RAPID COMMUNICATIONS

PHYSICAL REVIEW E VOLUME 48, NUMBER 1 JULY 1993

Determination of the noise level of chaotic time series

Thomas Schreiber*

Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

(Received 5 April 1993)

We propose a method to determine the amount of measurement noise present in a chaotic time series. If the data are embedded in a space of higher dimension than that strictly required to reconstruct the dynamics, the extra dimensions are dominated by the noise, which results in a certain shape of the correlation integral. For the case in which only Gaussian noise is present, this shape can be calculated analytically as a function of the noise level. Thus the noise level can be obtained from a simple function fit. The analytical result also shows that a noise level of more than 2% will obscure any possible scaling of the correlation integral and thus makes it impossible to estimate the correlation dimension.

PACS number(s): 05.45.+b, 06.50.−x

It is commonplace that every experiment is subject to measurement error. Equally obvious is that this uncertainty is desired to be as small as possible. But at first we have to know how severe the uncertainty actually is.

The measurement we consider here is a chaotic time series obtained from some nonlinear phenomenon (see [1] for a review on nonlinear time series analysis). Nonlinear noise-reduction methods [2] are available to suppress measurement noise to a considerable degree. But even the remaining noise will have an influence on any application, be it predictions, control, or the estimation of characteristic quantities. In some cases we know the difficulties that have to be expected depending on the nature and amplitude of the noise. For many experiments one can assume Gaussian noise with short correlations plus discretization error. Even when this is an adequate description of the nature of the noise, the amplitude of the Gaussian contribution can only be vaguely guessed from the experimental setup.

In this Rapid Communication we wish to present a method for obtaining a reliable estimate of the amplitude of the noise. The considered data sets have to meet some relatively weak requirements of length and complexity. We assume the noise to follow a Gaussian distribution, but we will be able to detect deviations from this assumption. As an application, we will use the method as a consistency check for nonlinear noise reduction [4] on experimental Taylor-Couette flow data. The data were provided by Buzug, Reimers, and Pfister [5].

The method presented in this Rapid Communication makes use of the correlation integral $C(\epsilon)$ introduced by Grassberger and Procaccia [6] in order to compute the correlation dimension of a strange set. Let

$$C(\epsilon)=\frac{1}{N^2}\sum_{i,j}\Theta(\epsilon-\|\mathbf{x}_i-\mathbf{x}_j\|) \tag{1}$$

denote the fraction of pairs of points on the attractor whose distance apart is less than ϵ. In the limit as $\epsilon\to 0$ and $N\to\infty$, and in the absence of noise, we have

$$C(\epsilon)\sim\epsilon^d , \tag{2}$$

where d is the correlation dimension of the attractor. Equation (2) can also be written as

$$d=\lim_{\epsilon\to 0}\lim_{N\to\infty} d(\epsilon),\quad d(\epsilon)=\frac{d}{d\ln\epsilon}\ln C(\epsilon) , \tag{3}$$

which is a more useful definition in practice.

Typically, one reconstructs the attractor in a suitable space and computes $C(\epsilon)$ and its slope $d(\epsilon)$ as functions of ϵ. When interested in the correlation dimension d, one would look for a range of ϵ values where $d(\epsilon)$ is relatively constant. Here we rather wish to study $d_m(\epsilon)$ as a function of ϵ and the embedding dimension m to extract information about the noise in the system.

One can distinguish four different types of behavior of $d_m(\epsilon)$ for different regions of length scales ϵ. For small ϵ (region I) the lack of data points is the dominant feature. Therefore, the values of $d_m(\epsilon)$ are subject to large statistical fluctuations. On the other hand, if ϵ is of the order of the size of the entire attractor (region IV), no scale invariance can be expected.

In between, we can distinguish two regions. Region II is dominated by the noise in the data: the reconstructed points are not restricted to the fractal structure of the attractor but fill the whole phase space available; thus we expect $d_m(\epsilon)\approx m$. Between regions II and IV we have region III, where the proper scaling behavior of the attractor may be observed: $d_m(\epsilon)\sim d$. Assume we have a scalar time series that originates in a low-dimensional dynamical system with a box-counting dimension [7] d_{box}. Then, according to Sauer, Yorke, and Casdagli [8], the correlation dimension d is preserved for a reconstruction using r delay coordinates [9], as long as r is at least equal to d_{box}. In this region we therefore expect $d_m(\epsilon)$ to increase with m until it saturates to $d_m(\epsilon)\approx d$ for $m\geq d_{\text{box}}$.

The algorithm we propose makes use of the fact that, for higher values of m, regions II and III will be very distinct: we will see a crossover between $d_m(\epsilon)\approx m$ below and $d_m(\epsilon)\approx d$ above the noise level. Later, we will analytically derive a formula describing this crossover.

But let us first establish the algorithm.

Say we use a reconstruction in $m > r$ dimensions of a signal which could be faithfully reconstructed using only the first r coordinates. Below we will show then that in regions II and III all the curves $[d_m(\epsilon)-d_r(\epsilon)]/(m-r)$ have the same functional form [Eq. (16)], parametrized by the noise level σ:

$$\frac{d_m(\epsilon)-d_r(\epsilon)}{m-r}=g\left[\frac{\epsilon}{2\sigma}\right], \quad g(z)=\frac{2}{\sqrt{\pi}}\frac{ze^{-z^2}}{\operatorname{erf}(z)} . \tag{4}$$

This formula can be used directly to determine σ: $d_m(\epsilon)$ and $d_r(\epsilon)$ can be obtained numerically from $C_m(\epsilon)$ and $C_r(\epsilon)$, respectively. An estimate for the noise level is given by the value of σ for which the function on the right-hand side fits best to $d_{m,r}(\epsilon)$.

To illustrate the use of Eq. (4), we generate a scalar time series $x_n=\operatorname{Re}(z_n)$ by measuring the real part of 100 000 iterates of the Ikeda map [10] $z_{n+1}=1+0.9z_n\exp[0.4i-6i/(1+|z_n|^2)]$. We add Gaussian uncorrelated measurement error of amplitude 0.005, corresponding to 1% of the total variance. We use *singular value decomposition* (SVD) [11,12] to consecutively embed the data in $m=2,\ldots,10$ dimensions using the first 2 to 10 singular vectors as basis vectors. Technically, we compute the covariance matrix $\Gamma_{ij}=\langle x_{n+i}x_{n+j}\rangle-\langle x_{n+i}\rangle\langle x_{n+j}\rangle$, where $\langle\ \rangle$ denotes the average over all iterates n. Then we use the orthonormal eigenvectors of Γ corresponding to the m largest eigenvalues as a new basis in m dimensions.

In these coordinates we compute the correlation integral $C_m(\epsilon)$ for $m=2,\ldots,10$ and values of $\{\epsilon_i\}$ in the range 0 to 0.025. We choose $r=2$ [13] and compute $d_{m,r}(\epsilon)$, $m=3,\ldots,10$. Further, we use Brent minimization [14] to find the value of σ for which

$$\sum_{m,i}\left[d_{m,r}(\epsilon_i)-\frac{\epsilon_i\exp\left[-\frac{1}{4\sigma^2}\epsilon_i^2\right]}{\sigma\sqrt{\pi}\operatorname{erf}\left[\frac{\epsilon_i}{2\sigma}\right]}\right]^2 \tag{5}$$

is minimal.

Figure 1 shows the curves $d_{m,r}(\epsilon)$ together with the best function fit, $\sigma_{\text{fit}}=0.005\,16$, which agrees very well with the actual noise level, $\sigma=0.005$. Observe also that the fluctuations for small ϵ are purely statistical [15] and thus in principle do not harm the least-squares fit. To suppress the largest fluctuations we exclude values of $d_{m,r}(\epsilon)$ obtained from $N^2C_m(\epsilon)<50$. An application of the method to an experimental data set will be discussed below.

To derive the formula (4) we compute the shape of $d_m(\epsilon)$ at the crossover between regions II and III. Recall that we reconstruct a signal in $m>r$ dimensions, although the first r coordinates would suffice. We denote the projection of a vector $\mathbf{x}$ in the m-dimensional phase space to the subspace spanned by the first r coordinates as $\bar{\mathbf{x}}$ and the projection to its orthogonal complement by $\tilde{\mathbf{x}}$. The clean signal can be thought of as drawn from a distribution

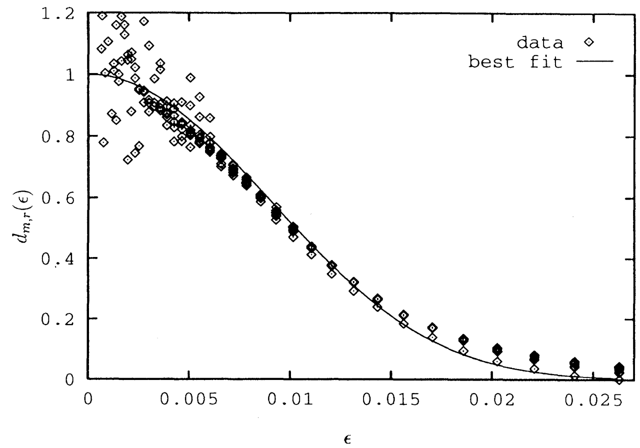

FIG. 1. Normalized effective dimensions $d_{m,r}(\epsilon)$ for 100 000 iterates of the Ikeda map with Gaussian noise of amplitude 0.005 added. Apart from the expected statistical fluctuations for small ϵ all the curves show the shape given in Eq. (16). This function is shown (solid line) with the parameter $\sigma=0.005\,16$, which gave the best fit.

$$\nu(\mathbf{x})=\nu_r(\bar{\mathbf{x}})\delta(\tilde{\mathbf{x}}-\tilde{\mathbf{x}}_0) . \tag{6}$$

$\tilde{\mathbf{x}}_0$ is uniquely determined by $\bar{\mathbf{x}}$ through the deterministic time evolution. In the presence of Gaussian uncorrelated noise of amplitude σ, the observed signal will follow a distorted distribution in m dimensions:

$$\mu_m(\mathbf{x})=\frac{1}{(\sigma\sqrt{2\pi})^m}\int d\mathbf{x}'\exp\left[-\frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{x}')^2\right]\nu(\mathbf{x}') . \tag{7}$$

Using Eq. (6) the integration over the δ functions can be carried out to yield

$$\mu_m(\mathbf{x})=\frac{1}{(\sigma\sqrt{2\pi})^{m-r}}\exp\left[-\frac{1}{2\sigma^2}(\tilde{\mathbf{x}}-\tilde{\mathbf{x}}_0)^2\right]\mu_r(\bar{\mathbf{x}}) . \tag{8}$$

All the information about the signal is contained in

$$\mu_r(\bar{\mathbf{x}})=\frac{1}{(\sigma\sqrt{2\pi})^r}\int d\bar{\mathbf{x}}'\exp\left[-\frac{1}{2\sigma^2}(\bar{\mathbf{x}}-\bar{\mathbf{x}}')^2\right]\nu_r(\bar{\mathbf{x}}'), \tag{9}$$

which we have to leave as it stands. If we had enough data to bin $\mu_m(\mathbf{x})$ reasonably fine in m dimensions we could "predict" $\tilde{\mathbf{x}}_0$ and determine the noise amplitude σ from the Gaussian functions in Eq. (8). For practical reasons this will not be our strategy.

We will rather study the correlation integral $C_m(\epsilon)$ of a distribution $\mu_m(\mathbf{x})$ of the form (8). A continuum definition of $C(\epsilon)$ is

$$C(\epsilon)=\int d\mathbf{x}\,\mu(\mathbf{x})\int_{\mathcal{B}(\epsilon,\mathbf{x})}d\mathbf{x}'\mu(\mathbf{x}') . \tag{10}$$

The second integration is over a "ball" $\mathcal{B}(\epsilon,\mathbf{x})$ of phase space centered at $\mathbf{x}$ and with radius ϵ. We will use the sup norm, in which case these "balls" are m-dimensional hypercubes of size 2ϵ.

Let us evaluate $d(\epsilon)$ for $\mu_m(\mathbf{x})$ given by (8). The correlation integral in m dimensions reads

$$C_m(\epsilon)=\frac{1}{(\sigma\sqrt{2\pi})^{2(m-r)}}\int d\mathbf{x}\exp\left[-\frac{1}{2\sigma^2}\overline{\mathbf{x}}^2\right]\mu_r(\overline{\mathbf{x}})\times\int_{\mathcal{B}(\epsilon,\mathbf{x})}d\mathbf{x}'\exp\left[-\frac{1}{2\sigma^2}\overline{\mathbf{x}}'^2\right]\mu_r(\overline{\mathbf{x}}')\ , \tag{11}$$

where we shifted $\mathbf{x}$ and $\mathbf{x}'$ to eliminate $\mathbf{x}_0$ and center the Gaussian functions. When we use the sup norm, we can carry out both integrals separately over the two subspaces:

$$C_m(\epsilon)=C_r(\epsilon)\frac{1}{(\sigma\sqrt{2\pi})^{2(m-r)}}\int d\tilde{\mathbf{x}}\exp\left[-\frac{1}{2\sigma^2}\tilde{\mathbf{x}}^2\right]\times\int_{\tilde{\mathcal{B}}(\epsilon,\tilde{\mathbf{x}})}d\tilde{\mathbf{x}}\exp\left[-\frac{1}{2\sigma^2}\tilde{\mathbf{x}}'^2\right]\ , \tag{12}$$

denoting the part containing the signal contribution by

$$C_r(\epsilon)=\int d\overline{\mathbf{x}}\mu_r(\overline{\mathbf{x}})\int_{\mathcal{B}_r(\epsilon,\overline{\mathbf{x}})}d\overline{\mathbf{x}}'\mu_r(\overline{\mathbf{x}}')\ . \tag{13}$$

The integrals in Eq. (12) can be evaluated through its components:

$$C_m(\epsilon)=C_r(\epsilon)\left[\frac{1}{(\sigma\sqrt{2\pi})^2}\int dx\exp\left[-\frac{1}{2\sigma^2}x^2\right]\int_{-\epsilon}^{\epsilon}dx'\exp\left[-\frac{1}{2\sigma^2}(x'+x)^2\right]\right]^{(m-r)}$$
$$=C_r(\epsilon)\left[\sqrt{2}\,\mathrm{erf}\left[\frac{\epsilon}{2\sigma}\right]\right]^{(m-r)}\ . \tag{14}$$

Substituting this into the definition of $d_m(\epsilon)$ [Eq. (3)] yields

$$d_m(\epsilon)=\frac{d}{d\ln\epsilon}\ln C_m(\epsilon)=d_r(\epsilon)+\frac{(m-r)\epsilon\exp\left[-\frac{1}{4\sigma^2}\epsilon^2\right]}{\sigma\sqrt{\pi}\,\mathrm{erf}\left[\frac{\epsilon}{2\sigma}\right]}\ . \tag{15}$$

Our main result is that all the curves $[d_m(\epsilon)-d_r(\epsilon)]/(m-r)$ have the same functional form, stretched by the noise level σ:

$$d_{m,r}(\epsilon)\equiv\frac{d_m(\epsilon)-d_r(\epsilon)}{m-r}=\frac{\epsilon\exp\left[-\frac{1}{4\sigma^2}\epsilon^2\right]}{\sigma\sqrt{\pi}\,\mathrm{erf}\left[\frac{\epsilon}{2\sigma}\right]}=g\left[\frac{\epsilon}{2\sigma}\right]\ , \tag{16}$$

where we introduced

$$g(z)=\frac{2}{\sqrt{\pi}}\frac{ze^{-z^2}}{\mathrm{erf}(z)}\ . \tag{17}$$

An estimate for the noise level is given by a least-squares fit of $g(\epsilon/2\sigma)$ to the values $d_{m,r}(\epsilon)$ available numerically.

We already discussed results with simulated data with known noise level. Moreover, we applied the method to experimental data, provided by Buzug and Pfister [5]. The data stem from the Taylor-Couette flow experiment. The noise in this particular set has been analyzed previously by Kantz *et al.* [4]. The sequence consists of 32 768 integers between 0 and 4095. The noise reduction scheme applied in [4] (which is the one proposed in [3]) gave a correction with an amplitude of 45 units, corresponding to 5% of the total variance.

We followed the same procedure as with the Ikeda data, except that we used delay 10 for the embedding to suppress effects of possible autocorrelation of the noise. The results for $d_{m,r}(\epsilon)$ with $r=5$ and the best fit are shown in Fig. 2. The main reason why the picture is less clear than for the Ikeda data is that we have to use fairly high embedding dimensions $(m>5)$. The estimated noise level is 41 units, in good agreement with the result of the explicit noise reduction procedure.

The remaining noise level after noise reduction (see [4] for details of the data processing) is more difficult to obtain (see Fig. 3). On the one hand, we cannot be sure that

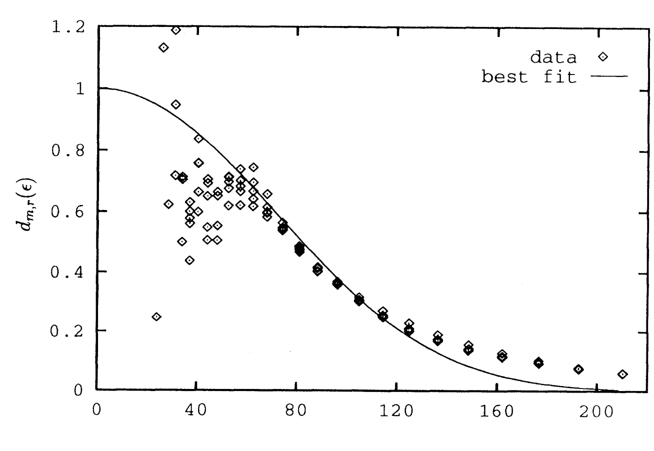

FIG. 2. Normalized effective dimensions $d_{m,r}(\epsilon)$ for experimental Taylor-Couette data, $m=6,\dots,10$. The best fit to Eq. (16) is obtained with $\sigma=41$ (solid line).

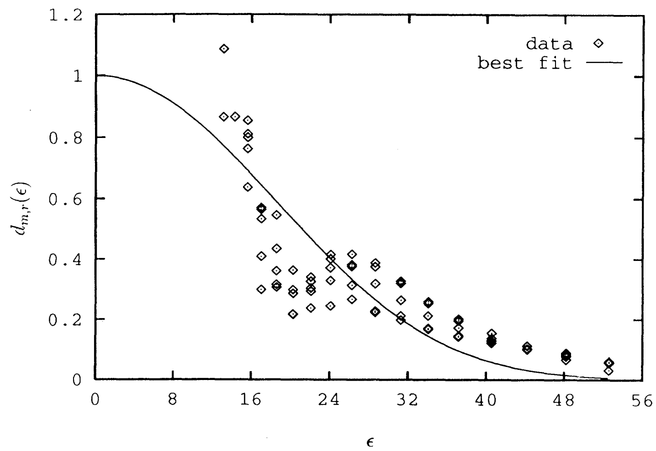

FIG. 3. Same as Fig. 2 but for the data after noise reduction (see [4] for details of the noise reduction). Although the characteristic shape [Eq. (16)] is not very pronounced, a fit is still possible, giving $\sigma=11$ (solid line).

the remaining error is still Gaussian. On the other hand, we expect the noise level to be smaller, which also means that we have to work further down in the fluctuation regime. Nevertheless, the values we obtain for $d_{m,r}(\epsilon)$ still yield a stable fit. We estimate the remaining noise to have an amplitude of about 11 (ca. 1.2% of the total variance). Thus the noise reduction procedure suppressed the noise by about a factor of 3–4.

In conclusion, we presented a method to estimate the noise level, which we believe can be a useful tool in the analysis of chaotic data. Since usually the correlation integral is computed anyhow in the search for scaling behavior, the additional cost for implementation and computation is very small. We found the SVD embedding most useful to separate signal and noise contributions: we could use smaller values of r than strictly required. Thus better statistics could be obtained.

An important consequence of our analytical result for the shape of the correlation integral [Eq. (15)] is that already a small amount of noise conceals possible scaling behavior: even if one uses an embedding in only one dimension more than strictly required for a faithful value of d [to make sure $d_m(\epsilon)$ converges with m], the effective dimension is raised by 0.2 even at $\epsilon=3\sigma$. That means, since even in the best case scaling can be expected up to ca. one-fourth of the attractor extent and down to three times the noise level, a data set with 2% noise can give at most a tiny scaling region of two powers of 2.

I want to thank Thorsten Buzug and Gerf Pfister for kindly providing the experimental data and Peter Grassberger and Holger Kantz for helpful discussions. The author receives a European Communities grant within the framework of the SCIENCE Programme, Contract No. B/SCI*-900557.

*Permanent address: Physics Department, University of Wuppertal, Gauss-Strasse 20, D-5600 Wuppertal 1, Germany.

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

[2] T. Schreiber and P. Grassberger, Phys. Lett. A **160**, 411 (1991); T. Sauer, Physica D **58**, 193 (1992); R. Cawley and G.-H. Hsu, Phys. Rev. A **46**, 3057 (1992); see also [3] and references therein.

[3] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, Chaos (to be published).

[4] H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L. G. Flepp, J. Simonet, R. Badii, and E. Brun, Phys. Rev. E (to be published).

[5] Th. Buzug, T. Reimers, and G. Pfister, in *Nonlinear Evolution of Spatiotemporal Structures in Dissipative Continuous Systems*, edited by F. H. Busse and L. Kramer (Plenum, New York, 1990).

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

[7] The box-counting dimension can be defined as follows: Let $n(\epsilon)$ be the minimum number of cubes of side length ϵ needed to cover the attractor. Then $n(\epsilon)\sim\epsilon^{d_{\text{box}}}$ as $\epsilon\to 0$, where d_{box} is the box-counting dimension. For more details, see J. D. Farmer, E. Ott, and J. A. Yorke, Physica D **7**, 153 (1983).

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

[9] Similar results hold for more general embedding methods, e.g., Fourier or SVD embeddings.

[10] K. Ikeda, Opt. Commun. **30**, 257 (1979).

[11] D. Broomhead and G. P. King, Physica D **20**, 217 (1986); A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp, Phys. Rev. A **38**, 3017 (1988).

[12] The method works with any embedding method that preserves distance. *Singular value decomposition* as used in [11] rotates phase space from usual delay coordinates to a space where the first coordinates contain most of the variance of the data. This gives usually a better separation of the noise dominated subspace.

[13] The box-counting dimension of the Ikeda attractor is slightly below 2. Thus a two-dimensional embedding preserves the correlation dimension. A faithful reconstruction of the dynamics, as required for the separation in Eq. (6), is obtained only for $r\geq 4$. Nevertheless, we empirically find $d_{m,r}(\epsilon)$ to have the expected shape already for $r=2$, which is preferable for better statistics.

[14] W. H. Press, B. P. Flannery, S. A. Teukolski, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1988).

[15] P. Grassberger, Phys. Lett. A **128**, 369 (1988).