Nova Série
BOLETIM
DA SOCIEDADE BRASILEIRA DE MATEMÁTICA

Bol. Soc. Bras. Mat., Vol. 29, N. 2, 197-228

The analysis of correlation integrals in terms of extremal value theory

Floris Takens

— Dedicated to the memory of Ricardo Mañé

Abstract. In this paper we first give an overview of the methods of analysis of time series in terms of correlation integrals, which were developed for time series generated by deterministic systems. From the extremal value theory one obtains asymptotic information on the behaviour of the correlation integrals of time series generated by non-deterministic (mixing) systems. This leads to an analysis in terms of correlation integrals which is complementary to the estimation of dimension and entropy.

Keywords: time series, (deterministic) chaos, correlation integral, predictability, extremal index.

1. Introduction

Correlation integrals were originally introduced in order to analyse time series generated by nonlinear deterministic dynamical systems. It was proved in [**T,1981**], [**GP,1983**], [**T,1983**] that some dynamical characteristics, like dimension and entropy of the relevant attractor, can be defined in terms of these correlation integrals. So the estimation of these characteristics can be done in terms of estimates of the correlation integrals. If the estimates of dimension and entropy do not 'converge', this is an indication that the time series is not generated by a deterministic dynamical system, at least not by such a system with sufficiently low dimension and entropy. For the precise meaning of 'converge', in this context, and other notions, see the next section.

Apart from this, the correlation integrals of a time series contain information about the predictability by nonlinear methods, compara-

Received 9 June 1998.

ble with the information contained in the autocorrelations about the predictability by linear methods. This means that the correlation integrals are relevant, independent of the fact whether we have a time series which is generated by a deterministic system or not. This is important because real and experimental systems are never strictly deterministic. In the deterministic case, there is a considerable redundancy in the values of the correlation integrals: the values of dimension and entropy contain in a concise form the most important (asymptotic) information about these correlation integrals. It is the purpose of the present paper to obtain an analogous reduction of the information for a large class of time series which are generated by stochastic systems. Such a reduction will be obtained by using ideas from extremal value theory. With this method we succeeded to interpret the estimated correlation integrals of time series from fluid bed reactors generated at the Technical University of Delft. The analysis of these time series was the starting point of the research reported in this paper.

2. Correlation integrals

In this section we give a review of the use of correlation integrals in the analysis of time series. We illustrate these with two examples, one from a (stochastic) autoregressive system and one from the (deterministic) Hénon-system.

2.1 Time series, stationarity, reconstruction distributions and order

Time series. In what follows we consider real valued time series. One can consider these as (potentially infinite) sequences of measurements which we denote by $\{x_i\}_{i\geq 1}$. In probability theory it is often convenient to consider an observed time series as a realization of some random process (in the case of a deterministic system, the only 'randomness' consists of the choice of the initial state, which determines the entire realization) in which case we write $\{x_i(\omega)\}_{i\geq 1}$, where ω is an element of the probability space (Ω, μ) of all possible realizations. For a discussion of these two approaches, see e.g. [**B,1981**] section 2.11. These two approaches are

essentially (i.e. under some further hypotheses, the most important of which is stationarity, see below) equivalent, but notationally one or the other is more convenient, depending on the situation.

Stationarity and reconstruction distributions. We always assume that we deal with stationary time series. The definition of this notion is different in the two approaches mentioned above. In the formalism of 'realizations' we have the finite dimensional distributions of $(x_{i_1}(\omega), \cdots, x_{i_k}(\omega))$ induced by the probability measure μ on Ω. Stationarity, or stationarity in the strict sense, means that these probability distributions are invariant under translations of time, i.e. that the distribution of the vectors $(x_{i_1}(\omega), \cdots, x_{i_k}(\omega))$ is the same as the distribution of $(x_{i_1+s}(\omega), \cdots, x_{i_k+s}(\omega))$ for any s. Without loss of generality we may restrict ourselves, in this definition of stationarity, to the distributions of the vectors of the form $(x_i(\omega), x_{i+1}(\omega), \cdots, x_{i+k-1}(\omega))$. For stationary time series, the whole probabilistic structure is given by the distributions of the vectors $(x_1(\omega), x_2(\omega), \cdots, x_k(\omega))$. These distributions are called the *k-dimensional reconstruction distributions* and are denoted by μ_k. This terminology is used because the vectors $(x_n(\omega), \cdots, x_{n+k-1})$ are often called reconstruction vectors because of their role in the reconstruction theorem [**T,1981**], [**SYC,1991**], which was the original motivation to study these probability distributions in the theory of nonlinear systems. Often one only needs to consider the reconstruction distributions μ_k for low values of k, see the discussion of the order of a time series below.

If we have only one time series, the reconstruction distributions, if they exist, are defined differently: We say that a time series $\{x_i\}_{i\geq 1}$ has a k-dimensional reconstruction distribution μ_k if for each positive continuous function $f : \mathbb{R}^k \to \mathbb{R}$ we have:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i, x_{i+1}, \cdots, x_{i+k-1}) = \int f \mathrm{d}\mu_k$$

(here we also speak of equality if both the left and right side diverge). In this case, where we have only one realization, we call the time series stationary if the reconstruction distributions exist in the above sense.

Order of dynamical systems. We say that a time series has order n if,

when predicting a new (future) element of the time series on the basis of the most recent values, with the knowledge of the general statistical properties of the time series (i.e. with the knowledge of the reconstruction distributions), one obtains an optimal results using the last n values in the sense that less than n values give in general a worse prediction and that more than n values don't help to improve the prediction. In order to make this more formal we have to introduce some notions.

For a probability distribution μ on $\mathbb{R}^l$ we define its projection π_1 on $\mathbb{R}^{l-1}$ by

$$(\pi_1(\mu))(A) = \mu(\{(x_1, \cdots, x_l) | (x_2, \cdots, x_l) \in A\}),$$

for arbitrary measurable subsets $A \subset \mathbb{R}^{l-1}$. The projection π_l is similarly defined, interchanging the roles of the first and the l-th coordinates.

If we consider a stationary time series with reconstruction distributions μ_k, these distributions satisfy $\pi_1(\mu_k) = \pi_k(\mu_k) = \mu_{k-1}$. Now we consider a fixed k and observe that there is a unique probability distribution μ'_{k+1} on $\mathbb{R}^{k+1}$ such that for any three subsets $A, B \subset \mathbb{R}$ and $C \subset \mathbb{R}^{k-1}$ we have that

$$\mu'_{k+1}(A \times C \times B) = \frac{\mu_k(A \times C)\mu_k(C \times B)}{\mu_{k-1}(C)},$$

where of course $\mu_{k-1} = \pi_1(\mu_k) = \pi_k(\mu_k)$.

Then our time series has order n if for all $k > n$ we have, in the above notation, $\mu'_{k+1} = \mu_{k+1}$. We should note that it is very common for stationary time series that, for k sufficiently big, μ'_{k+1} and μ_{k+1} are at least very close. This means that it is not a severe restriction to require that a time series has finite order. For estimation of the order of (nonlinear) time series see [**CT,1992**] and [**T,1996**].

Mixing. There are various definitions of mixing for dynamical systems and time series. Here we consider the *strong mixing* as introduced by Rosenblatt [**R,1956**]. We say that a stationary time series with reconstruction distributions μ_k is strongly mixing if there is a function $g : N \to \mathbb{R}$, such that $g(n)$ converges to zero for $n \to \infty$, and such that for each l, n, and m and measurable subsets $A \subset \mathbb{R}^l$ and $B \subset \mathbb{R}^m$, we

have:

$$|\mu_{l+n+m}(A \times \mathbb{R}^n \times B) - \mu_l(A)\mu_m(B)| < g(n).$$

So the function values $g(n)$ can be considered as a measure for the coherence of the behaviour of the time series during time intervals which are at least n units apart. Mixing means that this coherence disappears if the time distance increases.

Note that a *time series* which is generated by a deterministic system is (usually) not mixing, even if the *dynamical system* is mixing in the sense of ergodic theory.

2.2 Definition of correlation integral

Let X be a set in which there are defined a probability distribution μ and a metric ρ. We define the correlation integral at distance ε, $P(\varepsilon)$, as the probability that two independently chosen μ-random elements of X are within distance ε. (We shall disregard the possibility that the distribution of distances, or other distributions have points with positive probability, i.e., we tacitly assume that all our probability distributions are non-atomic, so that we don't care to distinguish between the strict and the non-strict inequality when defining distribution functions.) Applying this notion to the reconstruction distributions of a time series we obtain the correlation integrals of the time series:

Let $\{x_i(\omega)\}_{i\geq 1}$ be a stationary time series with reconstruction distributions μ_k. Then the k-dimensional correlation integral of this time series at distance ε, $P^k(\varepsilon)$, is defined as the correlation integral at distance ε in the above sense of $\mathbb{R}^k$ with the reconstruction distribution μ_k and metric $\rho_k((x_1, \cdots, x_k), (y_1, \cdots, y_k)) = \max_{i=1,\cdots,k} |x_i - y_i|$.

It is clear that if we have a finite segment of (one representative of) a stationary time series $\{x_i\}_{i=1}^N$, we can estimate the correlation integral $P^k(\varepsilon)$ as

$$\hat{P}^k(\varepsilon) = \frac{2}{(N-k+1)(N-k)} \#\{(i,j) | 1 \leq i < j \leq N-k+1$$
$$\text{and } \rho_k((x_i, \cdots, x_{i+k-1}), (x_j, \cdots, x_{j+k-1})) < \varepsilon\},$$

where # stands for 'the number of elements of'. Though this methods of

estimation is correct in the sense that for $N \to \infty$ the estimate converges to the correct value, there are refinements which are useful in case of limited values of N, see [**T,1990**].

2.3 Estimation of dimension and entropy

Dimension. If X is again a set with a probability distribution μ and a metric ρ, then we define its correlation dimension as

$$D(X) = \lim_{\varepsilon \to 0} \frac{\ln(P(\varepsilon))}{\ln(\varepsilon)}.$$

If the limit does not exist, then the correlation dimension is not defined (some authors define in that case an upper and lower correlation dimension by using limsup and liminf). For a stationary time series, we use this same definition and obtain D^k, the dimension of the k-dimensional reconstruction distribution μ_k. From the definition it is clear that this dimension can be estimated from the estimates of the correlation integrals. However, according to the definition we have to take the limit for $\varepsilon \to 0$, and, for small values of ε, $P^k(\varepsilon)$ becomes extremely small, and hence the estimate $\hat{P}^k(\varepsilon)$ becomes very inaccurate in the sense that the relative estimation error will be large. For this reason we introduce the notion of the dimension $D^k(\varepsilon)$ at a fixed length-scale ε:

$$D^k(\varepsilon) = \frac{\partial \ln(P^k(\varepsilon))}{\partial \ln(\varepsilon)}$$

which can be estimated as:

$$\hat{D}^k(\varepsilon) = \frac{\ln(\hat{P}^k(\alpha\varepsilon)) - \ln(\hat{P}^k(\varepsilon))}{\ln(\alpha)},$$

where α is some constant (in our numerical calculations we used $\alpha = .9$).

If these estimated values $\hat{D}^k(\varepsilon)$ are approximately independent of ε for low values of ε, then we can use these estimates as an estimate of the dimension D^k of the k-dimensional correlation distribution of the time series. In principle there may be no convergence of $D^k(\varepsilon)$ (or $\hat{D}^k(\varepsilon)$) for decreasing values of ε, but this is very rare.

If we now consider a time series which is generated by a deterministic system in the sense of [**T,1981**] or [**SYC,1991**], then it follows that

the dimensions D^k, for increasing values of k, become constant, i.e. independent of k. If for a given time series the dimensions D^k, as far as they can be estimated, converge, for increasing k, to a constant, then this constant is called the *dimension* of the time series. For a motivation of this definition we also refer to the above mentioned papers [**T,1981**] and [**SYC,1991**]. In the case that these estimated dimensions $\hat{D}^k$ do not tend to a constant, we say that the *dimension estimation does not converge.*

The typical situation for time series of a non-deterministic system is that for $\varepsilon \to 0$, $D^k(\varepsilon)$, and also $\hat{D}^k(\varepsilon)$, converges to k for $\varepsilon \to 0$, and hence the values of $\hat{D}^k = k$ diverge for increasing k.

Entropy. Next we come to the estimation of the entropy of a time series. This notion is derived from the notion of (Kolmogorov) entropy for dynamical systems. For a discussion of the various types of entropy of dynamical systems and their estimation from time series we refer to [**KS,1997**], Chapter 11, Section 4, and to [**GP,1983**] and [**T,1983**]. Here we only define the notion for (stationary) time series in terms of the correlation integrals. As in the case of dimensions, it is possible that for a time series the entropy is not defined and/or that the estimation of the entropy does not 'converge'. First the formal definition. The entropy of a time series with correlation integrals $P^k(\varepsilon)$ is defined as

$$H = \lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{-\ln(P^k(\varepsilon)}{k},$$

whenever this limit converges. As in the case of dimension estimation the estimation of entropy is done in two steps in which intermediate quantities are estimated. We first define the entropy at embedding dimension k and at length scale ε as

$$H^k(\varepsilon) = -\ln(P^{k+1}(\varepsilon)) + \ln(P^k(\varepsilon)).$$

It is clear that we estimate this quantity by replacing in the definition the correlation integrals by their estimates, so

$$\hat{H}^k(\varepsilon) = -\ln(\hat{P}^{k+1}(\varepsilon)) + \ln(\hat{P}^k(\varepsilon)).$$

If $H^k(\varepsilon)$ converges to a constant for $k \to \infty$, then this constant equals

$$H(\varepsilon) = \lim_{k\to\infty} (-\ln(P^{k+1}(\varepsilon)) + \ln(P^k(\varepsilon))) = \lim_{k\to\infty} \frac{-\ln(P^k(\varepsilon)}{k}.$$

In principle there may be no convergence at this point, but that is very rare. The estimation of $H(\varepsilon)$ is usually done by visual inspection of the graph of $\hat{H}^k(\varepsilon)$ as function of k. Some refinements are possible here but, as far as I know a systematic way of estimating $H(\varepsilon)$ has not been published.

Next, if $H(\varepsilon)$, for $\varepsilon \to 0$, converges to a constant, then this constant has to be equal to the entropy as defined before:

$$H = \lim_{\varepsilon\to 0} H(\varepsilon) = \lim_{\varepsilon\to 0} \lim_{k\to\infty} \frac{-\ln(P^k(\varepsilon)}{k}.$$

At this point there are often problems with the convergence. In fact, for a stochastic time series, one expects that $H(\varepsilon)$ tends to ∞ as ε tends to zero.

2.4 Numerical examples of estimates of correlation integrals and various derived quantities

For two time series, which were generated numerically, namely one time series from the Hénon-system and one from an autoregressive system we estimated the correlation integrals and give a graphical presentation of the related quantities as introduced above.

The first time series was generated by the rule

$$x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$$

and starting with random initial values close to zero. The second time series was generated by the rule $x_n = .9x_{n-1} +$ 'noise' where the noise is generated by a random generator producing independent realizations from a normal distribution with average zero and variance one. In both cases, in order to obtain time series which are practically stationary, we disregarded the first 100 values. The actual time series consists of the next 3000 elements. Before further calculations were carried out the time series were rescaled such that the maximal, respectively minimal, value equals one, respectively zero.

For both time series we first estimated the correlation integrals $P^k(\varepsilon)$ for $k = 1, \cdots, 20$ and $\varepsilon = 1, .9, .9^2, \cdots, .9^{59}$. For each value of k, we plotted, $\ln(\hat{P}^k(\varepsilon))$ as function of $\ln(\varepsilon)$, see figures 1 and 4. Since the values of $P^k(\varepsilon)$ and hence the values of $\ln(P^k(\varepsilon))$ decrease for increasing k, we have from the top to the bottom the values of k increasing from 1 to 20. Since the relative estimation errors of $\hat{P}^k(\varepsilon)$ become large as $P^k(\varepsilon)$ becomes low, we omitted those estimates which would have been based on less than 50 pairs (i, j) as in the formula for the estimation of correlation integrals.

Already from these graphical representations one gets an idea of the dimensions and entropies as defined above: the estimated dimensions are the slopes of the various graphs of $\ln(\hat{P}^k(\varepsilon))$ as function of $\ln(\varepsilon)$; the estimated entropies are the distances between the graphs of $\ln(\hat{P}^k(\varepsilon))$ and $\ln(\hat{P}^{k+1}(\varepsilon))$ for the various values of k and ε.

These estimates of dimensions and entropies as function of $\ln(\varepsilon)$ for various values of k are given in the figures 2, 3 and 5, 6.

In the case of the time series generated by the Hénon system, we see that both the estimates of dimension and entropy cluster, for low values of ε (is 'distance') around a constant value (the big exception for the entropy is for $k = 1$ which we may disregard since we have to consider the limit for $k \to \infty$). The results agree reasonably well with the estimates reported in the literature: for the dimension around 1.2 and for the entropy around .32.

In the case of the time series generated by the autoregressive system, we see that for low values of k, the values of $\hat{D}^k(\varepsilon)$ converge to k as $\varepsilon \to 0$ — for larger values of k this should also happen, but the estimates stop because of the low numbers of pairs on which the estimation of the correlation integrals would have been based. In the estimation of the entropies we see that there is essentially no dependence on k but that the values diverge to ∞ as ε tends to zero.

3. Extremal value theory

Extremal value theory is concerned with the following situation. Let

$\{x_i(\omega)\}$ be a stationary time series (again ω is an element of (Ω, μ), the probability space of all possible realizations). For such a time series we define $M_k(\omega)$ to be the maximum of $x_1(\omega), \cdots, x_k(\omega)$. The main interest of extremal value theorie is the study of the distribution of $M_k(\omega)$ for large values of k. We denote the distribution function of $M_k(\omega)$ by $\mathcal{M}_k(x) = P(M_k(\omega) < x)$.

One may think of the following example: Let x_i be the maximum sea level in year i (we assume that this is a realization of some stationary process of which we write the general realization as $\{x_i(\omega)\}$). Decisions on how high dikes have to be will be based on (estimates of) quantities like the value of x for which $\mathcal{M}_{5000}(x) = .5$ (if one accepts a flood about once every 5000 years).

This example shows that one is interested in asymptotic results about $\mathcal{M}_k(x)$ for large values of k, and in particular for those values of x where $\mathcal{M}_k(x)$ is bounded away from 0 and 1. This means that one uses in general a rescaling of the x variable which depends on k. This is also illustrated in the next example.

Extremal values of an iid time series. We consider a stationary time series $\{x_i(\omega)\}$ in which all elements are statistically independent. We call such time series *independently and identically distributed*, abbreviated as *iid*. The distribution function for x_1, and hence for each x_i, is denoted by $\mathcal{F}(x) = P(x_1(\omega) < x)$. It is easy to see that, due to the independence, the distribution function $\mathcal{M}_k$ of the maxima of k successive elements is the k-th power of $\mathcal{F}$:

$$\mathcal{M}_k(x) = \mathcal{F}(x)^k.$$

From this simple result, one sees that the rescaling mentioned above, is indeed useful because the asymptotic behaviour of $\mathcal{M}_k(x)$ for fixed x and $k \to \infty$ is uninteresting: whenever $\mathcal{F}(x) < 1$, $\mathcal{M}_k(x)$ converges to zero.

In this example we get a complete description of the distribution of extremal values for iid time series. No such simple result exists for time series in which the values at different times are dependent. We give however in subsection 3.2 some examples of time series with dependence

for which one can analytically compute the distribution functions $\mathcal{M}_k$. Then, in subsection 3.3 we introduce the extremal index which quantifies the effect of the dependence for the distribution of maxima. Before doing so, we discuss in subsection 3.1 the relation between the distributions of maxima and our correlation integrals as defined before. Finally, in subsection 3.4 we discuss the interpretation of correlation integrals in terms of extremal indices.

3.1 Correlation integrals and extremal values

In this subsection we describe some constructions for transforming a given time series into others; for this reason it is convenient to use a single symbol to denote a time series. Let $\mathbf{X} = \{x_i(\omega)\}$ be a stationary time series (again ω is in (Ω, μ), the probability space of all possible realizations). From this we define the time series of differences $\mathbf{D}X = \{dx_i(\omega_1, \omega_2) = x_i(\omega_1) - x_i(\omega_2)\}$. Note that the probability space of all possible realizations for $\mathbf{D}X$ is the Cartesian product of Ω with itself (with the product probability measure). We denote a general element of $\tilde{\Omega} = \Omega \times \Omega$ by $\tilde{\omega} = (\omega_1, \omega_2)$. Next we define the time series of absolute values of differences **ADX** by $\mathbf{ADX} = \{adx_i(\tilde{\omega}) = |dx_i(\tilde{\omega})|\}$. Finally we denote by $\mathcal{M}X_k$ the distribution function of the maxima of k successive elements of **ADX**.

These definitions are arranged in such a way that $\mathcal{M}X_k(\varepsilon)$ as defined above is just equal to the k-dimensional correlation integral of our time series **X** at distance ε. This is the motivation to try to interprete the correlation integrals in terms of extremal value theory (especially for stochastic time series).

Before we proceed to exploit the relation between correlation integrals and extremal values, we want to point to some simple relations between the time series **X**, **DX**, and **ADX**.

Since **X** is stationary, also **DX** and **ADX** are stationary. The time series **DX** has average 0 and variance equal to twice the variance of **X**. By a simple computation one can see that the autocorrelations of **X** and **DX** are the same. Finally, the reconstruction distributions of **DX** are invariant with respect to the involutions $\mathbb{R}^k \ni X \mapsto -X \in \mathbb{R}^k$, i.e. with

respect to multiplication by -1 in $\mathbb{R}^k$. This is due to the fact that in $\tilde{\Omega}$ the elements (ω, ω') and (ω', ω) have 'the same probability'.

Next we compare the extremal value properties of **DX** and **ADX**. As before, we denote by $\mathcal{M}X_k$ the distribution function of the maxima of k successive elements of **ADX**. For the present argument we denote by min_k, respectively max_k, the distribution functions of the minima, respectively maxima, of k successive elements of **DX**, i.e.,

$$min_k(x) = 1 - P(dx_1(\tilde{\omega}), \cdots, dx_k(\tilde{\omega}) > x)$$

and

$$max_k(x) = P(dx_1(\tilde{\omega}), \cdots, dx_k(\tilde{\omega}) < x).$$

(We remind: in the definitions of distribution functions, we assume everything to be non-atomic so that we do not have to distinguish between $<$ and $\leq$.) From the fact that the reconstruction distributions of **DX** are invariant under multiplication by -1, it follows that

$$max_k(x) = 1 - min_k(-x).$$

If we may assume that the distribution of maximal and minimal values in sequences of k successive elements of **DX** are *independent*, then it would follow, for $x > 0$, that $\mathcal{M}X_k(x) = max_k(x)(1 - min_k(-x)) = (max_k(x))^2$. This assumption of independence seems to be very reasonable for large values of k, and values of x which are not too close to 0 or negative. However it cannot hold for low values of k or values of x which are close to zero or even negative. E.g., for $k = 1$ and $x > 0$ we have:

$$\mathcal{M}X_1(x) = 1 - 2(1 - max_1(x)) = -1 + 2max_1(x) < (max_1(x))^2.$$

(Note that in this last expression $x > 0$, so that $max_1(x) \in [\frac{1}{2}, 1]$.) For $x = 0$ and $k = 1$, the independence assumption gives $\mathcal{M}X_1(0) = (1/2)^2 = 1/4$ though this quantity is clearly zero.

Conclusions. In terms of autocorrelations (and hence power spectra), there is no difference between the original time series **X** and the time series **DX** of its differences. Under an independence assumption, which seems to be justified for larger values of k and x, but which does not hold for low values of k or x, the extremal values of **ADX**, the time

series of the absolute values of the differences, is related in a simple way to those of **DX**.

There is however no reason to expect a simple relation between the time series **X** and **ADX**, neither in terms of autocorrelations nor in terms of extremal values.

3.2 Extrema of time series with correlations: examples

We return to the notation at the beginning of this section. So $\{x_i(\omega)\}$ is a stationary time series, $M_k(\omega)$ denotes the maximum of the first k values, and $\mathcal{M}_k$ its distribution function. We have seen already for time series, where the various $x_i(\omega)$ are independent (i.e. for iid time series), that $\mathcal{M}_k = (\mathcal{M}_1)^k$. Since we have no such general result for time series with dependence, we discuss here some examples. In the next subsection we discuss some asymptotic results from the literature.

Example 1. In this example we start with an iid time series $\{y_i(\omega)\}$ and a fixed integer m. We denote the distribution function of $y_1(\omega)$, and hence of each $y_i(\omega)$, by $\mathcal{F}$. We define the elements of $\{x_i(\omega)\}$ as

$$x_i(\omega) = \max_{j=i,\cdots,i+m-1} y_j(\omega).$$

In this case it is very easy to see that the distribution functions $\mathcal{M}_k$ of the maxima of k successive elements can be given analytically as

$$\mathcal{M}_k = \mathcal{F}^{m+k-1} = \mathcal{M}_1^{1+\frac{k-1}{m}}.$$

This result admits a simple heuristic explanation by comparing it with the result for iid time series. The distribution of maxima of k successive values of this time series looks like the maximum of $1 + (k-1)/m$ successive elements of an iid time series with distribution function $\mathcal{M}_1$. This is to be expected, at least whenever $k-1$ is a multiple of m, because then a maximum below x means that the maxima of $1+(k-1)/m$ successive blocks of length m of the original time series $\{y_i(\omega)\}$, which was iid, are all below x: one block for $x_1(\omega)$ and one for each of the successive blocks of m elements of the remaining $x_2(\omega), \cdots, x_k(\omega)$.

Example 2. For this example we consider an arbitrary probability distribution P on $\mathbb{R}$ with distribution function $\mathcal{F}$, and a number $p \in (0,1)$.

Instead of giving a formal definition of this example as a stochastic process, we describe how to obtain inductively the generic realization $\{x_i\}_{i\geq 1}$. This goes in the following way: x_1 is a random element from the probability distribution P ; for each $i > 1$, x_i is equal to x_{i-1} with probability $(1-p)$ and with probability p it is a new and independent random choice of our distribution P.

From this definition it is clear that for the distribution functions of the maxima $\mathcal{M}_k$ we find: $\mathcal{M}_1 = \mathcal{F}$ and for $k > 1$:

$$\mathcal{M}_k(x) = \mathcal{M}_{k-1}(x)(1 + p(\mathcal{F}(x) - 1)).$$

The latter formula follows from the following argument: if the first $k-1$ elements of our time series are smaller than x, then the probability that also the k-th element is still smaller than x is $(1-p)$ (the probability that x_k equals x_{k-1}) $+ p$ (the probability that x_k is a new random choice) times $\mathcal{F}(x)$ (the probability that the new choice is smaller than x). From this we obtain

$$\mathcal{M}_k(x) = \mathcal{F}(x)(1 + p(\mathcal{F}(x) - 1))^{k-1}.$$

From calculus we know that for z close to 1, $1 + p(z-1)$ is well approximated by z^p. This means that, where $\mathcal{F}(x)$ is close to 1, we have in good approximation

$$\mathcal{M}_k(x) \sim \mathcal{F}(x)^{1+p(k-1)}.$$

We recall that here $\mathcal{M}_1 = \mathcal{F}$. So we see that this approximate result has the same form as in example 1, with $1/m$ replaced by p.

Example 3. Our third example gives a slightly different result. It is the so-called max-autoregressive random sequence (see [**FHR,1994**], section 8.2). In this example we start with an iid time series $\{y_i(\omega)\}$ with distribution function $\mathcal{F}$ and a real number $c \in (0,1)$. The max-autoregressive sequence itself, $\{x_i(\omega)\}$, is obtained as follows: The first element x_1 is a random choice from some probability distribution with distribution function $\mathcal{G}$, which we will specify later. The subsequent elements are defined as

$$x_i(\omega) = c \cdot \max(x_{i-1}(\omega), y_i(\omega)) \text{ for } i > 1.$$

In order to have this time series stationary, we should have

$$\mathcal{G}(x) = \mathcal{G}(x/c)\mathcal{F}(x/c).$$

For this we have to take

$$\mathcal{G} = \prod_{i=1}^{\infty} \mathcal{F}(x/c^i).$$

This infinite product converges automatically to a non-decreasing function. In this case the distribution functions of the successive maxima are given by

$$\mathcal{M}_k(x) = \mathcal{G}(x)(\mathcal{F}(x/c))^{k-1}.$$

If $\mathcal{G}(x)$ is, to a good approximation, equal to $\mathcal{F}(x/c)$ (which is the case if x is sufficiently close to the upper bound of the support of the probability distribution defined by $\mathcal{G}$), then we have here the same situation as in the iid case. So, under this assumption, here the dependence has no effect on the distribution of maxima, at least for values of x which are close to the upper bound of the support of the distribution defined by $\mathcal{G}$.

3.3 Extremal value theory: the extremal index

In this subsection we recall a main asymptotic result on extremal value distributions for time series with correlations. So let $\{x_i(\omega)\}_{i\geq 1}$ be a stationary time series, which we assume to be mixing in the sense of subsection 2.1. As before we denote by $\mathcal{M}_k(x)$ the probability that the first k elements are all smaller than x. For $s \in (0,1)$ and $k \in \mathbb{N}$ we define $x(s,k)$ by the equation $(\mathcal{M}_1(x(s,k)))^k = s$, which has a solution whenever the probability distributions of $x_1(\omega)$, and hence of any $x_i(\omega)$, is non-atomic. If the solution is not unique we just take one of them. Note that, if the time series $\{x_i(\omega)\}$ were iid, then $\mathcal{M}_k(x(s,k))$ would be equal to s; this was in fact the reason for taking this definition. Then we define $\theta(s,k)$ by $\mathcal{M}_k(x(s,k)) = s^{\theta(s,k)}$. So $\theta(s,k)$ quantifies the effect of the dependence in the time series on the distribution of maxima. The result which we recall, see [**LLR,1983**], section (3.7), is the following. The quantity $\theta_- = \liminf_{k\to\infty} \theta(s,k)$ is independent of s. The same holds for θ_+ which is similarly defined with liminf replaced

by limsup. These quantities θ_- and θ_+ are called the *lower, respectively upper, extremal index.*

Observe that in the iid case, where $\mathcal{M}_k = \mathcal{M}_1^k$ we have $\theta_- = \theta_+ = 1$. Next we consider the examples in the previous subsection. In the first example, where we had $\mathcal{M}_k = \mathcal{M}_1^{1+\frac{k-1}{m}}$ it is easy to see that $\theta_- = \theta_+ = 1/m$. In the second example the calculation is somewhat more complicated, but one can prove, as suggested by the approximation $\mathcal{M}_k \sim \mathcal{M}_1^{1+p(k-1)}$, that $\theta_- = \theta_+ = p$. For the third example, assuming that the support of the probability distribution defined by $\mathcal{F}$ is bounded from above, so that $\mathcal{G}(x) = \mathcal{F}(x/c)$ whenever $\mathcal{F}(x/c)$ is sufficiently close to 1, we have that $\theta_- = \theta_+ = 1$.

From these examples we see that it is not uncommon that θ_- and θ_+ are equal. In that case they are denoted by θ which is called the *extremal index.*

Especially from the first two examples we see that much stronger relations hold: in these examples one has $\mathcal{M}_k \sim \mathcal{M}_1^{\theta k+\alpha}$ with $\theta k + \alpha = \theta(k-1) + 1$ or $\alpha = 1 - \theta$. By numerical experimentation with other examples, see the next section, we found that the former relation $\mathcal{M}_k \sim \mathcal{M}_1^{\theta k+\alpha}$ is often satisfied to a good approximation, even for values of k that are not extremely big, but that the latter relation, namely the relation between α and θ, is usually not satisfied. We call the value α in the above description the *offset.* (Note that these numerical experimentations concern correlation integrals, or, in other words, extremal distributions of absolute values of differences of a time series.)

3.4 The analysis of correlation integrals in terms of extremal values

As we mentioned in section 3.1, the correlation integrals of a time series $\mathbf{X} = \{x(\omega)_i\}$ can be interpreted as the values of the distribution functions of maxima of k successive elements of the corresponding time series of absolute values of differences.

As we observed already in the context of entropy estimation, the differences $\log(P^k(\varepsilon)) - \log(P^{k+1}(\varepsilon))$ usually are independent of k for sufficiently big k, say for $k \geq k_0$. The above considerations suggest even

more: they suggest that for properly chosen coefficients c_k we have for $k \geq k_0$ that $\log(P^k(\varepsilon)) \sim c_k \log(P^{k_0}(\varepsilon))$ (with c_k independent from the value of ε), and even that the c_k depend afinely on $k \geq k_0$. We will see that this is indeed the case in our numerical simulations. Also we expect that, for a suitable value of c_1, we may have that $\log(P^1(\varepsilon)) \sim c_1 \log(P^{k_0}(\varepsilon))$. In our simulations this last similarity is less convincing. Both these expectations can be verified by visual inspection once the coefficients c_1, c_{k_0+1}, etc. have been found. If the result is positive, both the extremal index and the offset, as defined in the previous subsection, can be estimated.

If the similarity between $\log(P^1(\varepsilon))$ and $c_1 \log(K^{k_0}(\varepsilon))$ is not convincing, then the estimated value of c_1 will depend strongly on the algorithm used, and hence the derived values of the extremal index and offset will not give much intrinsic information about the time series. In this case there are two ways in which one can proceed. One is to base the estimation of the coefficients c_i on the estimated correlation integrals $P^k(\varepsilon)$ whose values are very close to one, and hence for which ε is close to the maximal possible distance — this is suggested by the definition of the extremal index, but the relevance of the correlation integrals for large values of ε, e.g. for judging the predictability of the time series, is small. Therefore we try a different way. Instead of basing our estimation of the extremal index on the coefficients $c_1, c_{k_0}, c_{k_0+1}, \cdots$, we replace c_1 first by the value, obtained from $c_{k_0}, c_{k_0+1}, \cdots$ by linear extrapolation, and then estimate the extremal index. We refer to this as the estimated *modified extremal index*. After the substitution of c_1 as proposed here, the offset has no independent information, so that we don't have to estimate it.

4. Discussion of numerical estimations of extremal indices and offset

In this section we discuss the results of a number of numerical simulation in which the indices and the offset were estimated for three different time series: the time series of the Hénon system and the autoregressive system

which we also used before, and also of an experimental time series from a fluid bed reactor at the Technical University Delft (provided by the group of prof. Van den Bleek). First we explain which calculations were carried out, then we discuss the results on the basis of graphical output of these calculations.

4.1 Description of the calculations performed

The calculations are based on the estimated correlation integrals as described in subsection 2.4, i.e., embedding dimensions from 1 to 20 and at distances

$$\varepsilon = 1, .9, .9^2, \cdots, .9^{59}.$$

For these calculations we have to take a value, here, as in the previous section, denoted by k_0, which is the embedding dimension above which we expect the asymptotic results to be valid.

The first thing we calculate are the coefficients, or *multiplication factors* $c_1, c_{k_0+1}, c_{k_0+2}, \ldots, c_{20}$ such that the graphs of $c_i \cdot \log(P^{k_0}(\varepsilon))$ is as close as possible to the graph of $\log(P^i(\varepsilon))$. This means that for each of the values of ε, for which correlation integrals were estimated, we computed the quotient $\log(P^i(\varepsilon))/\log(P^{k_0}(\varepsilon))$, and computed a weighted average over the relevant values of ε. In further considerations we always assume that $c_{k_0} = 1$.

Here the graphical output consists of a plot of the values of c_i as function of i: this is for a first inspection whether these values indeed depend affinely on i, i.e. whether the plotted points are on a straight line.

Next we actually show all the graphs of $c_i^{-1} \cdot \log(P^i(\varepsilon))$ (as function of $\log(\varepsilon)$ in order to inspect whether the multiplication with the coefficients c_i^{-1} indeed makes all these graphs (practically) equal. We call this the *rescaled log-correlation-integrals*. One may expect the graph for $i = 1$ to be usually somewhat exceptional. Therefore, this graph is shown by a dotted line instead of by a solid line, which is used for the other embedding dimensions, at least if the value of k_0 is different from 1.

Then we repeat the same, but with the coefficients c_i replaced by

approximations $\hat{c}_i$ such that $\hat{c}_i$ depends affinely on i. This also can serve as a test whether k_0 is already big enough to give the asymptotic picture for embedding dimensions above k_0. Here we speak of *log-correlation-integrals with uniform scaling.* Also here the graph for the embedding dimension 1 (if $k_0 > 1$) is shown as a dotted line while for the other dimensions solid lines are used.

Finally, we give estimates of the indices and offset. These estimates are based on the comparisons of the correlation integrals at successive embeddingdimensions i and $i+1$ for $i = k_0$ to $i = 19$. The estimates are the following:

- The estimate of the extremal index is obtained as the quotient $(c_{i+1} - c_i)/c_1$. In the graphical output it is shown as a solid line.
- Next the offsets are estimated for the same embedding dimensions and shown as stars. The estimates are based on an affine extrapolation from c_i and c_{i+1}, i.e. our estimate of the offset is $(c_i - i(c_{i+1} - c_i))/c_1$.
- Finally we show in the same graphical output the estimates of the modified extremal indices. Again for each i as above, they are also based on an afine extrapolation from the values of c_i and c_{i+1}. These estimates are represented by '+' in the graphical output.

4.2 Discussion of the results

We first consider the time series of the autoregressive system. We show the 'multiplication factors' (figure 7) and the 'indices and offsets' (figure 8) for $k_0 = 1$. Then we show the 'rescaled log-correlation-integrals' (figures 9, 11, 13) and the 'log-correlation-integrals with uniform scaling' (figures 10, 12, 14) for various values of k_0. As successive values of k_0 we take 1 (figures 9 and 10), 5 (figures 11 and 12), and 10 (figures 13 and 14). Even the log-correlation-integrals with uniform scaling with embedding dimension ≥ 10 fit very well — the agreement with the rescaled curve for the embedding dimension 1 is however weak. So here we see that all the information about the correlation integrals with embedding dimension 10 and higher (or even 5 and higher) can be reduced to the one curve shown and the extremal index and offset.

From the figure with indices and offsets we see the extremal index is by far the most stable, in the sense that it is (rather) independent of the embedding dimensions.

In the following examples we have $k_0 = 3$.

Next we show the results for a time series from the fluid bed reactor mentioned before (figures 15, 16, 17, 18). Here we see even a better agreement with what we expect to hold asymptotically. In particular, the relation between, the results for embedding dimension one with the results for higher embedding dimensions is much better. This means that the information about the correlation integrals at embedding dimensions three and higher can be very well described in terms of the extremal index.

The same analysis for the time series from the Hénon attractor (figures 19, 20, 21, 22) shows a completely different picture: the multiplication factors are not on a straight line and the rescaled log-correlation-integrals for different embedding dimensions do not coincide (this does not improve if we take higher values for k_0). The same holds for the log-correlation-integrals with uniform scaling. So here this method of contracting the information contained is the correlation integrals by these ideas of extremal value theory does not work. This is not completely unexpected: this time series, as generated by a deterministic system, is not mixing in the sense of stochastic processes (see section 2.1). On the other hand, in this case the dimension and entropy converge and hence contain the asymptotic information about the correlation integrals.

Concluding remark. We want to stress that our purpose in this paper is to show that estimates of the extremal index gives for certain time series a good description of the asymptotic behaviour of the correlation integrals. We did not yet optimize the estimation of these indices nor of the coefficients c_k which were shown in the figures. This will be the subject of further investigations.

References

[**B,1981**] D.R. Brillinger, Time series, McGraw-Hill, New York,1981.

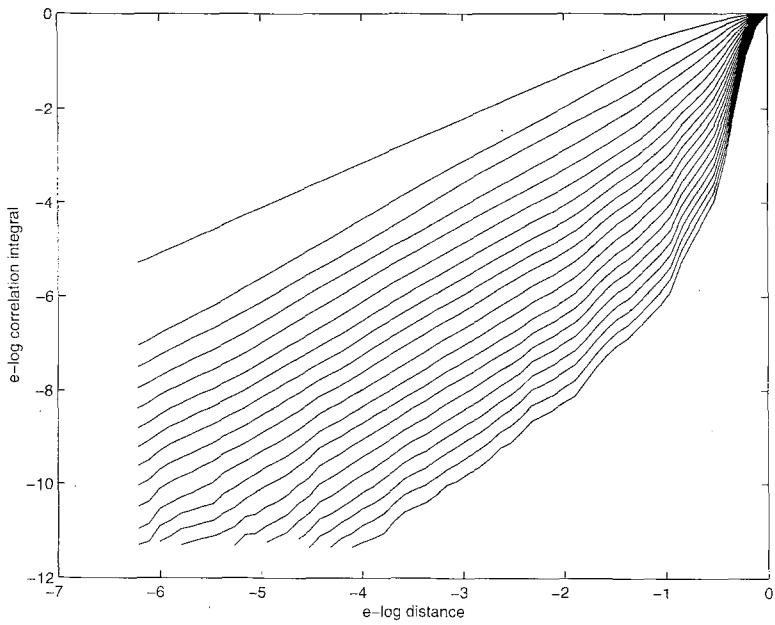

Figure 1: Henon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

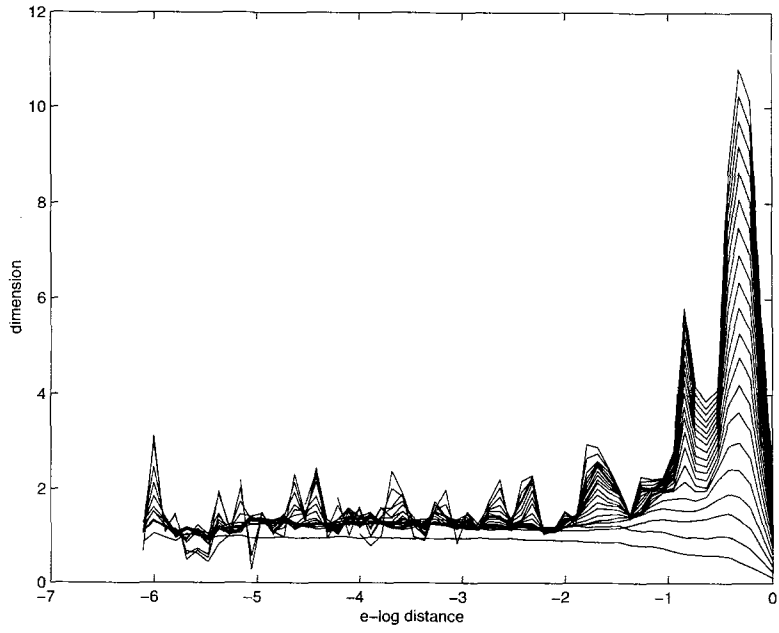

Figure 2: Henon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

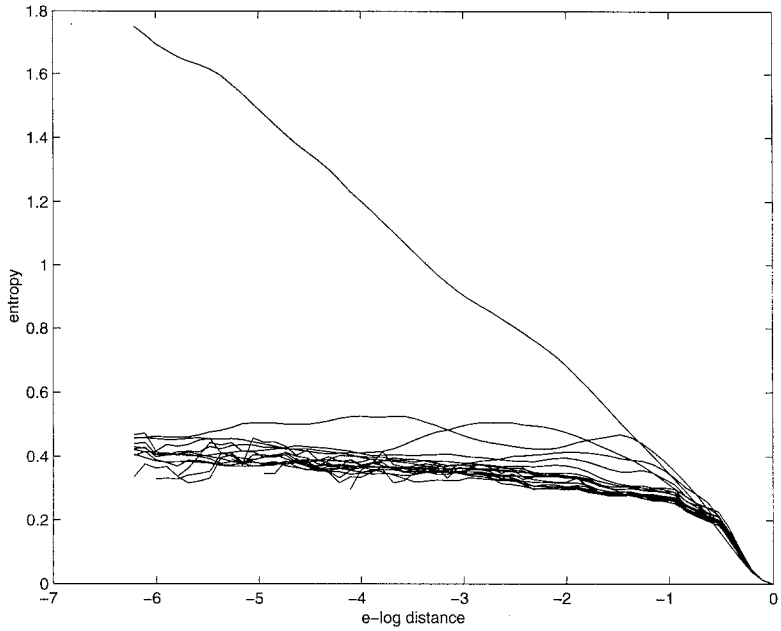

Figure 3: Henon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

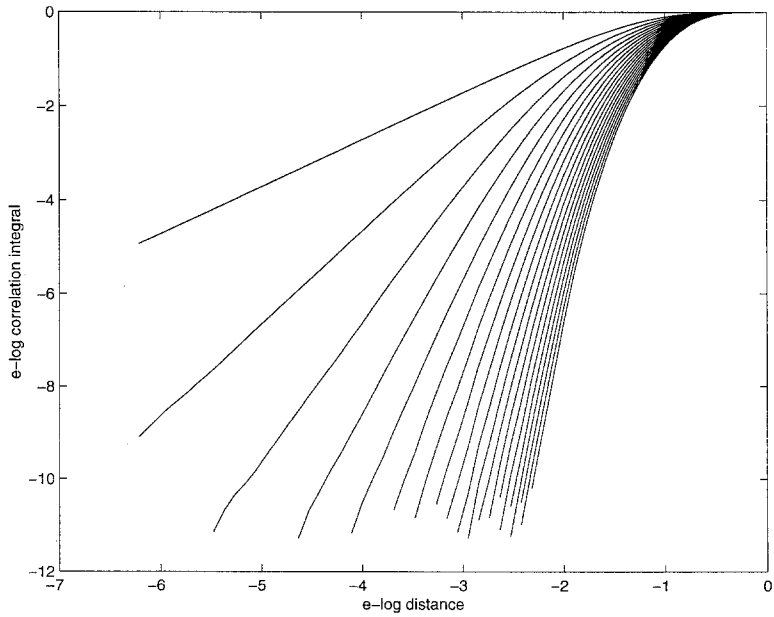

Figure 4: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

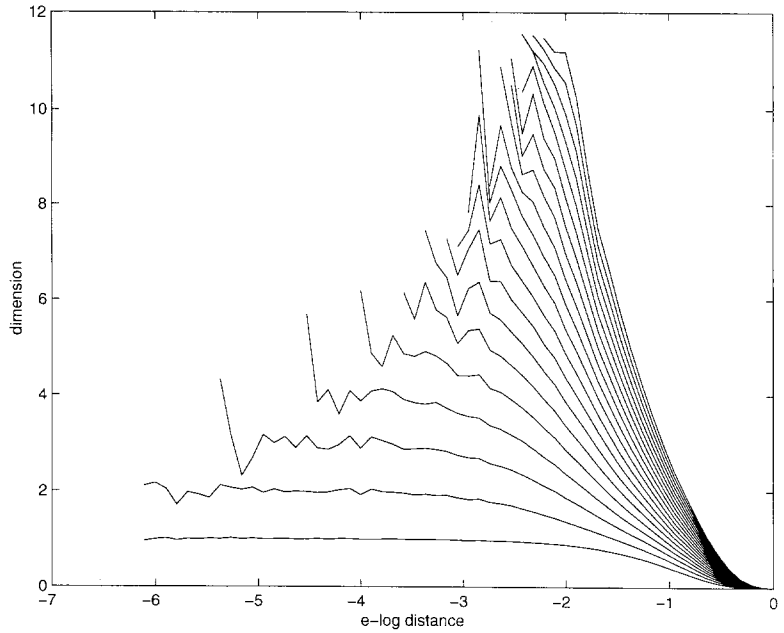

Figure 5: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

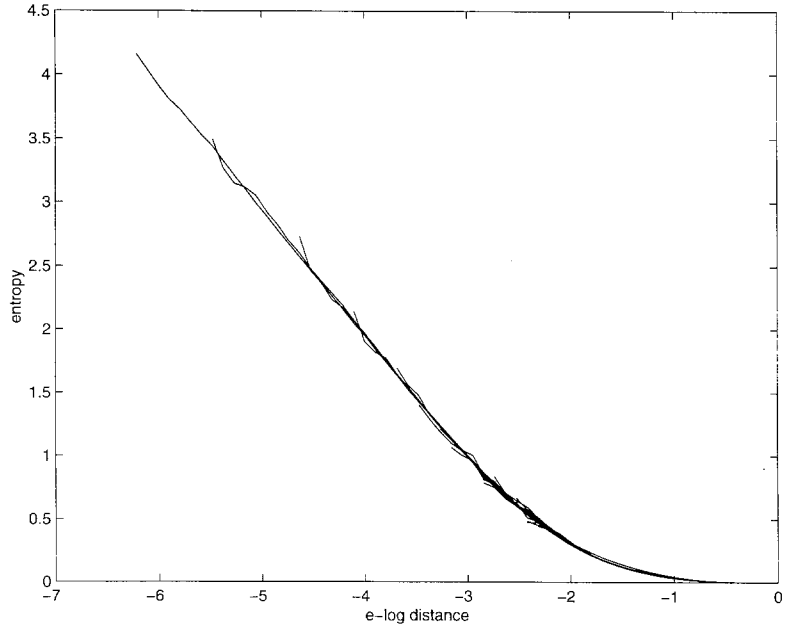

Figure 6: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

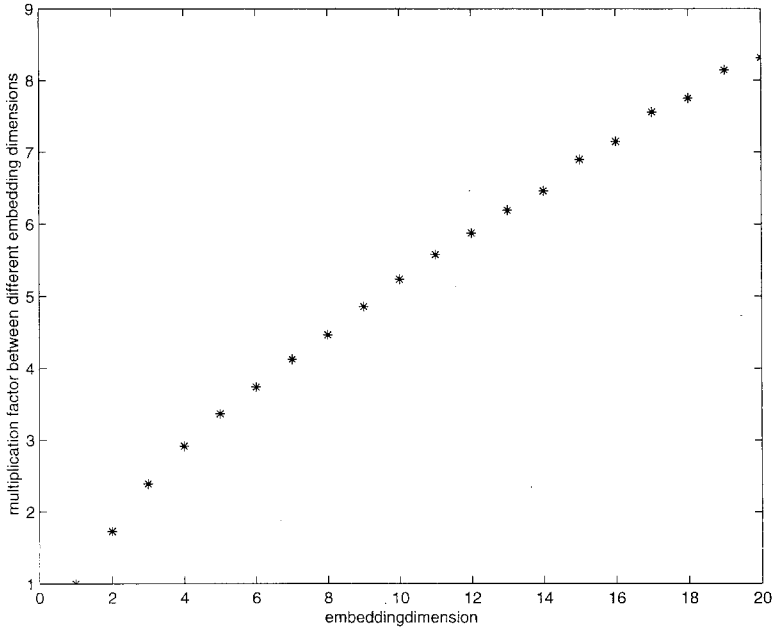

Figure 7: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

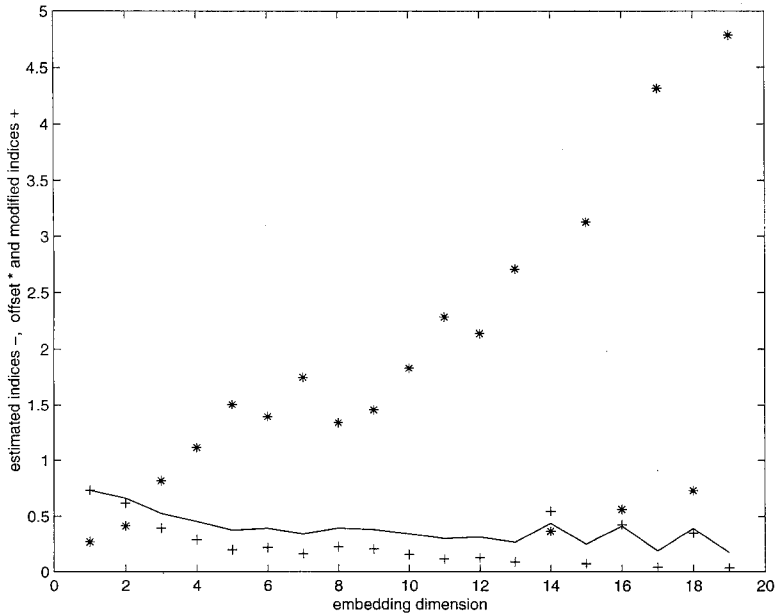

Figure 8: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

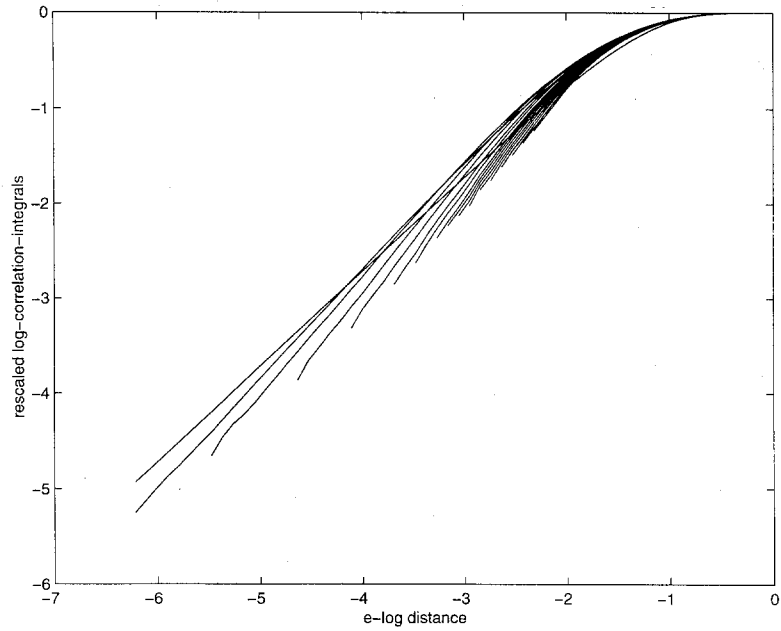

Figure 9: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

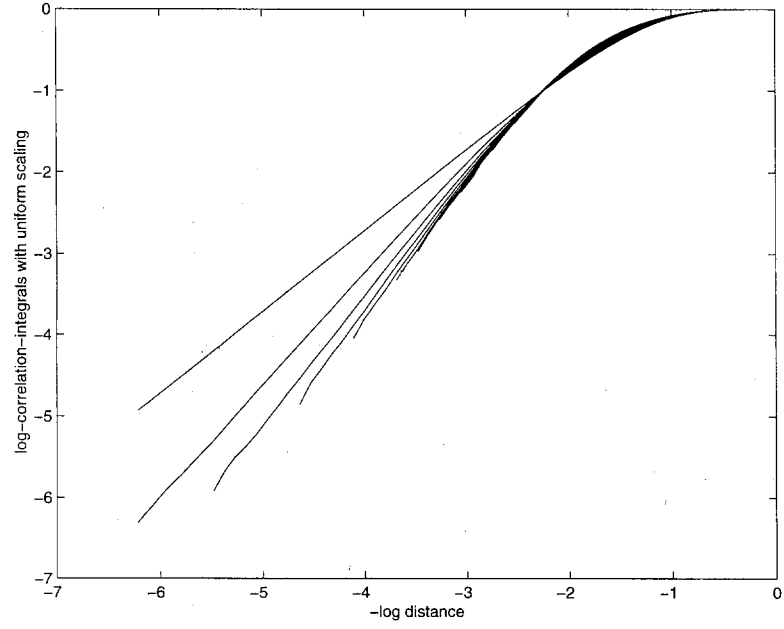

Figure 10: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

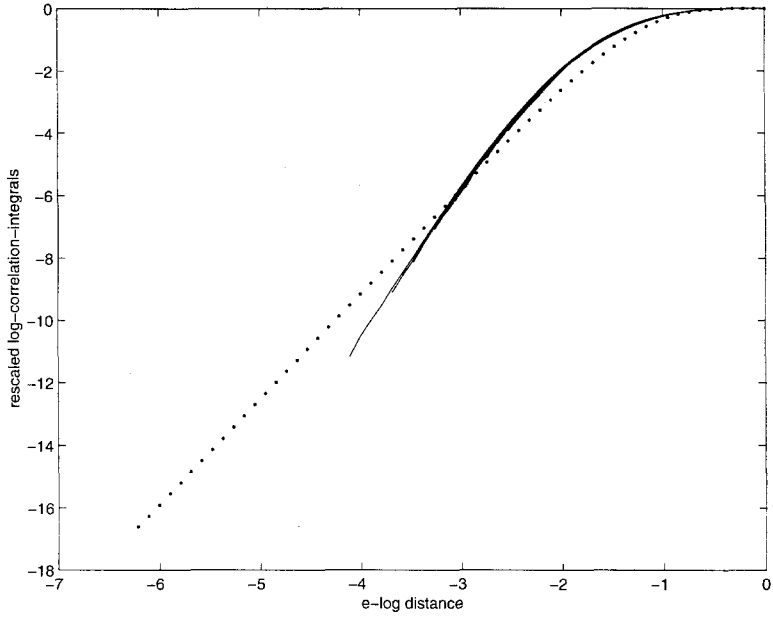

Figure 11: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

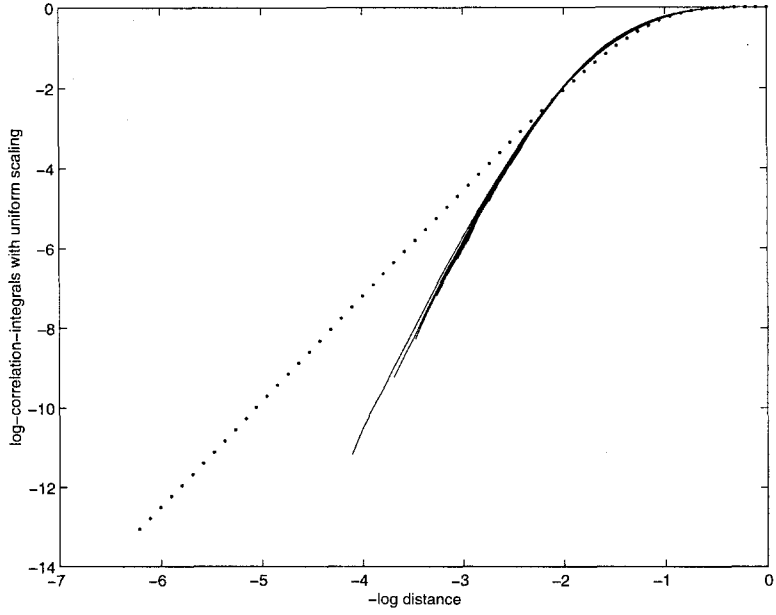

Figure 12: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

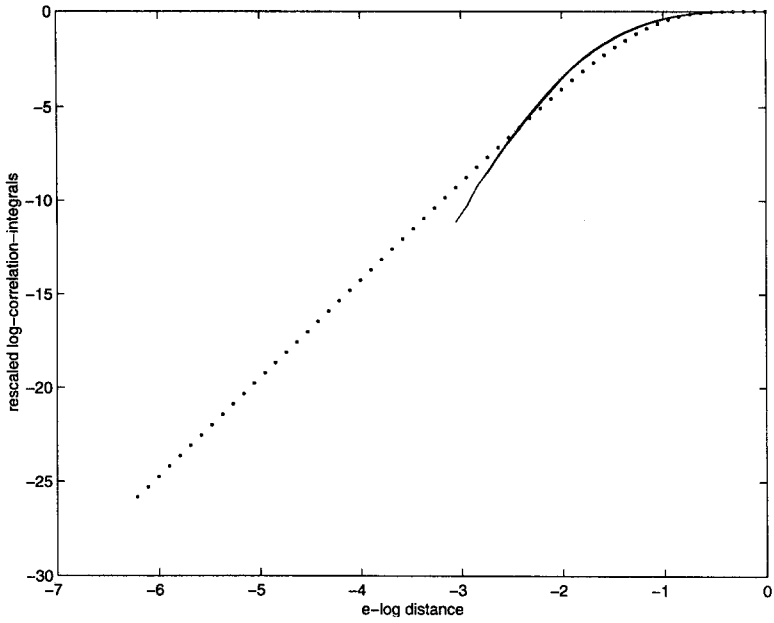

Figure 13: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

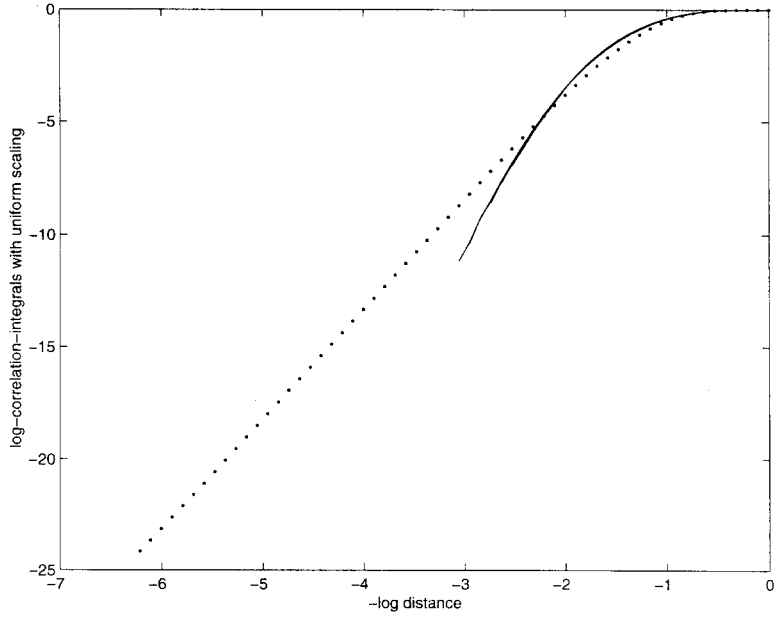

Figure 14: Autoregressive time series: $x_n = .9x_{n-1} + \text{noise}$.

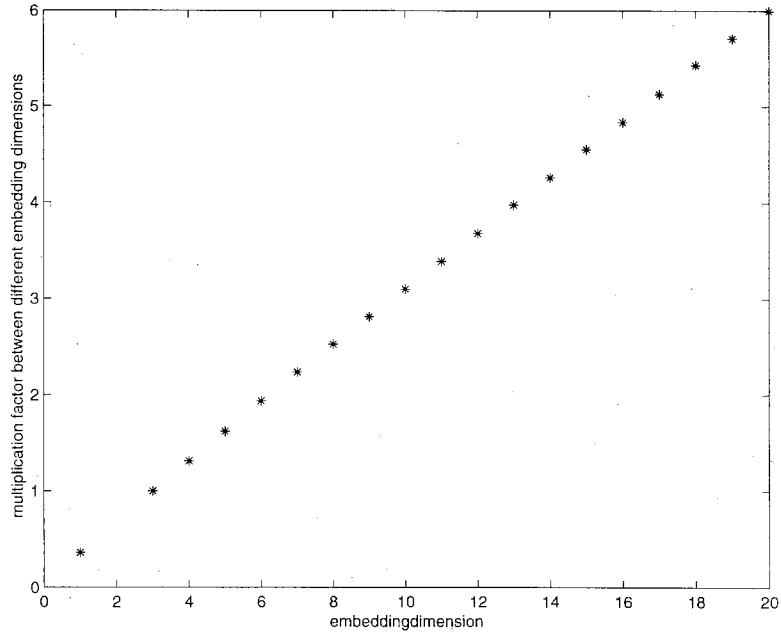

Figure 15: Time series from fluid bed.

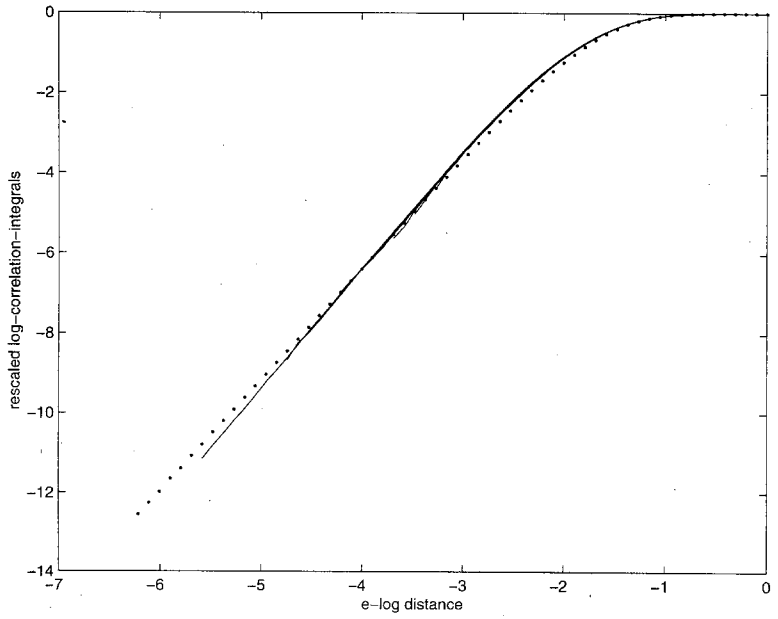

Figure 16: Time series from fluid bed.

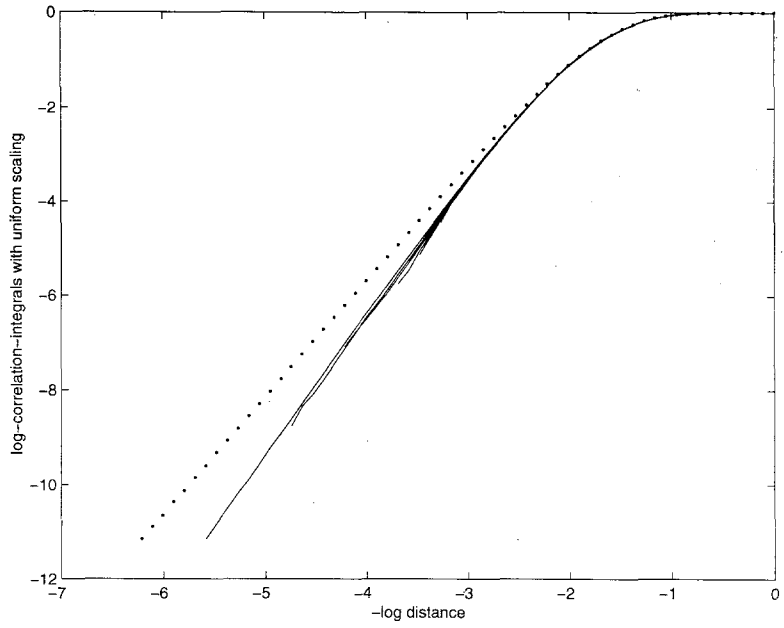

Figure 17: Time series from fluid bed.

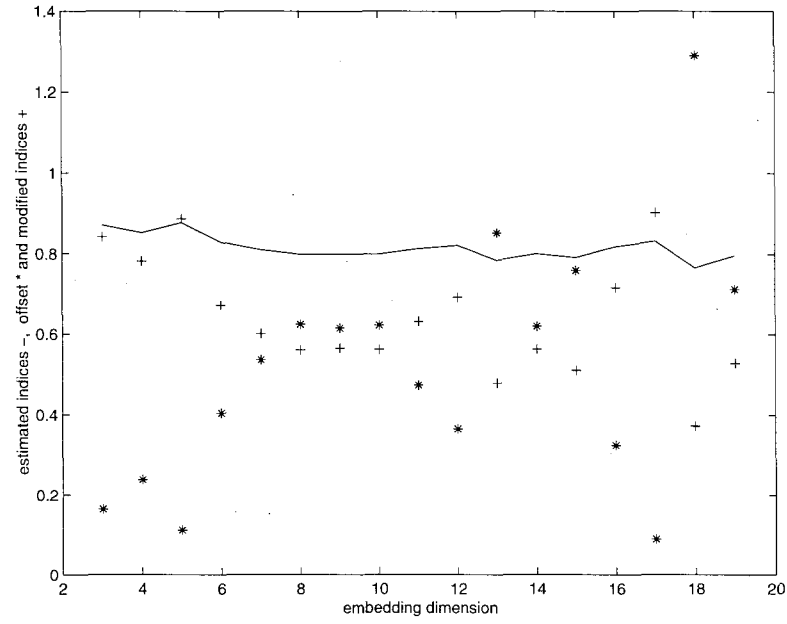

Figure 18: Time series from fluid bed.

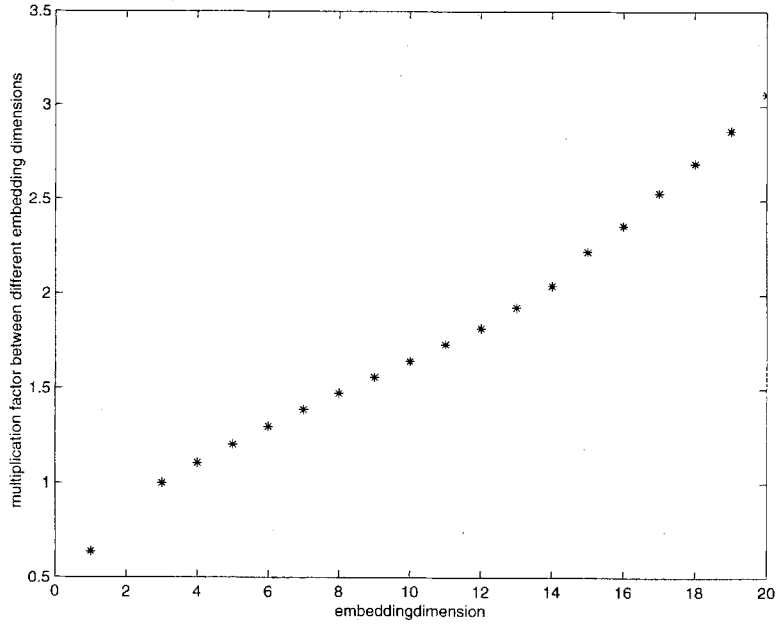

Figure 19: Hénon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

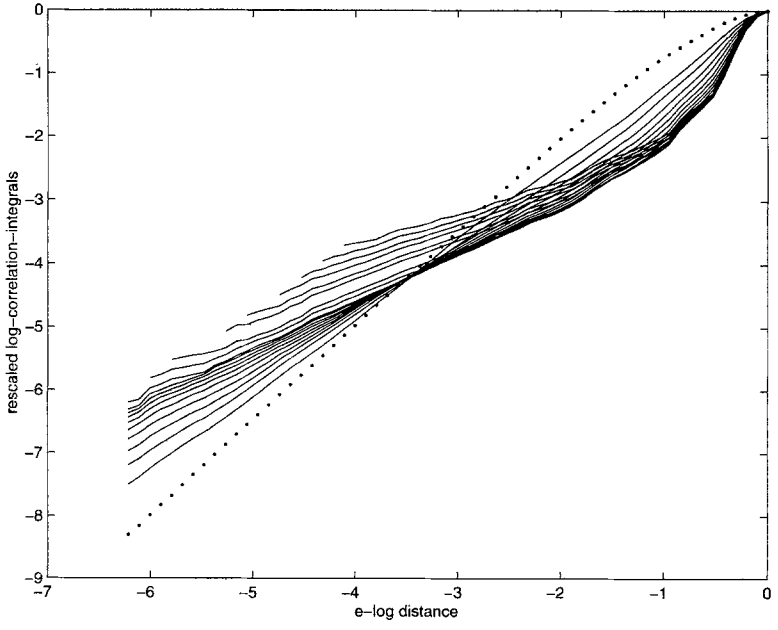

Figure 20: Henon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

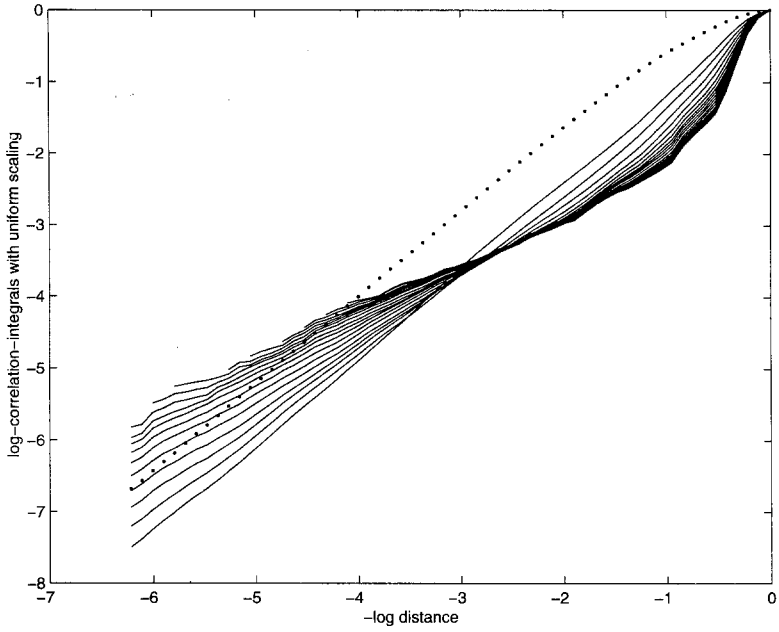

Figure 21: Henon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

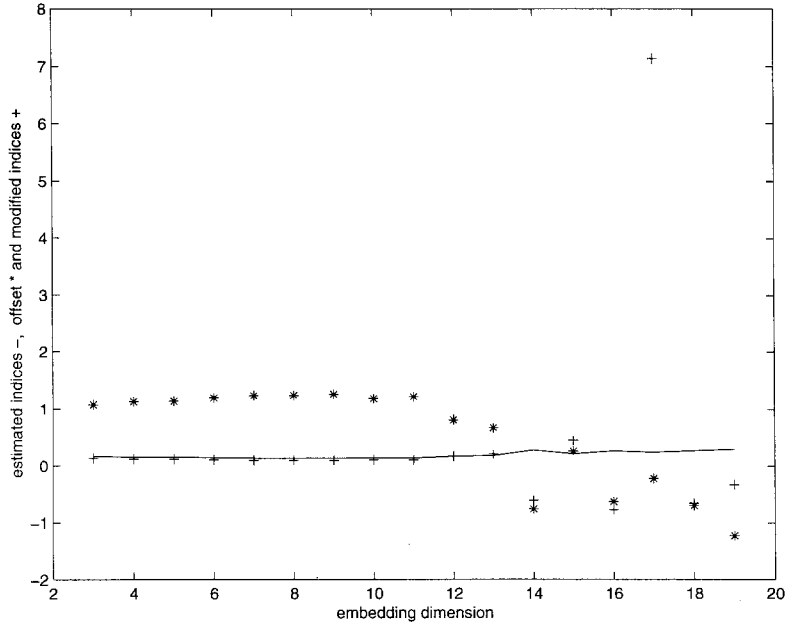

Figure 22: Henon-time series: $x_n = 1 - 1.4x_{n-1}^2 + .3x_{n-2}$.

[**CT,1992**] B. Cheng, H. Tong, On consistent nonparametric order determination and chaos, J. R. Statist. Soc. **B 54** (1992), 427 - 449.

[**FHR,1994**] M. Falk, J. Hüsler, R.-D. Reiss, Laws of small numbers: extremes and rare events, DMV Seminar, Band 23, Birkhüser, 1994.

[**GP,1983**] P. Grassberger, I. Proccaccia, Measuring the stangness of strange attractors, Physica **9D** (1983), 189 - 208.

[**GP,1983**] P. Grassberger, I. Procaccia, Estimation of Kolmogorov entropy from a chaotic signal, Phys. Rev. **A 28** (1983), 2591 - 2593.

[**KS,1997**] H. Kantz, T. Schreiber, Nonlinear time series analysis, Cambridge University Press, 1997.

[**LLR,1983**] M.R. Leadbetter, G. Lindgren, H. Rootzén, Extremes and related properties af random sequences and processes, Springer Series in Statistics, Springer-Verlag, 1983.

[**R,1956**] M. Rosenblatt, A central limit theorem and strong mixing condition, Proc. Nat. Acad. Sci. U.S.A. **42** (1956), 43 - 47.

[**SYC,1991**] T. Sauer, J.A. Yorke, M. Casdagli, Embedology, J. Stat. Phys. **65** (1991), 579 - 616.

[**T,1981**] F. Takens, Detecting strange attractors in turbulence, in: Dynamical systems and turbulence, eds. D. Rand, L.-S. Young, LNM 898 (1981), Springer-Verlag.

[**T,1983**] F. Takens, Invariants related to dimension and entropy, in: Atas do 13^o colóquio bras. de mat., IMPA, Rio de Janeiro (1983), 353 - 359.

[**T,1990**] J. Theiler, Estimating fractal dimension, J. Opt. Soc. Am. **7** (1990) 1055 - 1073.

[**T,1996**] F. Takens, Estimation of dimension and order of time series, in: Nonlinear dynamical systems and chaos, ed. H.W. Broer, S.A. van Gils, I. Hoveijn, F. Takens, Birkhäuser, 1996.

Floris Takens
Mathematisch Instituut
Rijksuniversiteit Groningen
PO box 800
9700 AV Groningen
Netherlands