Practical implementation of error estimation for the correlation dimension

M. Frank,[1] G. Keller,[2] and R. Sporer[1]
[1]*Physikalisches Institut, Friedrich-Alexander-Universität Erlangen–Nürnberg, D-91058 Erlangen, Germany*
[2]*Mathematisches Institut, Friedrich-Alexander-Universität Erlangen–Nürnberg, D-91058 Erlangen, Germany*
(Received 18 July 1995; revised manuscript received 3 November 1995)

We discuss the statistical error in the calculation of the sample correlation integral from a finite sample of points. For this purpose we introduce an estimator of the covariance matrix of these estimators. The application of the method is described and it is shown that only small modifications to a standard Grassberger-Procaccia algorithm are necessary. Testing the method with 100 independent runs of the Hénon system, we show that the errors obtained for the correlation integrals are in good accordance with the sample error. These results are extended to the application to time-continuous systems, in our case the Lorenz system.

PACS number(s): 02.50.−r, 07.05.Kf

I. INTRODUCTION

The calculation of correlation dimensions via the sample correlation integral has gained widespread attention over the last decade. However, little is known about the error of the obtained quantities. Our goal is to provide reasonable confidence intervals by a generalized least squares method. We wish to emphasize that we deal with the statistical error of the correlation integral only. Although by doing this we consider error sources such as time correlation and lack of data points, other sources of error, such as edge effects, are not involved.

There have been previous attempts to give error estimates [1–3], but unfortunately these methods normally require a very large data set, or multiple realizations of a system. In practice, it is often not possible to meet these conditions. A system may not be stationary over a long time, or only a limited sample of information on the system may be available. Therefore, we wish to predict the confidence interval for the dimension estimate for a single limited data set.

In the second section, we describe the practical implementation of an algorithm to estimate the covariance matrix of the correlation integral on the basis of the pointwise correlation integrals. Only a moderate modification of the standard Grassberger-Proccacia algorithm [4] is necessary to obtain reasonable error estimates for the correlation integral and the dimension estimates, respectively.

The third section includes a test on numerically simulated data. At first, we produced 100 independent runs of the Hénon system, each containing 10 000 points (and a shorter version of 1000 points), and calculated the correlation integral, the correlation dimension, and the variance, respectively. Then we compared the variance of the 100 independent results with the variance estimated for a single run. The same procedure is then applied to the Lorenz system, i.e., a time-continuous system. We conclude that it is possible to obtain reasonable error estimates and confidence intervals for the correlation integrals and dimension.

II. ESTIMATION OF CORRELATION DIMENSION

The outline of the problem of finding statistical error estimates for the correlation dimension is as follows. It is a common technique to derive error estimates for the fit procedure, if the errors for the underlying data points are known; see, e.g., [5]. These data points in our case are the *sample correlation integrals* $C_N(r,d)$, which are estimated from the given time series. If the underlying data are produced by a sufficiently mixing (i.e., chaotic) dynamical system, then the difference between the *sample correlation integral* and the correlation integral of the distribution μ_d (which would result from an infinitely long sample of the attractor) is approximately normally distributed with zero mean. The variance can be estimated on the basis of the covariance matrix of the $C_N(r,d)$. For a wide class of systems with exponentially decreasing correlation [6], this result can be proven rigorously. At this point it should be mentioned that normality of the error distribution is not a premise of the least squares fit. Nonetheless, the key to error estimation of the *correlation dimension* is an understanding of the variance of the $C_N(r,d)$, which is given by the covariance matrix of $C_N(r,d)$.

First, we review the standard procedure to obtain a dimension estimate [4]. In general, we use a sample of an attractor **A**, which is produced by an embedding procedure. Given a scalar time series $(x_1,x_2,\ldots)$, this is typically achieved by the Takens delay method [7],

$$\underline{x}_i=(x_i,x_{i+\tau},\ldots,x_{i+(d-1)\tau})\in\mathbb{R}^d,$$

where τ is the so-called delay time and d denotes the embedding dimension.

Geometric and dynamical information of the distribution μ_d can be derived from the correlation integral

$$C_\mu(r,d):=\int\int h(\underline{x},\underline{x}')d\mu_d(\underline{x})d\mu_d(\underline{x}'), \tag{1}$$

where $h(\underline{x},\underline{x}')=\theta(\|\underline{x}-\underline{x}'\|-r)$ and θ is the Heaviside function. $\|\underline{x}-\underline{x}'\|$ denotes the Euclidian (or any other suitable) distance of $\underline{x}$ and $\underline{x}'$. Quantities given by Eq. (1) are estimated by

$$C_N(r,d)=\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}h(\underline{x}_i,\underline{x}_j). \tag{2}$$

For the relation of $C(r,d):=\lim_{N\to\infty}C_N(r,d)$ to r and d, Grassberger and Procaccia [8] found that in many cases there are real numbers $\nu>0$ and $K_2>0$ such that $C(r,d)\approx r^{\nu}\exp^{-d\tau K_2}$ as $r\to 0$ and $d\to\infty$, leading to

$$\log C(r,d)=\nu\,\log r-hd+C+\delta(r,d); \tag{3}$$

ν is called the correlation dimension, h is an entropylike quantity related to K_2, C is a constant term, and $\delta(r,d)$ is a (hopefully small) nonlinear remainder. This model, although appropriate if $\|\underline{x}-\underline{x}'\|$, denotes the maximum norm, leads in the case of a Euclidean distance norm to improved convergence for h by replacing $\log r$ with $\log(r/\sqrt{d})$ [9].

Fitting a model (3) to given data points (2) is typically solved by a least squares fit. This procedure not only provides estimates for the parameters ν, h, and C but also gives estimates for the variance of the fitted parameters:

$$\mathrm{var}(\nu,h,C)=[(M^tM)^{-1}M^t]V[(M^tM)^{-1}M^t]^t. \tag{4}$$

Here the matrix V denotes the covariance matrix of the logarithmic correlation integrals, and $M=(\log \underline{r},-\underline{d},\underline{1})$ is frequently called the *design matrix*. To obtain error estimates for the correlation dimension ν and the entropy term h, we have to find an estimator for the covariance matrix V.

Finding an estimator for the covariance of $\log C_N(r,d)$ is a standard procedure of statistics. At this point we want to give the result (for details see [10]). The estimator for the covariance of the correlation integral is given in terms of the two matrices $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$:

$$\hat{\mathbf{K}}=[1+4(1+2t)N^{-1}]4N^{-1}\hat{\mathbf{P}}-N^{-2}\hat{\mathbf{Q}}. \tag{5}$$

In the next section, we describe the practical calculation of the matrices $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$. The relation between V in Eq. (4) and $\hat{\mathbf{K}}$ in Eq. (5) is given in Eq. (10).

III. IMPLEMENTATION OF THE METHOD

The basis of the estimation of the correlation integral $C_\mu(r,d)$ [Eq. (1)] is the *sample correlation integral* $C_N(r,d)$ [Eq. (2)]. If the underlying system possesses a nonvanishing time correlation, the term $C_N(r,d)$ has to be corrected by excluding pairs (i,j) that are closer in time than a specific time t. This correction, which was first introduced by Theiler [11], leads to a modified version of the sample correlation integral

$$U_N(r,d)=\frac{1}{(N-2t)(N-2t-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\|j-i|>2t}}^{N}h(\underline{x}_i,\underline{x}_j). \tag{6}$$

Note that the second sum represents the *pointwise correlation integrals*

$$B_N(r,d,i)=\frac{1}{N-2t-1}\sum_{\substack{j=1\\|j-i|>2t}}^{N}h(\underline{x}_i,\underline{x}_j). \tag{7}$$

These pointwise correlation integrals will be the basis for the estimation of the covariance matrix of $C_\mu(r,d)$. As described in the preceding section, the estimator $\hat{\mathbf{K}}$ includes the yet unknown matrices $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$. $\hat{\mathbf{P}}$ especially can be described easily by $U_N(r,d)$ and $B_N(r,d,i)$:

$$\begin{aligned}\hat{P}(r_1,d_1,r_2,d_2)&=\sum_{k=-t}^{t}\frac{1}{N-k}\sum_{i=1}^{N-k}[B_N(r_1,d_1,i)B_N(r_2,d_2,i+k)\\&\quad-U_N(r_1,d_1)U_N(r_2,d_2)].\end{aligned} \tag{8}$$

The notation r_i, d_i illustrates the fact that we want to fit the model (3) to a scaling region, which expands over a range of radii r_i, $i=1,\ldots,s_r$, and embedding dimensions d_j, $j=1,\ldots,s_d$, that have yet to be chosen. For example, a scaling region over four dimensions with ten radii each will leave us with a (40×40) covariance matrix, and, respectively, a (40×40) matrix for $\hat{\mathbf{P}}$.

As we mentioned earlier, the time t denotes a distance in time after which pairs $(\underline{x}_i,\underline{x}_j)$ are independent for $|i-j|>t$. The second sum in Eq. (8) is therefore some sort of ''correlation term'' between the pointwise correlation integrals $B_N(r_1,d_1,i)$ and shifted pointwise correlation integrals $B_N(r_1,d_1,i+t)$. It takes the form of an autocorrelation function of the $B_N(r,d,i)$. Here again the importance of term t becomes obvious. For a reasonable estimate of the covariance matrix, it is necessary to sum up all the right-hand sum terms in (8), which have not yet saturated to zero. With increasing time distance k, however, these terms will become smaller and smaller, which justifies the fixing of a ''cutoff'' time t. In practice the ''correlation term'' may oscillate with decreasing amplitude around zero for a long time. However, we observe that the partial sum $\hat{P}(r_1,d_1,r_2,d_2)$ will converge faster because of the decreasing amplitude of the right-hand sum $\sum_{i=1}^{N-k}[B_N(r_1,d_1,i)B_N(r_2,d_2,i+k)-U_N(r_1,d_1)U_N(r_2,d_2)]$. The matrix $\hat{\mathbf{P}}$ therefore represents the sum over the ''correlation term'' of the pointwise correlation integrals for $|k|\leqslant t$, while all other terms $|k|>t$ do not significantly change the sum (8). To conclude, the calculation of $\hat{\mathbf{P}}$ is straightforward. It consists simply in the evaluation of the pointwise correlation integrals and the execution of sum (8).

The still missing correction term $\hat{\mathbf{Q}}$ can be estimated by

$$\begin{aligned}\hat{Q}(r_1,d_1,r_2,d_2)=\frac{1}{\tilde{\pi}(N,t)}\sum_{\substack{i,j=1\\|j-i|>3t}}^{N}\sum_{p,q=-t}^{t}[&h^{(r_1,d_1)}\\&\times(\underline{x}_i,\underline{x}_j)h^{(r_2,d_2)}(\underline{x}_{i+p},\underline{x}_{j+q})\\&-U_N(r_1,d_1)U_N(r_2,d_2)],\end{aligned} \tag{9}$$

where $\tilde{\pi}(N,t)=(N-3t-1)(N-3t)$ denotes the number of all relevant pairs in (9). We concentrate on the explanation of the diagonal terms in (9). Again, we concentrate on the right-hand sum. $\hat{Q}(r_1,d_1,r_1,d_1)$ measures the joint probability that if $|\underline{x}_i-\underline{x}_j|$ is less than r_1, then also the distance $|\underline{x}_{i+p}-\underline{x}_{j+q}|$ is less than r_1. In the case of a dynamical system, the point $i+1$ is the ''next'' point on the trajectory; i.e., if the pair $(\underline{x}_i,\underline{x}_j)$ is close, it is likely that also the respective ''neighbors'' on the trajectory $(\underline{x}_{i+1},\underline{x}_{j+1})$ are also close together. The exact behavior is for the diagonal elements $p=q$, described by the Lyapunov exponents. The restriction on the

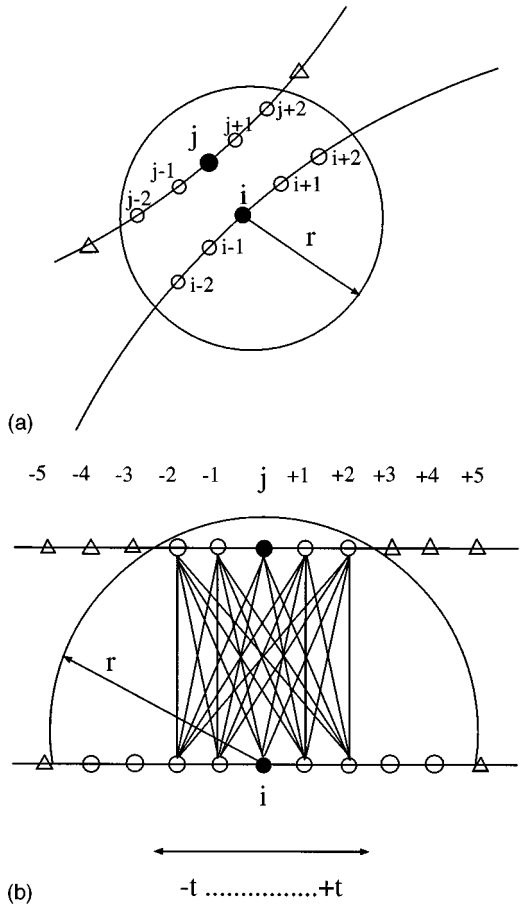

FIG. 1. (a) Points $(\underline{x}_i,\underline{x}_j)$ on different trajectories, within radius r. Respective neighborhood of $\underline{x}_{i,j}$: circles, points that contribute to the correction term; triangles, points that have a distance greater than r and therefore do not contribute to the correction term. (b) Lines indicate the pairs that contribute to the correction term within the chosen ''cutoff'' time t.

diagonal terms of (9) makes it easy to evaluate $\hat{Q}(r_1,d_1,r_1,d_1)$. (The covariance matrix that is obtained by that procedure has the form of $\hat{\mathbf{K}}$ on the diagonal and follows the structure of $\hat{\mathbf{P}}$ everywhere else. It is denoted $\hat{\mathbf{K}}'$.) Whenever a pair $(\underline{x}_i,\underline{x}_j)$ is found to be closer than r, i.e., whenever it will be counted in the original calculation of $B_N(r,d,i)$, the ''surrounding'' $(\underline{x}_{i+p},\underline{x}_{i+q})$ must be searched for pairs that fulfill the condition $|\underline{x}_{i+p}-\underline{x}_{j+q}|<r$. The described situation is illustrated in Fig. 1. This is a dynamical interpretation of $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$, but their justification is entirely based on statistical considerations [10].

We have now obtained estimates for $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ by making use of the standard procedure to estimate the pointwise correlation integral $B_N(r,d,i)$. Finally, we have to get the covariance matrix of the logarithm of $C_N(r,d)$. A Taylor expansion leads to the covariance matrix of $\log[C_N(r,d)]$,

$$\hat{V}'(r_1,d_1,r_2,d_2)=\frac{\hat{K}'(r_1,d_1,r_2,d_2)}{U_N(r_1,d_1)U_N(r_2,d_2)}. \qquad (10)$$

This covariance matrix is the basis for a least squares fit. The errors in the model parameters ν and h will follow from Eq. (4).

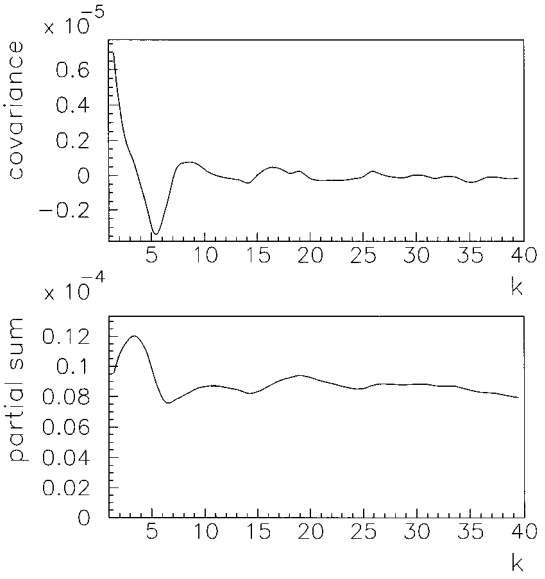

FIG. 2. Above: covariance $\Sigma_{i=1}^{N-k}H_{1,i}^{(u)}H_{1,i+k}^{(u)}$ for radius 25. Below: partial sum $\hat{P}_{u,u}$ for radius 25.

IV. RESULTS

A. Test with independent realizations

We now come to numerical tests of the method. First we want to test the procedure by comparing the results with a sample of 100 independent runs of two test systems. These are the Hénon system (standard parameters $a=1.4$ and $b=0.3$) and the Lorenz system (standard parameters $S=10$, $R=28$, and $b=\frac{8}{3}$). Our goal is to compare the mean and standard deviations of the 100 runs with the estimated values obtained from the individual runs. At first we use noise-free data.

1. Hénon system

For the Hénon system we compute the sum $\hat{P}(r_1,d_1,r_2,d_2)$ [see Eq. (8)] in order to find a suitable value for t. As seen in Fig. 2, the correlation is practically zero from $t=10$ on. Also, the sum of the correlation terms saturates to a fixed value. This is not surprising, as the underlying system shows no longtime correlations, i.e., the same behavior can be observed for the autocorrelation of the time series itself.

Then we calculate the correlation integral $\hat{P}(r_1,d_1,r_2,d_2)$ and $\hat{Q}(r_1,d_1,r_2,d_1)$ (which we from now on denote simply $\hat{P}$ and $\hat{Q}$) with $N=10\,000$ and $N=1\,000$ and r_i for $i=15$ to 35 (giving an effective range of $r_{15}=0.001$ to $r_{35}=0.067$ in units of global attractor size) for each individual run. The embedding dimension for the 100 runs is fixed to 4. Since the dimension d is now a fixed parameter, the term hd in our model is included in the constant C. For each of the 100 single runs j, we calculate the estimator for the logarithms of the correlation integrals $Z_j(r_i)=\log C_j(r_i)$ and their estimated standard deviations $\hat{\sigma}_j(r_i)$. The average of the standard deviations is denoted $\bar{\sigma}(r_i)$. The sample standard deviation of the $Z_j(r_i)$ is denoted $S_i(Z)$. This is the statistical fluctuation of the $Z_j(r_i)$ due to the limited data sets. In Fig. 3 the values for $S_i(Z)$ are compared with the uncorrected values proportional to $\hat{P}$ and

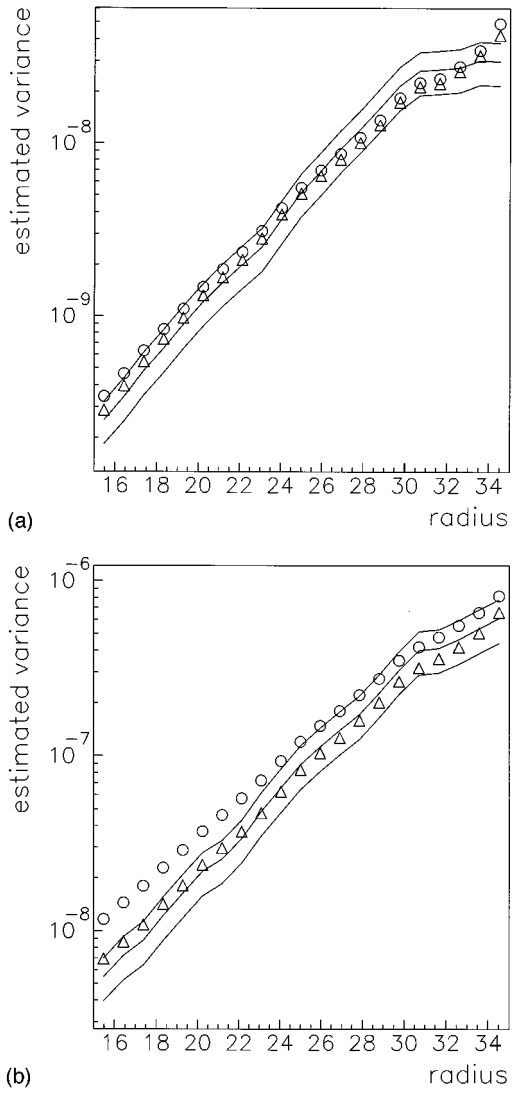

FIG. 3. (a) Sample variance S (line), uncorrected covariance (circles), and corrected covariance (triangles) for 10 000 points. (b) Same as (a) for 1000 points (Hénon).

with the corrected estimators, the correction term being $\hat{Q}$. Additionally we give the 95%-confidence region $S_i(Z) \pm 1.96\sigma[S_i(Z)]$. Figure 3 shows clearly that the calculated estimators for the variance of $\log C_j(r_i)$ are very close to the variation between the 100 values of $\log C_j(r_i)$, which is caused by the statistical uncertainty due to the limited data length. On the average, the corrected values (triangles) are closer to the sample standard deviation, and except for a few radii, all lie in the 95%-confidence interval. This is not the case for the uncorrected values, which are too large, especially for small radii, and do not lie in the error interval. This effect is naturally more obvious for the calculations based on only 1000 points. We derive from this that the estimated variance terms are able to describe the statistical fluctuation of the correlation integral and that the introduced correction term is necessary to give a good estimate.

We now come to the error estimates for the correlation dimension, i.e., the slope of the correlation integrals. The average dimension for the two scaling regions is $\overline{\nu}_{15,25}=1.194$ and $\overline{\nu}_{25,35}=1.208$ for 10 000 points [a standard Grassberger-Procaccia (GP) algorithm yields 1.21 [12]]. The standard deviation of the 100 runs is $S(\nu)_{15,25}=0.0113$ and $S(\nu)_{25,35}=0.0077$. This has to be compared with the mean of the estimated variance terms. These are $\overline{\sigma}_{15,25}=0.0120$ and $\overline{\sigma}_{25,35}=0.0081$ for the corrected estimator. The values for the variance are, in both scaling regions, very close. However, this procedure has its drawbacks. The estimated values are the mean over 100 runs. Although they are a very good approximation of the statistical error for the correlation dimension, we are interested in the behavior of the estimate based on single runs. Therefore, we calculate for each single run a 95%-confidence interval given by $J_j=[\hat{\nu}_j-1.96\hat{\sigma}_j(\nu),\ \hat{\nu}_j+1.96\hat{\sigma}_j(\nu)]$, with $\hat{\sigma}_j(\nu)$ the estimated standard deviation. We do not know the real value of ν; therefore, we use the mean $\overline{\nu}$ of all 100 $\hat{\nu}_j$ as the asymptotic value of the correlation dimension. Then we count the number of times that the $\overline{\nu}$ lies in our individual interval J_j. If the calculated estimates for the variance of $\hat{\nu}$ are reasonable, we expect the mean $\overline{\nu}$ to lie in these confidence intervals 95 times. Table I shows these probabilities for two radius regions $r_{15}-r_{25}$ and $r_{25}-r_{35}$ and for runs based on 1 000 and 10 000 points.

TABLE I. Fraction $\overline{\nu} \in J_j$ in different radius regions and corrected and uncorrected estimators for J_j (Hénon system).

	$N=1\,000$	$N=1\,000$	$N=10\,000$	$N=10\,000$
Radius region	$r_{15}-r_{25}$	$r_{25}-r_{35}$	$r_{15}-r_{25}$	$r_{25}-r_{35}$
Uncorrected	0.99	0.95	0.98	0.98
Corrected	0.93	0.93	0.96	0.96

As we can see, the uncorrected values have a probability for $\overline{\nu} \in J_j$ that is very close to 100% because the estimated confidence intervals J_j are too pessimistic, i.e., too large. This is especially obvious for $N=1000$ and small radii. The corrected values, on the other hand, show very good agreement with the 95% level for $N=10\,000$. For the case $N=1000$, the observed values of 93% must be accepted from a statistical viewpoint. We conclude that (for $N=10\,000$) the estimated confidence interval will include the average value of ν with a probability of 96%, i.e., the estimated confidence interval of a single run well represents the fluctuation in the hundred independent $\hat{\nu}_j$. It is therefore possible to give an error estimate and a confidence interval for the correlation dimension.

2. Lorenz system

The same procedure that was applied to a discrete map will now be extended to a time-continuous system, the Lorenz equations. By applying the same procedure as in Fig. 2 we choose $t=20$. The 100 independent runs are embedded in a six-dimensional embedding space. Then the same procedure as for the Hénon data is repeated. Again we show in Fig. 4 the sample standard deviation $S_i(Z)$ of the 100 $Z_j(r_i)$, together with the uncorrected values corresponding to $\hat{P}$ and the $\hat{Q}$ corrected by estimators. The radius region for the radii is now r_i, with $i=15,\ldots,30$ ($r_{15}=0.0016$ and $r_{30}=0.025$). The results show that the method also yields good error estimates for the variance of the correlation integral for the Lorenz system.

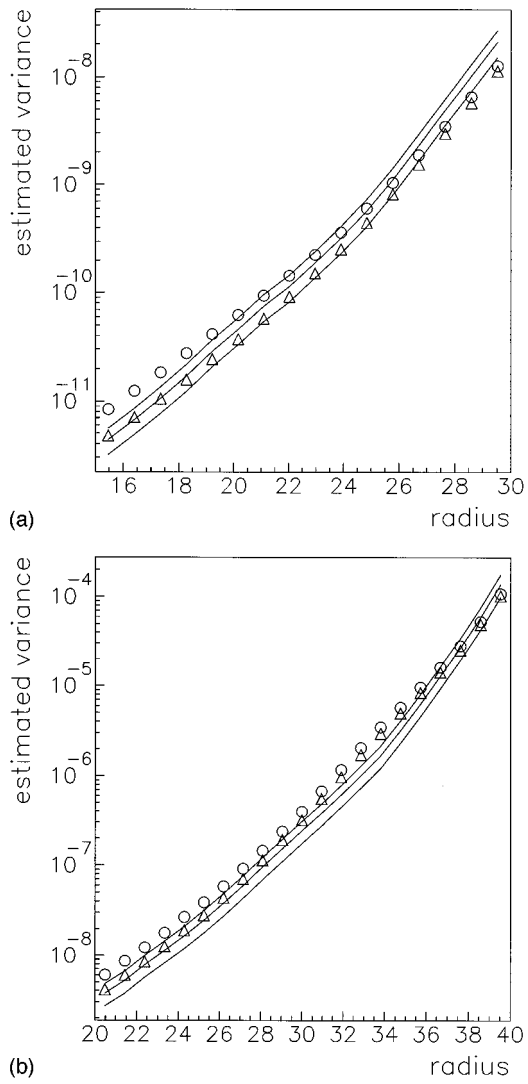

FIG. 4. (a) Sample variance S (line), uncorrected covariance (circles), and corrected covariance (triangles) for 10 000 points. (b) Same as (a) for 1000 points (Lorenz).

Again we determine the estimated slope of the 100 runs. The average dimension for the two radius regions is $\overline{\nu}_{15,25}=2.084\pm0.053$ and $\overline{\nu}_{20,30}=2.067\pm0.024$ for 10 000 points (a standard GP algorithm yields 2.06 [12]). The mean of the estimated variance for the slope is $\overline{\sigma}_{15,25}=0.057$ and $\overline{\sigma}_{20,30}=0.023$. This is in good agreement with the statistical fluctuation.

To test the estimates for single runs, we again investigate the 95%-confidence interval $J_j=[\hat{\nu}_j-1.96\hat{\sigma}_j(\nu), \hat{\nu}_j+1.96\hat{\sigma}_j(\nu)]$. Table II shows the probabilities of $\overline{\nu\in J_j}$ for two radius regions, $r_{15}-r_{25}$ and $r_{20}-r_{30}$, $N=10\,000$, and $r_{20}-r_{30}$ and $r_{30}-r_{40}$, $N=1\,000$.

TABLE II. Fraction $\overline{\nu\in J_j}$ in different radius regions and corrected and uncorrected estimators for J_j (Lorenz system).

	$N=1\,000$	$N=1\,000$	$N=10\,000$	$N=10\,000$
Radius region	$r_{20}-r_{30}$	$r_{30}-r_{40}$	$r_{15}-r_{25}$	$r_{20}-r_{30}$
Uncorrected	0.97	0.98	1.00	0.98
Corrected	0.92	0.93	0.95	0.95

Again the corrected values for 10 000 points are in very good agreement with the 95% level, i.e., $\overline{\nu\in J_j}$ in 95% of the 100 runs, which corresponds very well to the 95%-confidence level of J_j. The uncorrected values result in a probability of $\overline{\nu\in J_j}$, which is too high because the respective confidence interval is too large.

To conclude, the method described is capable of giving good estimates of the statistical fluctuation of the correlation integral and the correlation dimension for both the Hénon system and the Lorenz system. It especially allows confidence bounds for the estimated values.

B. Influence of noise

To determine the influence of noise, we use a model that is a slight modification of Eq. (3). For pure random data, (3) can be expanded as follows. Given that $\|\underline{x}-\underline{x}'\|$ denotes the maximum norm of $\underline{x}-\underline{x}'$, then $C(r,d)$ equals $[C(r,1)]^d$ such that

$$\log C(r,d)=d\ \nu_1[\log r+O(\log r)] \quad \text{as } r\to 0, \tag{11}$$

where ν_1 denotes the correlation dimension of the distribution μ_1 of x_i; see [13]. The equations above can be combined in the following linear model of $\log C(r,d)$:

$$\log C(r,d)=\nu\ \log r-hd+\nu_1 d\ \log r+C'+\delta'(r,d). \tag{12}$$

The influence of random behavior is now measured by the term ν_1. For $\nu_1\approx 0$ the interpretation is *no noise*; $\nu_1\neq 0$ means *presence of noise*. To fit all parameters in (12), especially to determine the entropy term h, we calculate the covariance matrix for the Hénon system for radius $r_{15}=0.001,\ldots,r_{25}=0.011$ and dimension $d=4,\ldots,7$. Only single runs of 10 000 points are performed. We compare two situations: no noise and 1% noise added after the iteration (measurement noise). The ''noise'' was added as uniform distributed random numbers.

For the case without noise, we find that ν_1 is not significantly different from 0 (interpretation: no noise). The intervals denote the 95%-confidence region for the given values. We find that $\nu_1\in[-0.013,0.025]$. Since the value for ν_1 is not different from zero, we exclude ν_1 from the model. The respective confidence intervals after that are $\nu\in[1.077, 1.188]$ and $h\in[0.277,0.313]$. Note especially the possibility of testing the null hypothesis: ''h is different from 0.''

For the case of 1% measurement noise, the intervals are $\nu\in[-0.216,0.224]$, $h\in[2.717,3.167]$, and $\nu_1\in[0.531, 0.631]$. ν_1 is clearly different from zero. The entropy term h is now larger than in the situation without noise, describing the higher information production of the system in the presence of noise. However, it should be noted that the model is not able to distinguish between the entropy produced by the noise and the entropy produced by the deterministic system. In the case $\nu_1\neq 0$, there is no proper interpretation of h as a possible entropy. The detection of $\nu_1\neq 0$ is an indication that $\log C_N(r,d)$ follows the model assumption (11). Since the situation in (11) is produced by random behavior, one can therefore argue that the dependency of $\log C_N(r,d)$ on $\log r$ and d is produced by some sort of random distribution in the data.

We conclude that the method described above is capable of detecting even a moderate level of random behavior (noise) in the data. Furthermore, there is the possibility of testing the assumption ''h is different from 0.''

V. CONCLUSION

We present a method to determine the statistical precision of the correlation integral and the correlation dimension. Compared to other approaches providing confidence intervals (e.g., [1,2]), we neither need parametric assumptions on the dynamical system nor very large sample sizes to obtain reasonable estimates.

The algorithm is easy to implement and requires only moderate modifications to an existing standard Grassberger-Proccacia approach. The main difference lies in the additional terms of $\hat{Q}$. These can be determined by a separate routine during calculation of the correlation integral.

We have shown that the results for a single run well represent the error in calculating 100 independent samples of the same system. The 95%-confidence interval, both for the correlation dimension and for the correlation integral itself, gives reliable bounds for the statistical error. Therefore, the method presented is a good tool to determine the statistical error and provides the basis for additional tests, e.g., detection of noise or testing whether the entropy is zero. Also, a test on linearity of the model [14] can be performed using a χ^2 test based on the least squares fit. Such a test is extremely sensitive to even moderate nonlinearities and distinguishes between systematic and statistical error [10]. Restrictively, we note that by ''systematic'' we mean the error in the model assumption. We do not address geometric effects [15] or other sources of error.

Finally, we want to emphasize that the procedure is capable of analyzing experimental data. As we have shown for the Lorenz system, our approach gives good results and provides the possibility of determining a confidence interval for the estimated quantities.

To conclude, we have presented a tool to estimate the statistical error of the correlation integrals and the respective dimension for a single data set of moderate length. It provides a wide range of applications and is therefore a good tool for the experimentalist.

ACKNOWLEDGMENT

We wish to thank Professor Dr. W. Kreische for his encouraging support.

[1] K. Judd, Physica D **56**, 216 (1992).
[2] J. B. Ramsey and H.-J. Yuan, Phys. Lett. A **134**, 287 (1989).
[3] J. Theiler, Phys. Rev. A **41**, 3038 (1990).
[4] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).
[5] C. D. Cutler, in *Nonlinear Time Series and Chaos, Vol. I: Dimension Estimation and Models*, edited by H. Tong (World Scientific, Singapore, 1993).
[6] M. Denker and G. Keller, J. Stat. Phys. **44**, 67 (1986).
[7] F. Takens, Lect. Notes Math. **898**, 366 (1980).
[8] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).
[9] M. Frank *et al.*, Physica D **65**, 359 (1993).
[10] G. Keller and R. Sporer (unpublished).
[11] J. Theiler, Phys. Rev. A **34**, 2427 (1986).
[12] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).
[13] W. A. Brock, W. D. Dechert, and J. A. Scheinkman (unpublished).
[14] G. Keller and M. Künzle(unpublished).
[15] M. A. H. Nerenberg and C. Essex, Phys. Rev. A **42**, 7065 (1990).