underline{X}_i=(x_i,x_{i+\gamma},x_{i+2\gamma},\ldots,x_{i+(m-1)\gamma})^T\ ,$$

where γ is the delay in the number of sample time steps τ_s between successive vector elements.

It is advantageous to relate the length of the reconstructed vector to a specific time scale T in the time series [7]. This time scale or embedding *time window* is a segment $[t,t+T]$ of the time series with length in time of $T=m\gamma\tau_s$. It is convenient to take the delay γ equal to unity so that at a given time window length, the number of points m in the window is only determined by the sampling frequency. As mentioned by Grassberger, Schreiber, and Schaffrath [8], in the case of infinitely precise measurements, any $T>0$ will do to characterize the state of the dynamic system. However, in practice the length of the time window is limited from below and from above. First of all, T should not be too large so that the first value x_i and the last value x_{i+m-1} in the time window are practically uncorrelated. But, on the other hand, T should be sufficiently large to cover the dominant frequency of the motion [8]. In addition, the proper choice of the window width may also be affected by the noise in the signal, as Casdagli *et al.* [9] have shown in the case of measurement noise.

Independent of the choice of the length of the time window is the choice of the number of measurement points to be contained in the time window. This number of points m determines the number of elements of the vector $\underline{X}$ in the reconstructed state space. m can be considered as the embedding dimension that should be chosen in such a way that a faithful and smooth embedding is obtained to have the data points sufficiently open up the details of the attractor. Thus, m should be chosen

as large as possible but at least large enough to obey the embedding theorem (Sauer, Yorke, and Casdagli [10]). As soon as T and m have been chosen, the sampling frequency automatically follows from $f_s = m/T$.

It may also be convenient to choose an upper bound l_0 of the scaling distance l. This is due to the fact that the scaling relationship l^D only holds for sufficiently small l. Moreover, at larger l, the correlation integral will be dominated by saturation effects [$C(l) \rightarrow 1$] when l becomes of the order of the size of the attractor. The necessity of a maximum scaling distance is even more clear when one needs to normalize distances when using Takens's maximum-likelihood estimator of the correlation dimension [11]. First, l_0 must be chosen larger than the maximum possible noise distance l_n between two points on the (noisy) attractor. Furthermore, l_0 should be sufficiently large, so that the sample of distances with $l_n \leq l \leq l_0$ is not too small; however, this sample size can be raised by considering more pairs $(\underline{X}_i, \underline{X}_j)$. Also, l_0 should not be too big: as mentioned earlier, the scaling law l^D will certainly be violated for values of l that are too large. Especially when the correlation dimension D is high, the last two requirements may force one to consider very large numbers of pairs $(\underline{X}_i, \underline{X}_j)$.

The distance between two reconstructed vectors in an m-dimensional state space can be calculated using the Euclidean norm or the maximum norm. Because the results based on both norms in terms of dimension are equivalent [5], it is advantageous from a computational viewpoint to use the maximum norm. We will show further that application of the maximum norm also facilitates the estimation of the error level of a noisy attractor directly from the correlation integral. The distance between two points $\underline{X}_i$ and $\underline{X}_j$ on the attractor is thus defined as

$$\|\underline{X}_i - \underline{X}_j\| = \max_{0 \leq k \leq m-1} |x_{i+k} - x_{j+k}| , \tag{2}$$

with

$$\underline{X}_i = (x_i, x_{i+1}, \ldots, x_{i+m-1})^T$$

and

$$\underline{X}_j = (x_j, x_{j+1}, \ldots, x_{j+m-1})^T .$$

For a numerical determination of D, we generally start with generating a large number of randomly chosen pairs $(\underline{X}_i, \underline{X}_j)$ from the sequence $X(t)$ and count the number of pairs for which the distance is smaller than l for various values of l with $l \leq l_0$. In this way we obtain estimates of $C(l)$ for which it is expected that $\ln[C(l)]$ depends linearly on $\ln(l)$ for small l (the scaling region). The correlation dimension then follows from the slope of $\ln[C(l)]$ versus $\ln(l)$. This method has some disadvantages when it is applied to experimental time series.

INFLUENCE OF NOISE

The most important difficulty arises with the presence of noise that may corrupt the scaling behavior at all length scales. In practical cases, the log-log plot will then not even show a linear part at all. This means that the power law relationship l^D no longer gives a good representation of the interpoint correlations.

The influence of noise on the correlation integral can be evaluated as follows: Let us consider two points $\underline{X}_i$ and $\underline{X}_j$ that are located on the reconstructed attractor on different orbits. These points are not disturbed by noise, so they may be considered as true points satisfying the exact (chaotic) dynamics of the system. The maximum norm distance between these points is given by Eq. (2). Let now each point in the time series be corrupted by noise that is bounded in magnitude with a maximum possible amplitude of $\pm\frac{1}{2}\delta x_{\max}$. In that case the elements ($z_{i,k}$ and $z_{j,k}$) of the noise-corrupted vectors $\underline{Z}_i$ and $\underline{Z}_j$ each are assumed to be composed of a noise-free part ($x_{i,k}$ and $x_{j,k}$) and a noisy part ($\delta x_{i,k}$ and $\delta x_{j,k}$) according to $z_{i,k} = x_{i,k} + \delta x_{i,k}$ and $z_{j,k} = x_{j,k} + \delta x_{j,k}$, with

$$-\tfrac{1}{2}\delta x_{\max} \leq \delta x_i \leq +\tfrac{1}{2}\delta x_{\max}$$

and

$$-\tfrac{1}{2}\delta x_{\max} \leq \delta x_j \leq +\tfrac{1}{2}\delta x_{\max} .$$

Basically, we assume thus that there exists a trajectory satisfying the true dynamics of the chaotic system, which is sufficiently close to the measured, noise-corrupted trajectory. If this is not the case, the simple linear addition of noise-free and noise-corrupted contributions is not allowed.

In principle, δx can be due to any time of noise; however, we will assume here that in the case of dependent (dynamic) noise, the time scale of the correlations is considerably smaller than the time scale T of the time window used in the reconstruction. Theoretically, when the number of vector elements is infinite, $m \rightarrow \infty$, the probability of finding two corresponding elements $z_{i,k}$ and $z_{j,k}$ that are maximally corrupted with $-\frac{1}{2}\delta x_{\max}$ and $+\frac{1}{2}\delta x_{\max}$, respectively, will then be unity. Furthermore, one needs the maximally corrupted pair also to be the pair for which $|x_{i,k} - x_{j,k}|$ is maximal. If the embedding dimension is sufficiently large, while $x_{i,k}$ and $x_{j,k}$ depend smoothly on k, then this coincidence can be well approximated (see also Appendix A). The maximum norm distance between the corrupted vectors is thus found from

$$\begin{aligned} l_z &= \lim_{m \rightarrow \infty} \max_{0 \leq k \leq m-1} |z_{i,k} - z_{j,k}| \\ &= \lim_{m \rightarrow \infty} \max_{0 \leq k \leq m-1} |(x_{i,k} + \delta x_{i,k}) - (x_{j,k} + \delta x_{j,k})| \\ &= \lim_{m \rightarrow \infty} \max_{0 \leq k \leq m-1} |x_{i,k} - x_{j,k}| + \delta x_{\max} = l_x + l_n . \end{aligned} \tag{3}$$

Here, l_z is the corrupted distance, l_x is the noise-free distance, and $l_n = \delta x_{\max}$ is the maximum noise distance.

Equation (3) illustrates that the probability of finding interpoint distances l_z below $l_n = \delta x_{\max}$ is zero. In other words, this means that $C(l_z \leq l_n) = 0$ and $C(l_z > l_n) > 0$, which implies that the maximum noise scale can be directly obtained from the correlation integral. This can be seen as a sort of *filtering* of the data where application of the maximum norm *prevents* interpoint distances

below the maximum noise distance $\delta x_{\max}$ from being included in the computation of the correlation integral. In practice, m will be bounded by practical limitations. This implies that a good estimate of the maximum noise level will only be obtained when m can be taken sufficiently large. Principally, the choice of an appropriate value of m depends on the type of noise and on its resolution; see Appendix A for an order of magnitude estimation of m in the case of uniform white noise. Generally, it is appropriate to choose m as large as possible. In practice, a trivial approach would be, of course, to increase m to observe its effect on the correlation integral.

When the power law dependency holds for the noise-free distances l_x according to $C(l_x)\sim(l_x)^D$, we can write that

$$C(l_z|l_z>l_n)\sim(l_z-l_n)^D ,$$

because $l_x=l_z-l_n$ [Eq. (3)]. With the requirements that $C(l_z=l_n)=0$ and $C(l_z=l_0)=1$, we obtain

$$C(l_z)=\left[\frac{l_z-l_n}{l_0-l_n}\right]^D , \quad l_n\le l_z\le l_0 . \tag{4}$$

We normalize all distances with respect to the maximum scaling distance l_0, using $r=l_z/l_0$ and $r_n=l_n/l_0$, in which case it follows that

$$C(r)=\left[\frac{r-r_n}{1-r_n}\right]^D , \quad r_n\le r\le 1 . \tag{5}$$

This expression illustrates that we have effectively rescaled the corrupted distances l_z in order to let the correlation integral obey the power law function again.

Equations (4) and (5) can also be derived by the following reasoning: If we embed a noisy chaotic attractor in an m-dimensional state space, we may expect that the noisy part will be space filling on length scales smaller than the noise scale. When it is space filling, $C(l)$ scales like l^m. We may thus expect that a plot of $\ln[C(l)]$ as a function of $\ln(l)$ will have a slope of D down to length scales characterized by the noise strength ($l>l_n$) and then a slope of m for distances $l\le l_n$, as has been demonstrated, for example, in Ref. [12]. This suggests splitting the correlation integral into two parts,

$$C_1(l\le l_n)\sim l^m$$

and

$$C_2(l>l_n)\sim(l-l_n)^D .$$

Using the requirements that $C_1(0)=0$, $C_1(l_n)=C_2(l_n)=l_n^m$, and $C_2(l_0)=1$ and after normalizing, we obtain that

$$\begin{aligned} &C_1(0\le r\le r_n)=r^m , \\ &C_2(r_n\le r\le 1)=r_n^m+(1-r_n^m)\left[\frac{r-r_n}{1-r_n}\right]^D . \end{aligned} \tag{6}$$

When $m\to\infty$, it is immediately seen that $C_1(r\le r_n)=0$ and Eq. (6) effectively reduces to Eq. (5). If m is taken at least of the order of 50, it is clear that for moderate values of r_n (0 to 0.5), $C_1(r\le r_n)=r^m$ will be very small (smaller than 10^{-15}) and Eq. (6) reduces again to Eq. (5).

DIMENSION ESTIMATION

When l_n is known *a priori*, all distances can simply be rescaled. The correlation dimension can then be estimated from the slope of $\ln[C(l)]$ versus $\ln(l-l_n)$ (or $\ln[C(r)]$ versus $\ln(r-r_n)$) or, alternatively, D can be computed using the maximum-likelihood method [11] from which the maximum-likelihood dimension $D_{ML,n}$ of the noise-free chaotic attractor is found as (see Appendix B)

$$D_{ML,n}=\left[\frac{-1}{M}\sum_{i=1}^{M}\ln\left[\frac{r_i-r_n}{1-r_n}\right]\right]^{-1} . \tag{7}$$

Here, M is the sample size of interpoint distances r_i with $r_n\le r_i\le 1$.

However, in nearly all practical situations, l_n will be unknown or, in the best case, not known to sufficient accuracy. In that situation the parameters r_n and D can be estimated from a nonlinear least-squares fit of the integral function Eq. (5) [or (6)] to the experimentally determined correlation integral. Because the function $C(r)$ is known analytically, it is convenient to use the Levenberg-Marquardt least-squares method to estimate r_n and D. This standard curve fitting method is elegant and quick and works very well in practice [13].

Of course, in principle, it is possible also to formulate maximum-likelihood estimators for D and r_n. However, in practice, it has been found that the numerical calculation of these estimators is not as straightforward as and much more time consuming than with Takens's best estimator [11]. It requires the solving of two nonlinear algebraic equations—which is not always simple—while in the meantime, it also necessitates the storage of all distances r_i on which the best estimation is based. That has been the reason we preferred the fitting procedure. Of course, after the fit has been obtained, the resulting correlation integral should always be compared with the original one in order to observe how good the fit is.

When we apply Takens's maximum-likelihood estimator of the correlation dimension straightforwardly *without* rescaling of the noise-corrupted distances, the dimension estimate D_{ML} is given as [11]

$$D_{ML}=\left[\frac{1}{M}\sum_{i=1}^{M}-\ln(r_i)\right]^{-1} . \tag{8}$$

Because the computation of D_{ML} is now based upon the noise-corrupted distances r_i, it will effectively overestimate the true correlation dimension of the uncorrupted attractor. This can readily be seen from

$$\frac{D_{ML}}{D_{ML,n}}=\frac{\sum_{i=1}^{M}\ln(r_i-r_n)-\sum_{i=1}^{M}\ln(1-r_n)}{\sum_{i=1}^{M}\ln(r_i)}>1\ ,$$

$$\forall r_n \le r_i \le 1\ . \quad (9)$$

In that case the approach in Ellner [14] might be considered as an alternative.

D_{ML} and D will be related because these invariants are obtained from the same correlation integral. Their relationship can be derived by calculating the expectation value of $\ln(r)$ with probability distribution $p(r)dr$ on [0,1] which is determined from the cumulative distributions $C_1(r)$ on $[0,r_n]$ and $C_2(r)$ on $[r_n,1]$ (see Appendix C). It is found that for sufficiently high values of m that D_{ML} and D are approximately related according to

$$D_{ML}=\frac{D+r_n}{1-r_n}\ . \quad (10)$$

So D_{ML} can readily be calculated when D and r_n have been determined from a least-squares fit of the correlation integral. Furthermore, it is seen from Eq. (10) that D_{ML} is larger than D in the case of the presence of noise ($r_n>0$), which of course is to be expected.

RECONSTRUCTION PARAMETERS

So far, we have not given any indication as to a proper choice of the length of the time window T and the maximum scaling length l_0. The choice of these parameters is actually a matter of empiricism because no clear-cut theoretically founded choices are available. The specific choice of these parameters is not crucial for the demonstration of our method, but to illustrate its functioning on real data, we nevertheless have to choose some values of T and l_0. For that reason we will only give here some suggestions that seem to work quite well for the evaluation of *our* data. For example, we have used these specific choices of the reconstruction parameters successfully in our studies of chaotic pressure fluctuations in gas-solid fluidized beds [15,16]. For us, the most important criterion for the choices of T and l_0 is that they should be unambiguous with the exclusion of any subjective interpretation. In this way it is possible to compare results between different measurements in a relevant way.

It is our experience that the *average cycle time* T_c of the time series provides a robust and characteristic measure for the length of the time window T. The average cycle time is defined as the *average* time that is needed to complete a full cycle after the first passage through the average of the signal:

$$T_c=\frac{[\text{length of time series (units of time)}]}{[(\text{number of crossings of average of time series})/2]}\ . \quad (11)$$

This choice is predominantly based on the assumption that the structural information needed to reconstruct a point on the attractor is mainly contained in the data sampled during one average time period of length T_c. Generally, this period of time can be compared to an average orbital period on the attractor. Choosing a longer time segment T would add only a little information to that already contained in the components considered in the window of length T_c (see also Ref. [17]).

The value of T_c is unambiguous for each given time series and can be readily calculated by counting the number of crossings of the time series' average, Eq. (11). We also suggest using T_c as the characteristic time in the exclusion of interpoint distances to avoid dynamic correlations in the computation of the correlation integral [18].

In the derivation of our algorithm to estimate D and r_n, we suggested taking m at least of the order of 50 points per time window (see also Appendix A). In general this will be quite sufficient to satisfy the requirement that $m \ge 2D+1$ in order to accommodate the attractor in the state space. In most cases it seems appropriate to take m in the range 50–200. This immediately fixes the required sampling frequency f_s at 50 to 200 times the average cycle frequency f_c. Of course, it is of no use to have f_s exceeding the response frequency f_r of the measuring device. Thus, in practice, f_r imposes an upper limit on the choice of m ($m=f_s/f_c \le f_r/f_c$), where, first of all, of course, the response frequency should satisfy the requirement that $f_r/f_c \ge 50$.

These necessarily high values of m will generally imply that the data have to be oversampled. In combination with white noise, this is not the usual regime where noise reduction methods are applied because in this regime a straightforward low-pass filter may do equally well. However, when the characteristic time scales of the noise and of the data are similar, these simple filters become problematic.

According to our experience, it is furthermore very practical to take the maximum scaling length l_0 equal to the *average absolute deviation* of the time series. This is in general a robust estimator of the data's width around the mean. In particular, application of the average absolute deviation Δx is very convenient in combination with the maximum norm; it is computed from

$$\Delta x=\frac{1}{N}\sum_{i=1}^{N}|x_i-\bar{x}|\ . \quad (12)$$

The *average* of the time series is obtained from

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i\ . \quad (13)$$

To eliminate any effect of origin or scale, we further suggest working always with *normalized* data points. Data points are conveniently normalized with respect to the average absolute deviation to create a standarized time series with data points y_i according to

$$y_i=\frac{x_i-\bar{x}}{\Delta x}\ . \quad (14)$$

The data points y_i in the normalized time series thus will have a mean equal to zero and an average absolute deviation equal to unity.

Of course, in specific cases it may be possible that the correlation integral, from which the dimension is estimated, is dependent on the choice of the reconstruction parameters (T and m). That means that the estimated dimension may also depend on this choice. Furthermore, the definition of the correlation dimension is only valid in the limit case, which means specifically for small distances l_0. In principle, this dependence can be checked by repeating the estimation for various combinations of the reconstruction parameters and the scaling region. Especially in the case of time series with a relatively large entropy, the effect of the dependence of the dimension on the reconstruction parameters may be clearly observed specifically with long time windows where data points have become practically uncorrelated within one window. In this latter case it can even be expected that the dimension will increase with longer time windows irrespective of the presence of noise.

SUMMARY OF THE ALGORITHM

We will now briefly summarize the various steps in our algorithm to estimate D and r_n for a given time series. This algorithm is the basis for a user-friendly and menu-driven software package RRCHAOS that has been written by the authors [19].

Step 1. First of all, before we start acquiring the time series, we have to choose a proper sampling frequency to obtain at least $m=50$ points per average cycle (preferably somewhere in the range 50–200 points). This can be done by carrying out the experiment once at some initial sampling frequency, adjusting the sampling frequency f_s in the appropriate direction, and repeating the experiment to obtain m at least in the range 50–200 points per average cycle. Be sure that $f_s < f_r$.

m is computed as follows: Calculate the time series' average [Eq. (13)] and count the number of crossings $N_{\text{crossings}}$ of this average. Calculate the number of cycles N_{cycles} in the time series from

$$N_{\text{cycles}} = N_{\text{crossings}}/2 \ .$$

Calculate m from $m = N/N_{\text{cycles}}$, N being the number of points in the time series; of course, m should be rounded off to obtain an integer number.

Step 2. Calculate the average absolute deviation from Eq. (12) which gives l_0.

Step 3. Choose randomly pairs of points

$$(\underline{Z}_i, \underline{Z}_j) \ ,$$

$$\underline{Z}_i = (z_i, z_{i+1}, \ldots, z_{i+m-1})^T \ ,$$

and

$$\underline{Z}_j = (z_j, z_{j+1}, \ldots, z_{j+m-1})^T \ .$$

To avoid dynamic correlations, be sure that $|i-j| > m$. Determine the maximum norm distance l_z. Create a histogram of the number of pairs with distances $l_z \le l_0$. No simple rule can be given for the size of the bins in the histogram. It is our experience that bin sizes Δl of the order of $l_0/400 \le \Delta l \le l_0/100$ give good results.

Also, no clear rule can be given for the total number of pairs M to be included in the histogram. The more pairs are taken into account, the more accurate the correlation integral will be. As a first indication, we base the total number of pairs on the relative standard error of Takens's maximum-likelihood estimator (see Appendix B). We require this standard error to be less than 1%; this means that M should be at least of the order of 10 000.

Step 4. Calculate the normalized cumulative histogram. This cumulative histogram may be considered equal to the correlation integral when the (normalized) bin size Δr is sufficiently small (viz., $\frac{1}{400} \le \Delta r \le \frac{1}{100}$, as suggested in step 3).

Step 5. Estimate D and r_n from fitting the correlation integral to Eq. (5) or Eq. (6) using the Levenberg-Marquardt least-squares method.

EXAMPLES

Let us now illustrate the applicability of our method with two different sets of time series.

Lorenz attractor

First, we will demonstrate our method using time series of the X variable of the Lorenz system that are calculated with a fourth-order Runge-Kutta method. The parameters are $\sigma = 10.0$, $b = 2.67$, and $R = 28.0$; a total of 75 000 points have been computed with a sampling frequency of 50 Hz; the time step in the Runge-Kutta procedure is 0.001; initial conditions are $X=1$, $Y=1$, and $Z=1$; the first 10 000 points are omitted before the time series (65 000 points) is acquired; and data points are discretized with 16-bit resolution. Six time series (L1–L6) have been computed in which the original data points (X variable) are disturbed with additive, independent, and uniformly distributed noise with, respectively, $\delta x_{\max}$ equal to 0.25, 0.5, 0.75, 1.0, 1.5, and 3.0 in units of the variable X. These maximum levels of the added noise correspond with 3.9% (L1) to 46.1% (L6) of the average absolute deviation. An overview of the results is given in Table I. In general, the rescaled correlation integral function gives a much better fit to the data. The sum of least squares of the rescaled function is generally a factor 10 to even up to 1000 times smaller. In Table I it is observed that the rescaled correlation integral leads to good estimates of the dimension even at a noise level up to 7.7% (L2) where the maximum-likelihood method already leads to an overestimation. At higher noise levels (L3–L6), the maximum-likelihood method overestimates the dimension even more, while the rescaled correlation integral function tends to underestimate the dimension. As expected (see Appendix A), the level of noise in the signal (r_n) is underestimated.

TABLE I. Results of estimation of dimension and noise level of Lorenz data sets.

Lorenz series	X	L1	L2	L3	L4	L5	L6
δx_{max}	0	0.25	0.5	0.75	1.0	1.5	3.0
$\delta x_{max}/\Delta x$ (%)	0	3.9	7.7	11.6	15.4	23.1	46.1
m	178	158	132	112	93	73	39
T_c (sec)	3.56	3.16	2.64	2.24	1.86	1.46	0.78
M	14 167	22 790	40 225	66 750	98 610	149 060	150 968
D_{ML}	2.14	2.14	2.22	2.28	2.35	2.52	3.17
D	2.14	2.14	2.13	2.05	2.01	1.85	1.60
r_n (%)	0.0	0.0	2.2	5.6	8.7	16.8	35.6

Fluidized bed data

The second example is based on time series of pressure fluctuation data that have been measured in two gas-solid fluidized beds operated with air at ambient conditions. The experimental conditions are summarized in Table II. The pressure fluctuations have been recorded with a Kistler piezoelectric transducer, low-pass filtered and discretized with 16-bit precision. The pressure fluctuations were measured with respect to the average pressure at the tip of the pressure sensor. The time series consist of 65 000 points; the results are summarized in Table III. It is observed that D is in all cases considerably lower than the maximum-likelihood dimension D_{ML}. The estimated noise level is relatively high, viz., about one-third of the average absolute deviation, which is, for example, much higher than the inaccuracy of the pressure sensor [even in the worst case (FB3), this inaccuracy is still smaller than 3% of the average absolute deviation]. This suggests that another type of (dynamic) noise is also present which the method corrects for. As mentioned also in the case of the Lorenz data, the fits of the rescaled correlation integral function are much better than that of the original function. A typical example is shown in Figs. 1(a)–1(c) where the rescaled and original functions are compared with the measured correlation integral (data set FB3). Figure 1(c) in particular illustrates that the difference between the rescaled correlation function [Eq. (6)] and the experimental correlation integral is much smaller than in the case of the original correlation function based on the maximum-likelihood dimension.

CONCLUDING REMARKS

We have derived a simple expression for the correlation integral that accounts for the effect of noise which is strictly bounded in magnitude. This function is based on a rescaling of the interpoint distances on the attractor. The assumption is that the noisy trajectories on the attractor stay within the proximity of the uncorrupted trajectories of the underlying (chaotic) dynamic system. In that case separation between the noisy component and the chaotic component is possible and useful. The rescaled correlation integral can be used to estimate the correlation dimension belonging to the underlying dynamic phenomenon. At the same time the noise level is obtained, however, generally this will be underestimated.

We suggest estimating the correlation dimension and the noise scale with this new method and reporting these values possibly together with the maximum-likelihood dimension. In this way a useful impression is obtained about the presence of noisy components in the time signal as well as about their influence on the increase of the correlation dimension.

TABLE II. Experimental conditions of fluidized bed data sets.

	Fluidized bed series		
	FB1	FB2	FB3
Bed diameter (m)	0.284	0.15	0.15
Static bed height (m)	0.60	0.30	0.30
Superficial gas velocity (m/s)	1.04	1.53	0.051
Particle type	Polystyrene	Polystyrene	Glass beads
Average particle diameter (mm)	1.78	1.78	0.1
Particle density (kg/m^3)	1102	1102	2800
Minimum fluidization velocity (m/s)	0.61	0.61	0.02
Sampling frequency (Hz)	200	200	1000
Low-pass filter frequency (Hz)	60	40	340

TABLE III. Results of estimation of dimension and noise level of fluidized bed data sets.

	Fluidized bed series		
	FB1	FB2	FB3
m	155	102	153
T_c (sec)	0.78	0.51	0.15
M	53 267	41 117	32 776
D_{ML}	4.83	5.18	6.05
D	3.08	3.14	3.73
r_n (%)	30	33	33

Finally, we would like to make a general remark about the meaning of the correlation dimension. In the case of a self-similar attractor, the correlation dimension D may well be able to quantify the self-similarity in a "scaling region" where $\ln[C(r)]$ is proportional to $\ln(r-r_n)$. Self-similarity, however, does not necessarily have to be always present in practical situations; in that case it is useful to interpret the correlation dimension in a broader sense. Therefore, we would like to emphasize that the purpose of our method is not to look for a scaling region in the first place, but to describe the correlation integral as adequately as possible by a two-parameter model that corrects for the presence of noise. This implies that one should not conclude on the basis of the model fit *only* that the data are samples from a strange attractor with noise added.

APPENDIX A

An order of magnitude estimate of m can be derived analytically for the simple case when the time segments that represent the true vectors $\underline{X}_i$ and $\underline{X}_j$ are more or less parallel in state space. This is the case when $|x_{i,k}-x_{j,k}| \approx \text{const}$ with $1 \le k \le m$. The maximum norm distance will now be obtained for that combination of $x_{i,k}$ and $x_{j,k}$ with the largest noise amplitudes $\delta x_{i,k}$ and $\delta x_{j,k}$ with $1 \le k \le m$. However, when the time segments are not parallel, there will effectively always be a shorter range of successive corresponding elements $z_{i,k}$ and $z_{j,k}$ wherein the maximum norm distance is obtained: $k_1 \le k \le k_2$ with $k_1 - k_2 < m-1$ or it may even be the case that $k_1 - k_2 \ll m-1$. In that situation the maximum noise level is estimated from a smaller subset $m_0 < m$ of corresponding vector elements. This "effective embedding" m_0 will be strongly dependent on the specific time series that is considered and moreover it will differ from segment to segment. This makes it difficult to estimate m_0. Nevertheless, m_0 provides a lower bound of m which we can use to estimate the order of magnitude of m in a specific case.

Let us suppose now that we have measured a time series that we discretize with 16-bit precision. That means that the time series consists of integer numbers between 1 and 65 536. Suppose that in the case of uniformly distributed noise, the maximum measurement error in integer units equals $\pm\frac{1}{2}(N_p-1)$ (for convenience we take N_p to be an odd number) with equal probability $1/N_p$ for each data point in the time series. The probability that two corresponding elements in two reconstructed vectors are maximally corrupted with maximum noise amplitudes of, respectively, $+\frac{1}{2}(N_p-1)$ and $-\frac{1}{2}(N_p-1)$ equals $1/N_p^2$ [because there is only one combination leading to the maximum difference of (N_p-1)]. Thus the probability that two corresponding elements are *not* maximally corrupted equals $(1-1/N_p^2)$. So with m successive vector elements, the probability that none of the corresponding elements is maximally corrupted equals $(1-1/N_p^2)^m$. This leads finally to the probability that *at least* one set of two corresponding vector elements is maximally corrupted:

$$P(\text{maximum norm of noise part of distance equals } l_n) = 1 - \left[1 - \frac{1}{N_p^2}\right]^m . \tag{A1}$$

For example, with a noise resolution $N_p = 200$ and a number of vector elements $m = 100$, the probability that exactly the maximum noise amplitude l_n is found equals 2.5×10^{-3}. This probability is extremely small and m should be of the order of $m > 185\,000$ to obtain probabilities larger than 0.99. Obviously, m should even be (much) larger when $N_p > 200$. This order of magnitude of m of 10^5 is not realistic in practical measurements where m will be much smaller. That means that l_n will generally be underestimated. The question now is to what extent l_n is underestimated. To answer that, we will determine the expectation value of the actual measured distance l_a

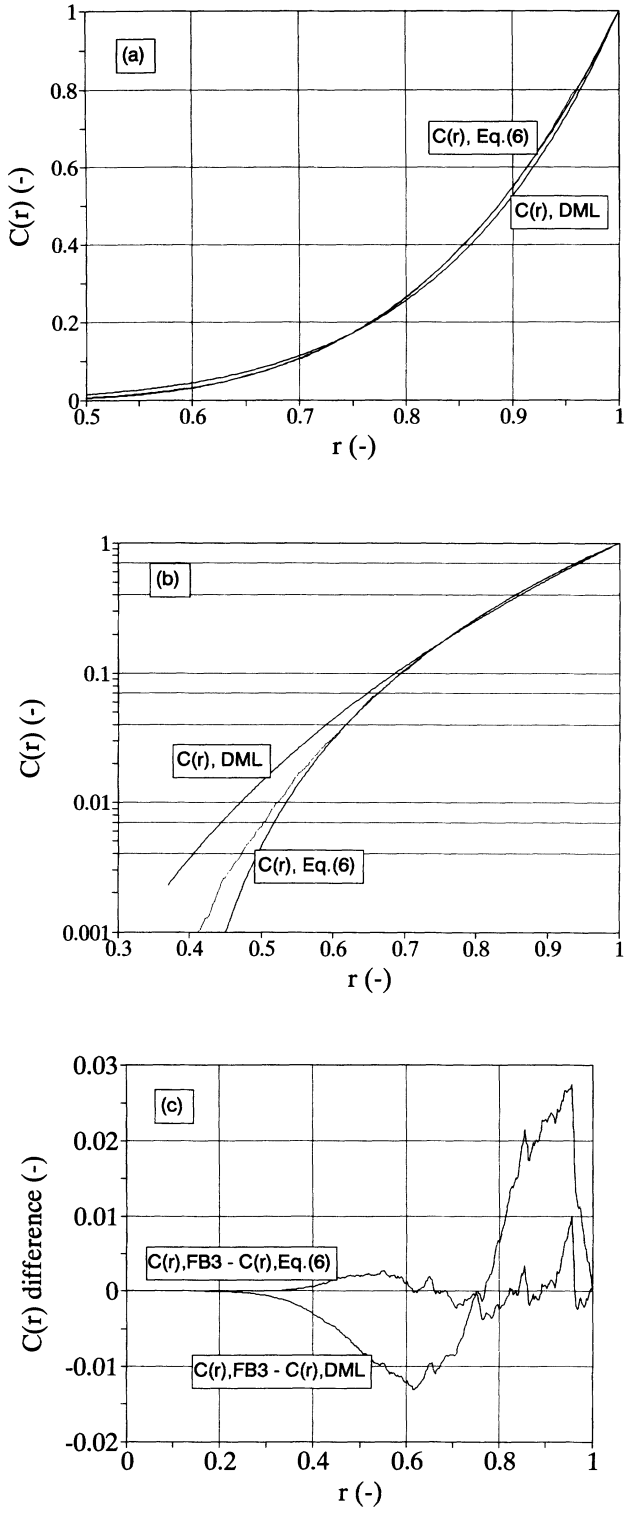

FIG. 1. The correlation integral of fluidized bed series FB3: comparison between the rescaled correlation integral function ($D=3.73$, $r_n=0.33$) and the maximum-likelihood result ($D_{ML}=6.05$); the dotted lines in (a) and (b) represent the correlation integral of experimental time series FB3. (a) $C(r)$ as a function of r. (b) $\ln[C(r)]$ as a function of r. (c) Differences between $C(r)$ of the experimental time series FB3 and the original and rescaled integral functions.

[l_a given in integer units, with $l_a < l_n$ and $l_n=(N_p-1)$].

First, from an evaluation of all possible combinations $N_p \times N_p$ of differences $a=(x_{i,k}-x_{j,k})$ between the noise parts of two corresponding vector *elements*, it can be derived that

$$P(a)=\frac{N_p-|a|}{N_p^2}\ , \tag{A2}$$

with

$$-(N_p-1)\le a\le(N_p-1)\ .$$

In other words, $P(a)$ is the probability that the difference between corresponding elements of two vectors equals a.

From this probability distribution, the cumulative distribution $P(A\le a)=\sum P(a)$ is derived as

$$P(A\le a)=\begin{cases}\dfrac{(N_p-|a|)(N_p-|a|+1)}{2N_p^2}\ , & \text{(A3a)}\\ -(N_p-1)\le a\le 0\\ \dfrac{2N_p^2-(N_p-a)(N_p-a-1)}{2N_p^2}\ , & \text{(A3b)}\\ 1\le a\le N_p-1\ .\end{cases}$$

The probability $P_m(A\le a)$ that with m successive corresponding vector elements a distance smaller than or equal to l_a is found is simply given by $P_m(A\le a)=[P(A\le a)]^m$. The probability that a distance *larger* than l_a is found is thus obtained as

$$P_m(A>a)=1-P_m(A\le a)^m\ .$$

The expectation value of the actual measured noise distance l_a will be $E(l_a)=l_n=\delta x_{\max}$ when m is infinite. When m is not infinite, the expectation value of l_a is obtained from

$$E(l_a)=\sum_{a=-(N_p-1)}^{N_p-1} aP_m(a)\ . \tag{A4}$$

$P_m(a)$ is the probability that *exactly* the noise distance l_a between two vectors is found. This probability is obtained from

$$P_m(a)=P_m(A\le a)-P_m(A\le a-1)\ ,$$

with, of course, $\sum P_m(a)=1$:

TABLE IV. The expectation value of r_a, Eq. (A6), at different values of the noise resolution (N_p) and the number of points (m) per time window in the case of independent and uniformly distributed noise.

N_p	$m=2$	$m=5$	$m=10$	$m=20$	$m=50$	$m=100$	$m=150$	$m=200$	$m=250$
10	0.257	0.526	0.681	0.797	0.903	0.954	0.974	0.985	0.991
20	0.245	0.501	0.649	0.761	0.864	0.917	0.940	0.953	0.963
50	0.238	0.487	0.630	0.739	0.840	0.892	0.915	0.929	0.939
100	0.236	0.482	0.624	0.732	0.832	0.884	0.907	0.921	0.930
200	0.235	0.480	0.621	0.728	0.828	0.879	0.902	0.916	0.925
500	0.234	0.478	0.619	0.726	0.826	0.877	0.900	0.913	0.923
1000	0.234	0.478	0.618	0.726	0.825	0.876	0.899	0.912	0.922
2000	0.233	0.477	0.618	0.725	0.824	0.876	0.898	0.912	0.921
5000	0.233	0.477	0.618	0.725	0.824	0.875	0.898	0.912	0.921
10000	0.233	0.477	0.618	0.725	0.824	0.875	0.898	0.912	0.921

$$P_m(a)=\begin{cases}\left[\dfrac{(N_p-|a|)(N_p-|a|+1)}{2N_p^2}\right]^m-\left[\dfrac{(N_p-|a-1|)(N_p-|a-1|+1)}{2N_p^2}\right]^m, & -(N_p-1)\le a\le 0 \quad \text{(A5a)}\\ \left[1-\dfrac{(N_p-a)(N_p-a-1)}{2N_p^2}\right]^m-\left[1-\dfrac{(N_p-a+1)(N_p-a)}{2N_p^2}\right]^m, & 1\le a\le(N_p-1)\ . \quad \text{(A5b)}\end{cases}$$

The normalized expectation value of the actual measured noise distance r_a is thus derived as

$$E(r_a)=\sum_{a=-(N_p-1)}^{N_p-1}\frac{a}{N_p-1}P_m(a)\ , \tag{A6}$$

where $r_a=l_a/l_n$ has been normalized with respect to the maximum noise distance $l_n=(N_p-1)$.

Some examples are given in Table IV. This table illustrates that the expectation value of r_a is already of the order of about 0.8 or higher at moderate values of the number of vector elements m (>50). As mentioned above, in practice the effective number of vector elements m_0 will be smaller than m with the obvious but rare lower bound at $m_0=1$. At this lower bound of $m_0=1$, the expectation value of r_a is, of course, zero; however, at $m_0=2$ or 5 values of, respectively, about 0.23 and 0.48 are already obtained. From Table IV it can be concluded that in this specific case of uniformly distributed noise, m_0 should be of the order of 50 or higher (preferably >100) to have $E(l_a)>0.8$. In practice, with an unknown noise distribution, a trivial approach would be, of course, to increase m to observe the effect on $E(l_a)$.

APPENDIX B

Another way of estimating the correlation dimension from the correlation integral of an attractor with known noise level l_n can be obtained from Takens [11]. Here the maximum-likelihood rule is applied to derive an unbiased estimate of the correlation dimension with minimal variance in case of a noise-free attractor. Takens's derivation can straightforwardly be applied when we use the substitution $u=(r-r_n)/(1-r_n)$ with $C(u)=u^D$, $0\le u\le 1$. This leads to

$$D_{ML,n}=\left[\frac{-1}{M}\sum_{i=1}^{M}\ln(u_i)\right]^{-1}, \tag{B1}$$

with

$$u_i=(r_i-r_n)/(1-r_n),\quad r_n\le r_i\le 1\ .$$

Using the same substitution, the relative standard error in $D_{ML,n}$ is obtained as $M^{-1/2}$; it is only dependent on the sample size M and independent of the estimate of the correlation dimension $D_{ML,n}$ and the noise level r_n. For example, the maximum-likelihood estimate of the correlation dimension will have a relative standard error of 10% if the sample size M is 100, while a relative standard error of 1% is obtained with a sample size of 10000. This would mean that, respectively, 100 and 10000 pairs of points have to be included with $l_n\le l\le l_0$.

APPENDIX C

D_{ML} and D will be related because they are invariants obtained from the same correlation integral. This relationship can be determined by calculating the expectation value of $\ln(r)$ that, using Eq. (8), will be equal to

$$E(\ln(r))=E\left[\int_0^1\ln(r)p(r)dr\right]=-D_{ML}^{-1}\ . \tag{C1}$$

We calculate the probability density $p(r)dr$ on [0,1] from the cumulative distributions $C_1(r)$ on $[0,r_n]$ and $C_2(r)$ on $[r_n,1]$, from which it follows that

$$E(\ln(r))=\int_0^1 \ln(r)p(r)dr=\ln(r)C(r)|_0^1+\int_0^1 \frac{C(r)}{r}dr$$

$$=\int_0^1 \frac{C(r)}{r}dr=\int_0^{r_n} r^{m-1}dr+\int_{r_n}^1 \frac{r_n^m+(1-r_n^m)\left[\frac{r-r_n}{1-r_n}\right]^D}{r}dr$$

$$=r_n^m\left[\frac{1}{m}-\ln(r_n)\right]+\frac{(1-r_n^m)}{(1-r_n)^D}\int_{r_n}^1 \frac{(r-r_n)^D}{r}dr\ . \qquad \text{(C2)}$$

For sufficiently large values of m, it is clear that $r_n^m \to 0$ and $1/m \to 0$ and Eq. (C2) can be approximated by

$$E(\ln(r))=-1/D_{ML}\simeq\frac{1}{(1-r_n)^D}\int_{r_n}^1 \frac{(r-r_n)^D}{r}dr\ . \qquad \text{(C3)}$$

The integral in Eq. (C3) cannot be solved analytically. From a numerical evaluation of the integral for a range of values of r_n and D, it is concluded that D_{ML}, D, and r_n are approximately related as

$$D=D_{ML}-r_n(D_{ML}+1)\Leftrightarrow D_{ML}=\frac{D+r_n}{1-r_n}\ . \qquad \text{(C4)}$$

[1] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, Chaos **3**, 127 (1993).

[2] T. Schreiber, Phys. Rev. E **48**, R13 (1993).

[3] H. Oltmans, J. C. Schouten, F. Takens, P. J. T. Verheijen, and C. M. van den Bleek, Delft University of Technology Internal Report Section CRE, 1993 (unpublished).

[4] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).

[5] F. Takens, in *Atas do 13° Colóqkio Brasileiro de Matemática* [Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, Brazil, 1983].

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

[7] A. M. Albano, A. Passamante, and M. E. Farrell, Physica D **54**, 85 (1991).

[8] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos **1**, 521 (1991).

[9] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, Physica D **51**, 52 (1991).

[10] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[11] F. Takens, in *Dynamical Systems and Bifurcations,* edited by B. L. J. Braaksma, H. W. Broer, and F. Takens, *Lecture Notes in Mathematics,* Vol. 1125 (Springer-Verlag, Berlin, 1984), p. 99–106.

[12] A. Ben-Mizrachi, I. Procaccia, and P. Grassberger, Phys. Rev. A **29**, 975 (1984).

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University, Cambridge, England, 1986).

[14] S. Ellner, Phys. Lett. A **133**, 128 (1988).

[15] J. C. Schouten, M. L. M. van der Stappen, and C. M. van den Bleek, in *Powders and Grains 93,* edited by C. Thornton (A. A. Balkema, Rotterdam, 1993), pp. 395–400.

[16] M. L. M. van der Stappen, J. C. Schouten, and C. M. van den Bleek, in *Proceedings of the 1993 International Conference on Fluidized Bed Combustion,* edited by Lynn N. Rubow [American Society of Mechanical Engineers (ASME), New York, 1993], Vol. 1, pp. 129–140.

[17] U. Hübner, W. Klische, N. B. Abraham, and C. O. Weiss, in *Measures of Complexity and Chaos,* edited by N. B. Abraham *et al.* (Plenum, New York, 1990), pp. 133–136.

[18] J. Theiler, Phys. Rev. A **34**, 2427 (1986).

[19] J. C. Schouten and C. M. van den Bleek, *RRCHAOS: A Menu-Driven Software Package for Chaotic Time Series Analysis* (Reactor Research Foundation, Delft, The Netherlands, 1994).