Estimating invariants of noisy attractors

C. Diks
Department of Physiology, Leiden University, P.O. Box 9604, 2300RC Leiden, The Netherlands
(Received 30 November 1995)

We propose a method for estimating the correlation dimension and correlation entropy of a time series. It is based on a generalization of the correlation integral that is specifically useful when the time series is corrupted with Gaussian measurement noise. From computational experiments we conclude that reasonable estimates for the noise level, correlation dimension, and correlation entropy can be found for time series with up to 20% noise. The method appears to be fairly robust with respect to the noise distribution. [S1063-651X(96)51405-4]

PACS number(s): 05.45.+b, 02.50.−r

The characterization of nonlinear time series in the presence of measurement noise is a problem of great current interest. For the ideal case of noise-free deterministic time series, the reconstruction theorem [1,2] has led to a number of powerful characterization methods. Measurement noise, however, is known to put severe limitations on the estimation of dynamical invariants from time series with these methods.

The most commonly used characterization method of noise-free time series is that of Grassberger and Procaccia [3]. It allows the determination of both the correlation dimension D and the correlation entropy K of an attractor after constructing the m-dimensional delay vectors $\vec{x}_n=(x_n,x_{n+\tau},\ldots,x_{n+(m-1)\tau})$ with delay τ from a time series $\{x_n\}_{n=1}^N$. The correlation integral $C_m(r)$ is defined in terms of the distribution $\rho_m(\vec{x})$ of delay vectors as

$$C_m(r)=\int d\vec{x}\rho_m(\vec{x})\int d\vec{y}\rho_m(\vec{y})\Theta(r-|\vec{x}-\vec{y}|) \tag{1}$$

where $\Theta(\ )$ denotes the Heaviside function and $|\ |$ a norm. The correlation dimension D and correlation entropy K are defined as $D=\lim_{r\to 0}\lim_{m\to\infty}d\log C_m(r)/d\log r$ and $K=\tau^{-1}\lim_{r\to 0}\lim_{m\to\infty}(\{-\log[C_m(r)]\}/m)$. In practice, the limits cannot be taken, and a log-log plot of the estimated $C_m(r)$ versus r is usually made, in order to look for a range in the lower r region where the curves are approximately linear and parallel for consecutive values of m. For small values of r, measurement noise gives rise to an increased slope in the log-log plot of the correlation integral. For low noise levels, a scaling region can usually still be found, whereas this is impossible if the noise level is high.

Basically two approaches can be distinguished concerning the analysis of time series with measurement noise. The first is to separate the noise and the underlying time series with a noise reduction method (for surveys see [4,5]). The second is based on characterizing the modified delay vector distribution. By calculating the effect of noise on the correlation integral, Schouten *et al.* [6] obtained a method for estimating the correlation dimension in the case of bounded independent, identically distributed (IID) noise. Schreiber [7] has proposed a method for estimating the noise level of a deterministic time series contaminated with unbounded IID Gaussian measurement noise. The effect of this noise on the correlation integral has also been investigated by Smith [8], who used an approximation of the correlation integral to estimate D for small noise levels. The analytic difficulties which prevent the estimation of invariants at higher noise levels appear to be related to the contrast between the smooth Gaussian noise distribution on the one hand and the abrupt nature of the kernel function $\Theta(\ )$ in the correlation integral (1) on the other.

In this Rapid Communication we will show that, in the context of IID Gaussian noise, a more natural formalism is obtained by examining a function from the same family as the correlation integral, but which is tailored for Gaussian measurement noise. We start by considering the correlation integral as a member of a generalized class of kernel integrals. Then a Gaussian kernel member is picked from this class and its behavior is derived analytically in the presence of Gaussian measurement noise. We then give some example applications to noisy deterministic time series for which D, K and the noise level are estimated.

The correlation integral defined by (1) can be generalized to

$$T_m(h)=\int d\vec{x}\rho_m(\vec{x})\int d\vec{y}\rho_m(\vec{y})w(|\vec{x}-\vec{y}|/h) \tag{2}$$

where $w(\ )$ is a kernel function. The correlation integral (1) is retained when the kernel function $w(x)$ is taken to be $\Theta(1-x)$. The parameter h will be referred to as the bandwidth.

Using the Gaussian kernel function

$$w(x)=e^{-x^2/4}, \tag{3}$$

a version of the correlation integral,

$$T_m(h)=\int d\vec{x}\rho_m(\vec{x})\int d\vec{y}\rho_m(\vec{y})e^{-|\vec{x}-\vec{y}|^2/(4h^2)}, \tag{4}$$

is obtained which will be referred to as the Gaussian kernel correlation integral. Ghez and Vaienti [9,10] used a Gaussian kernel function for the estimation of dimensions and entropies of noise-free time series. Gaussian kernel functions are also used in a statistical test for the reversibility of time series [11] and for comparing the delay vector distributions of two time series [12].

The following properties of $T_m(h)$ will be used below. If we take m fixed and consider a deterministic time series with correlation dimension D, then the scaling law

$$T_m(h) \sim h^D \quad \text{for } m \text{ fixed}, \quad h \to 0, \tag{5}$$

holds according to the results of Ghez and Vaienti [9]. More generally, any kernel $w(x)$ which decreases monotonically in x for $x \geqslant 0$ and for which $\lim_{h\to 0} h^{-p} w(x/h) = 0$ pointwise for $x > 0$ and for any $p \geqslant 0$, implies the scaling law (5).

The m dependence of (5) is found by expressing $T_m(h)$ as $T_m(h) = \int dr\, \eta_m(r) w(r/h)$ where $\eta_m(r) = \mathrm{d}C_m(r)/\mathrm{d}r$ is the distribution of the interpoint distances r. It was shown by Frank *et al.* [13] that the correlation integral calculated with the Euclidean norm behaves as

$$C_m(r) \sim e^{-K\tau m} (r/\sqrt{m})^D \quad \text{for } r \to 0, \ m \to \infty, \tag{6}$$

which implies $\eta_m(r) \sim e^{-K\tau m} m^{-D/2}$ for fixed r. The m dependence thus is described by the factor $e^{-K\tau m} m^{-D/2}$. We therefore find

$$T_m(h) \sim e^{-K\tau m} m^{-D/2} h^D \quad \text{for } h \to 0, \ m \to \infty, \tag{7}$$

for Gaussian kernel correlation integrals in the noise-free case with the Euclidean norm.

Following Frank *et al.*, we could remove the factor $m^{-D/2}$ in (7) by defining an m-dependent bandwidth. There is, however, a practical reason for not using this freedom and proceeding with (7). Due to the finiteness of the attractor there usually is an upper bandwidth up to which the behavior (7) is observed, and it is approximately independent of m. By using dimension scaled bandwidths we would be forced to go to smaller upper bandwidths for increasing m.

With the Euclidean norm, the Gaussian kernel correlation integral can be written as

$$T_m(h) = (2h\sqrt{\pi})^m \int d\vec{x} [\rho_m^h(\vec{x})]^2 \tag{8}$$

where

$$\rho_m^h(\vec{x}) = (h\sqrt{2\pi})^{-m} \int d\vec{y} \rho_m(\vec{y}) e^{-|\vec{x}-\vec{y}|^2/(2h^2)}. \tag{9}$$

This is best demonstrated by deriving (4) from (8) and (9). After substituting (9) into (8) we obtain

$$T_m(h) = (h\sqrt{\pi})^{-m} \times \int d\vec{x} \int d\vec{y} \int d\vec{z} \rho_m(\vec{y}) \rho_m(\vec{z}) \times e^{-[|\vec{x}-\vec{y}|^2 + |\vec{x}-\vec{z}|^2]/(2h^2)}. \tag{10}$$

Expanding the expression between square brackets in the exponent as $\sum_{i=1}^m 2[x_i - \frac{1}{2}(y_i + z_i)]^2 + 1/2(y_i - z_i)^2$, the integral over $\vec{x}$ is easily performed and results in (4).

The distribution $\rho_m^h(\vec{x})$ given in (9) allows an interpretation as a convolution of the delay vector distribution $\rho(\vec{x})$ with a normalized Gaussian distribution with a standard deviation equal to h. This property may be exploited in the presence of Gaussian measurement noise which itself acts as a Gaussian convolution on the delay vector distribution.

To see how IID Gaussian measurement noise effects $T_m(h)$ it is useful to make a clear distinction between the noisy delay vector distribution $\rho(\vec{x})$ and its underlying noise-free distribution $\overline{\rho(\vec{x})}$. The relation between $\rho(\vec{x})$ and $\overline{\rho(\vec{x})}$ can be described by a convolution with a normalized Gaussian distribution with standard deviation σ [14]. The distribution $\rho_m^h(\vec{x})$ in turn is obtained from $\rho(\vec{x})$ by a convolution with a normalized Gaussian distribution with standard deviation h as described by (9). The two consecutive convolutions can be summarized by the single convolution

$$\rho_m^h(\vec{x}) = (s\sqrt{2\pi})^{-m} \int d\vec{y} \overline{\rho(\vec{y})} e^{-|\vec{x}-\vec{y}|^2/(2s^2)}, \tag{11}$$

where $s = \sqrt{h^2 + \sigma^2}$. Substituting (11) into (8) and rewriting this in the form of (4), we obtain

$$T_m(h) = \left(\frac{h^2}{h^2 + \sigma^2} \right)^{m/2} \times \int d\vec{x} \overline{\rho_m}(\vec{x}) \int d\vec{y} \overline{\rho_m}(\vec{y}) e^{-|\vec{x}-\vec{y}|^2/(4h^2 + 4\sigma^2)}. \tag{12}$$

Equation (12) describes $T_m(h)$ in terms of the underlying distribution $\overline{\rho(\vec{x})}$ in the presence of IID Gaussian noise with standard deviation σ.

The behavior of the double integral in (12) is found from the definition of $T_m(h)$ given in (4) together with the noise-free scaling law (7), leading to

$$T_m(h) \simeq \phi \left(\frac{h^2}{h^2 + \sigma^2} \right)^{m/2} e^{-K\tau m} m^{-D/2} \sqrt{h^2 + \sigma^2}^{\,D} \quad \text{for } \sqrt{h^2 + \sigma^2} \to 0, \quad m \to \infty, \tag{13}$$

where ϕ is a normalization constant.

In practice the standard deviation σ of the noise level is fixed at a nonzero value. We are thus not able to let $\sqrt{h^2 + \sigma^2}$ go to zero. Nevertheless, we expect relation (13) to hold good in a range of small values of h if the noise level σ is not too large. Note that the Gaussian kernel correlation integral for small values of h and m fixed behaves as $T_m(h) \sim h^m$, which is a manifestation of the m-dimensionality of the set of noisy delay vectors. Taking the limit $\sigma \to 0$ on the other hand, gives back the scaling relation (7) of the noise-free case.

Before we describe the application of our method to time series, we want to make a remark about our normalization conventions. All time series are rescaled to have a standard deviation of 1 and the quoted noise levels denote the noise levels after rescaling. This allows for a convenient comparison of the bandwidth parameter h and the noise levels σ for the different time series to be considered. The case $\sigma = 0$ corresponds with a clean noise-free time series whereas a noise level of $\sigma = 1$ implies a time series consisting of pure IID samples.

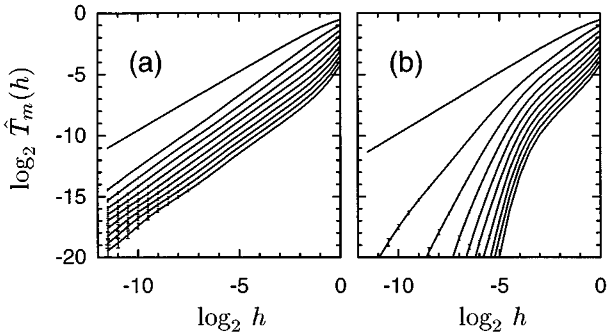

FIG. 1. The estimate $\hat{T}_m(h)$ as a function of the bandwidth h on a log-log scale for a Hénon time series without noise (a) and with a (normalized) noise level of 0.05 (b). The different lines correspond to the cases $m=1$ (upper line) up to $m=10$ (lower line). The bars denote the estimated standard error.

The Gaussian kernel correlation integrals $T_m(h)$ can be consistently estimated by replacing the integrals over the delay vector distributions in (4) with an average over delay vectors which are assumed to be independently distributed according to $\rho_m(\vec{x})$. The estimate $\hat{T}_m(h)$ becomes

$$\hat{T}_m(h)=\frac{1}{N_p}\sum_i \sum_{j\neq i} \psi_{ij}(h) \tag{14}$$

where N_p is the number of (i,j) pairs used and

$$\psi_{ij}(h)=e^{-|\vec{x}_i-\vec{x}_j|^2/(4h^2)}. \tag{15}$$

In Fig. 1(a), the estimate $\hat{T}_m(h)$ is drawn as a function of the bandwidth h on a log-log scale for a noise-free time series generated with the Hénon model, of length $N=4000$. The curves are obtained by choosing at random 1000 reference indices i and using all values of j for which $j\neq i$. The bandwidth parameters were chosen equidistant on a logarithmic scale with 2 values per binade. This choice has the advantage that it is sufficient to perform the numerically time-consuming evaluation of the exponential function in $w_{ij}(h)$ only for the largest value of the bandwidth parameter. The value of $w_{ij}(h)$ at the smaller bandwidths can then be found efficiently using the relation $w_{ij}(h/\sqrt{2})=w_{ij}^2(h)$.

The scaling relation (7) for low-dimensional deterministic time series implies parallel linear curves for small values of h and large values of m. In Fig. 1(a) this behavior can be observed for a large range of h values for $m\geqslant 2$. Figure 1(b) shows a log-log plot of $\hat{T}_m(h)$ versus h for the same Hénon time series with IID Gaussian noise with a standard deviation of $\sigma=0.05$. It can be observed in Fig. 1(b) that the noise gives rise to an increased slope for small bandwidth values h.

For different values of the noise level, a Marquardt nonlinear fit procedure [15] for the parameters ϕ, K, D, and σ was performed in the range where $h\leqslant 0.25$, and $\hat{T}_m(h)>2/N_p$. For each m, the values $\hat{T}_m(h)$ and $\hat{T}_{m+1}(h)$ were fitted simultaneously to the model function (13). The standard deviations of the estimates $\hat{T}_m(h)$ were taken as weights in the fit procedure. Assuming independence of the distances, the variance $\mathcal{V}$ of $\hat{T}_m(h)$ is estimated as

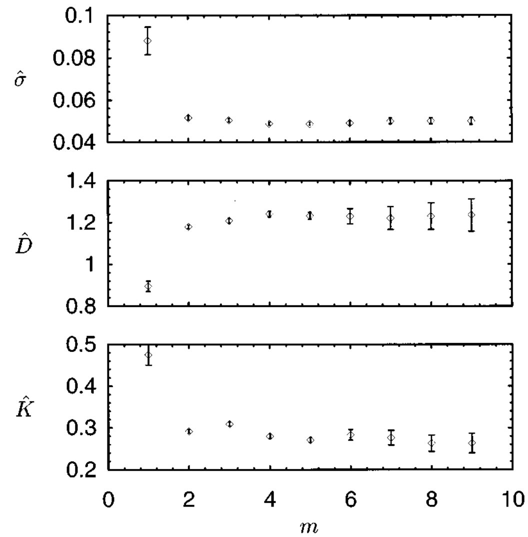

FIG. 2. Estimated values $\hat{\sigma}$, $\hat{D}$ and $\hat{K}$ as a function of m for a Hénon time series with a noise level σ of 0.05. The bars denote the estimated 95% confidence interval (2 standard errors).

$$\mathcal{V}[\hat{T}_m(h)]=\frac{1}{N_p}[\overline{\psi_{ij}^2}(h)-\overline{\psi_{ij}}(h)^2] \tag{16}$$

where the bars denote averaging over the pairs (i,j).

The estimates of the model parameters at a noise level of $\sigma=0.05$ as a function of m are shown in Fig. 2. Reasonable values of $\hat{\sigma}$, $\hat{D}$, and $\hat{K}$ (the estimates of σ, D, and K respectively) are obtained at moderately small values of m ($m=3$ or $m=4$).

The results for different noise levels ranging up to $\sigma=0.20$ are summarized in Table I. All values were estimated at $m=4$. The estimates of σ are close to their true values and for noise levels up to 0.10, the values of $\hat{D}$ and $\hat{K}$ are close to the values found in the literature ($D\approx 1.22$, see [3] and $K\approx 0.29$, see [13]). The estimated values of the standard error, however, seem to be on the small side. This is possibly due to cross-correlations between the estimates of the Gaussian kernel correlation functions for different values of h and m.

We applied the method to a noise-free time series ($N=10\,000$, sample time 0.5, $\sigma=0.00$) generated with the Rössler model (see Ref. [16]). The estimated parameters at $m=9$ for $\tau=3$ were $\hat{\sigma}=0.0007\pm 0.0002$, $\hat{D}=1.97\pm 0.01$ and $\hat{K}=0.066\pm 0.009$. Application of the method to a noisy

TABLE I. Estimated values $\hat{\sigma}$, $\hat{D}$ and $\hat{K}$, $\pm$ the estimated standard errors for Hénon time series contaminated with different noise levels σ. All values are estimated at $m=4$.

σ	$\hat{\sigma}$	$\hat{D}$	$\hat{K}$
0.00	0.00004 ± 0.00009	1.196 ± 0.002	0.296 ± 0.003
0.01	0.0102 ± 0.00006	1.205 ± 0.002	0.293 ± 0.002
0.02	0.0195 ± 0.0002	1.200 ± 0.004	0.294 ± 0.003
0.05	0.0487 ± 0.0003	1.240 ± 0.007	0.280 ± 0.003
0.10	0.0996 ± 0.0008	1.26 ± 0.02	0.309 ± 0.003
0.20	0.206 ± 0.003	1.16 ± 0.05	0.278 ± 0.003

Rössler time series ($N=10\,000$, sample time 0.5, $\sigma=0.05$) gave the estimates $\hat{\sigma}=0.0507\pm0.0003$, $\hat{D}=1.94\pm0.03$ and $\hat{K}=0.061\pm0.001$. An upper limit for the value of D is the information dimension which was estimated by Grassberger *et al.* [16] as 2.00 ± 0.01. Using the method of Schouten *et al.* [17], we have found a correlation entropy of 0.07 for the Rössler time series. These examples show that our method can also be applied to time series obtained from continuous time dynamical systems.

The sensitivity of our method with respect to the type of measurement noise, was tested on a Hénon time series with independent uniformly distributed noise with a standard deviation σ of 0.05. The estimated parameters are $\hat{\sigma}=0.0454\pm0.0003$, $\hat{D}=1.220\pm0.007$ and $\hat{K}=0.294$ ±0.003. Although the estimated noise level is about 10% too small, the estimates of D and K are still very reasonable. This suggests that the method can be of use for different noise distributions.

In this Rapid Communication, we have introduced the Gaussian kernel correlation integral $T_m(h)$ which is tailored for the characterization of delay vector distributions in the presence of Gaussian measurement noise. For small noise levels σ, the behavior of the Gaussian kernel correlation integral is derived analytically in terms of the noise level σ and the invariants D and K. This allows the estimation of the noise level of a time series simultaneously with these invariants using a Marquardt type estimation method. The first results obtained with maps and continuous time dynamical systems are in good agreement with the noise-free values. The standard errors of the parameters appear to be somewhat underestimated, possibly due to the neglected cross-correlations among the estimates of the Gaussian kernel correlation integrals at different values of h and m. The method works well up to noise levels σ of about 0.20. Although a trial with uniformly distributed noise suggests that the method is fairly robust against the type of measurement noise, we expect an improvement with the use of linear combinations of several coordinates like in embeddings based on singular value decompositions. Also the IID requirement may be relaxed when a delay τ of the order of the autocorrelation time of the noise is used.

In practice an appropriate choice of the upper bandwidth has to be made. Also *a priori* it is not known whether an experimental time series consists of a low-dimensional component corrupted with measurement noise. In order to prevent spurious estimates, the quality of the fit below the upper bandwidth chosen should be investigated. Furthermore, the stability of the estimated parameters upon changing the embedding parameters τ and m should be assessed.

The author wishes to thank Professor F. Takens and Dr. J. DeGoede for stimulating discussions.

[1] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).

[2] F. Takens, in *Dynamical Systems and Turbulence, Warwick, 1980*, edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, Berlin, 1981), p. 366.

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

[4] E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).

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

[6] J. Schouten, F. Takens, and C. van den Bleek, Phys. Rev. E **50**, 1851 (1994).

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

[8] R. L. Smith, J. R. Statist. Soc. B **54**, 329 (1992).

[9] J. M. Ghez and S. Vaienti, Nonlinearity **5**, 777 (1992).

[10] J. M. Ghez, E. Orlandini, M.-C. Tesi, and S. Vaienti, Physica D **63**, 282 (1993).

[11] C. Diks, J. C. van Houwelingen, F. Takens, and J. DeGoede, Phys. Lett. A **201**, 221 (1995).

[12] C. Diks, W. R. van Zwet, F. Takens, and J. DeGoede, Phys. Rev. E **53**, 2169 (1996).

[13] M. Frank, H.-R. Blank, J. Heindl, M. Kaltenhäuser, H. Köchner, W. Kreische, N. Müller, S. Poscher, R. Sporer, and T. Wagner, Physica D **65**, 359 (1993).

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

[15] W. H. Press, S. A. Teukolski, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

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

[17] J. Schouten, F. Takens, and C. van den Bleek, Phys. Rev. E **49**, 126 (1994).