Physica D 65 (1993) 373–398
North-Holland

SDI: 0167-2789(92)00028-S

PHYSICA D

Estimating the correlation dimension of an attractor from noisy and small datasets based on re-embedding

Klaus Fraedrich and Risheng Wang

Institut für Meteorologie, Universität Hamburg, W-2000 Hamburg 13, Germany
and Institut für Meteorologie, Freie Universität Berlin, W-1000 Berlin 41, Germany

Received 24 January 1992
Revised manuscript received 20 November 1992
Accepted 28 November 1992
Communicated by A.V. Holden

A modified version of the Grassberger–Procaccia algorithm is proposed to estimate the correlation dimension of an attractor. Firstly, a measured time series is embedded into an M dimensional phase space spanned by time delay coordinates. This, in turn, is linearly transformed into an equivalent space spanned by an orthogonal basis derived from singular value decomposition. Secondly, a subspace composed of the directions of the (first few) principal eigenvectors, is again embedded into a higher dimensional space, which is called re-embedding. Finally, the Grassberger–Procaccia algorithm is applied on a re-embedding space instead of the Takens' embedding and thereby the correlation dimension (D_2) is calculated. This leads to a modified version of the Grassberger–Procaccia algorithm, which is aimed at dealing with the estimation of the D_2 from noisy and relatively small data sets. In order to make full use of the available data, the delay time for the first embedding is always set to the sampling time. In order to reduce the noise level, only the principal components which are clearly above the "noise level" are used for the re-embedding. This modified algorithm is tested using low dimensional dynamical models with random noise of different levels. Here we have used the Lorenz model with D_2 about 2.0 and the Mackey–Glass equation with D_2 about 5.0. The results show that the present procedure gives a clearer scaling region in the $D_2(M, r)$–$\ln(r)$ diagram and thus a better estimate of D_2, especially when the data set is noisy and relatively small. This modified algorithm is applied to meteorological data and some of the problems associated with estimating the dimension of the weather and climate attractors are discussed based on the results.

1. Introduction

In recent decades, the theoretical understanding of some complex dynamical systems has undergone a rapid development, owing largely to the discovery of deterministic chaos [36,53,54, 10,11,13,26,58] and fractals [38]. It is now widely accepted that a nonlinear dynamical system may display a chaotic attractor whose dynamic behaviour is both deterministic and unpredictable and whose properties in phase space can be characterized in part by fractal dimensions. (Note that by *fractal dimensions*, we mean the dimension of a fractal object in a general sense as distinct from the dimension of a usual object. In specific contexts, the fractal dimension is specified by the concepts such as Hausdorff dimension, information dimension, correlation dimension, etc.) Encouraged by this progress, much effort has been devoted to detecting low dimensional attractors from measured signals of various dynamical systems. Most remarkable is the work on reconstructing the attractors using univariate time series based on the embedding theorem [63,46] and the algorithms for calculating the dimensions of the attractor (see for example [21,48,5,62]), among which the most commonly used one is the Grassberger–Procaccia algorithm [21]. These algorithms are a useful tool for identifying chaotic attractors from obser-

vations and therefore have been applied to measurements from various laboratory and natural and dynamical systems, particularly those relating to weather and climate time series. However, these algorithms normally require a very large and almost noise free data set in order to produce a reliable estimate of the dimensions [22,58]. This casts a certain doubt on the results of dimension analyses based on observations, since there is rarely a sufficient quantity of data and the data can never be noise free. Indeed, there are reports, based on various observations, that attractors of low dimension, ranging roughly from 3 to 7, exist in the weather or climate system [42,17,12,66,28,32], but there has been no indication of clear-cut scaling regions displayed in the form of the slope versus distance (scale) diagram (see section 2). Most of the existing algorithms work very well when dealing with low dimensional mathematical models or carefully controlled physical experiments and the scaling region of the attractors can be clearly demonstrated. However, when dealing with real observations, some kind of averaging or regression on a narrow scaling region must be performed in order to obtain an estimation of the dimension. Here some uncertainty occurs as the results are usually quite dependent on the scaling region being chosen. Moreover, there are no general prescriptions (theoretical or experimental) for the choice of time delay or the maximum embedding when the dimension is not known a priori. The effect of noise on the calculation of the dimension results leads to another source of uncertainty. Many studies published in recent years have been concerned with these problems; attempts have been made to estimate dimensions based on small data sets (e.g. [2,3]), or deal with noise reduction (e.g. [31,15,57]).

Broomhead and King [6] (see also [16,68]) propose a different approach by applying the singular value decomposition (or singular spectrum analysis [68]) to the trajectories in the embedding space to identify the principal directions spanned by the associated eigenvectors, which are linearly independent. The number of principal directions, which is determined by the associated eigenvalues above the "noise level", is thought to be related to the number of linearly independent variables involved in the observed dynamics and thus to its dimensionality [16,68]. (Note that the "noise level" refers to the flat tail of the eigenvalues associated with white noise. For coloured noise, matters are more complicated.) However, this number usually depends on the sampling time, the maximum embedding dimension and the time delay for the embedding [47]. Therefore it is difficult to estimate the dimension of the underlying attractor based on single observables using solely singular value decomposition, although it provides a powerful approach of time series analysis in terms of unstable, unharmonic limited cycle statistics, which has been shown to provide useful information about the underlying dynamics [18,20,52, 68].

As an alternative approach, Albano et al. [1] calculated D_2 with the Grassberger–Procaccia algorithm by using a state space spanned by the principal eigenvectors derived from the embedding space (instead of using the original time delay coordinates). The criterion for the choice of the number of principal components (PCs) involved in the calculation is somewhat arbitrary, and the result could be quite dependent on the choice of the window (M) and the number of PCs, especially when the data is noisy. In our approach, we will only use the first few PCs, those above the noise level, to estimate D_2. However, a re-embedding procedure is proposed to retain the static properties of the attractor under consideration (see section 2.4). We will show that, by applying the correlation integral to the re-embedding phase space, the resulting D_2 does not change with changes in the number of PCs, as long as they are all clearly above the noise level and given that the scaling exists.

In fact, the dynamics reliably represented by an observational data set is determined by the resolution in time and the length of the time

series, especially when the physical nature of the dynamics changes within time scale. Therefore, it is desirable to extract the dynamics with the reliably represented scale range before analysing its dimensionality. As chaotic signals (related to turbulence) usually obey a certain power law scaling, the larger scale dynamics carry more variance than the smaller ones. Therefore the singular value decomposition provides a useful approach for this purpose, because the variance distribution along the directions of the eigenvectors is determined by the associated eigenvalue (see section 2.2). The eigenvectors are linearly independent, but they are actually related in the sense that their cross(lag)-correlation is significant in most cases. This leads to the introduction of the re-embedding in section 2.4, which is aimed at capturing the dynamics reliably represented by the observed signals. Thus the modified version of the Grassberger–Procaccia algorithm introduced in this paper is different from the original version in that the correlation integration is performed in the re-embedding space (section 2.4). It is aimed at making full use of the available data and to reduce the noise levels.

The paper is structured as follows. As the time delay embedding technique, the singular value decomposition and the Grassberger–Procaccia algorithm are all basic to our approach, a brief review of each will be given in section 2. This will be followed by a description of the re-embedding procedure and the modified version of the Grassberger–Procaccia algorithm, concentrating on the correlation dimension (D_2) since it is the most commonly used measure for quantifying the static properties of a chaotic attractor. Section 3 will be devoted to testing the algorithms using the Lorenz model and the Mackey–Glass equation with different noise levels and different numbers of data points. In section 4, an application of this method to observational meteorological data will be presented and some of the problems associated with estimating the weather and climate attractors will be discussed. Finally, concluding remarks will be given in section 5. For the sake of convenience, all the details about the data used for testing and application will be given in the appendix.

2. Methodology

2.1. The time-delay embedding and the Grassberger–Procaccia algorithm

Assume the system under consideration is governed by a set of differential equations with K state variables ($\zeta \in \mathbb{R}^K$) and satisfies the well-known condition of existence and uniqueness theorem such that a solution curve, known as trajectory or orbit, passes uniquely through any given initial value $\zeta_0 \in \mathbb{R}^K$. Assume further that the trajectory of the system settles on an attractor or an attractive manifold (once the transients have decayed). An observable of a dynamical system can be viewed as a projection of a trajectory or a generic smooth function from the manifold to the reals. Let $x \in \mathbb{R}^1$ be one of the observables that is measured at a sampling interval δ during the course of an experiment, leading to a single time series with a total number of N discrete data points that covers a span of time $N\delta$,

$$\{x_t, t = 1, 2, 3, \ldots, N\} . \tag{1}$$

The attractor can be reconstructed, based on the embedding theorem [46,63,56], by a univariate generic time series (1) using a time delay coordinate with any generic time delay τ and a sufficiently large embedding dimension M. In the context of the singular value decomposition, the time span $M\tau$ is often referred to as the *window*:

$$\boldsymbol{Y}(M, t) = (x_{t-(M-1)\tau}, x_{t-(M-2)\tau}, \ldots, x_t) , \tag{2}$$

where $\boldsymbol{Y}(M, t)$ represents an M dimensional vector at a given time t, which forms $N_M \times M$ matrix as t runs from 1 to N:

$$\mathbf{Y} = \{\mathbf{Y}(M, t), t = 1, 2, 3, \ldots, N_M\} \tag{3}$$

where $N_M = N - (M-1)\tau$ and $\mathbf{Y}$ is a subset of $\mathbb{R}^M$. If $\mathbf{Y}$ is an embedding of the attractor, then some metric properties of the attractor, such as the dimensionality and characteristic exponents, will be preserved in $\mathbf{Y}$ (see for example [63,11]). This forms the basic idea for extracting dimensions from measured signals.

The dimensionality of an attractor characterizes its geometrical properties, measured by a certain dimension. If a measure ρ is assigned to the attractor, the dimension measures the power-scaling behaviour in the phase space where it sits, with respect to ρ. As a result, there are different definitions of dimension according to the definitions of ρ and specific aspects of the attractor one wishes to stress (see for example [40]). In this paper, we will concentrate on the correlation dimension which is defined on the basis of probability measure (see [21]). The correlation integral is defined in the M-dimensional reconstructed space (3) as the probability of finding a pair of vectors whose distance is not larger than r:

$$C(M, r) = \frac{1}{N^2} \times \sum_{|i-j|>\tau_0} \Theta(r - |(\mathbf{Y}(M, i) - \mathbf{Y}(M, j)|)\,, \tag{4}$$

where Θ is the Heaviside step function and τ_0 is the correlation time due to dynamics [64]. According to Grassberger and Procaccia [21], the correlation dimension D_2 is derived from

$$D_2 = \lim_{r\to 0} D_2(M, r) \tag{5}$$

for sufficiently large M, where $D_2(M, r)$ is the slope of $C(M, r)$:

$$D_2(M, r) = \frac{\mathrm{d}\ln(C(M, r))}{\mathrm{d}\ln(r)}\,. \tag{6}$$

Fig. 1a gives an example using the first variable (x) of the Lorenz model (see appendix). Since the calculation of D_2 for very small r is dominated by noise, particularly when dealing with observational data, there is no convergence of the slope for very small r. In practice, for a given M, D_2 is defined in a scaling region (r_1, r_2),

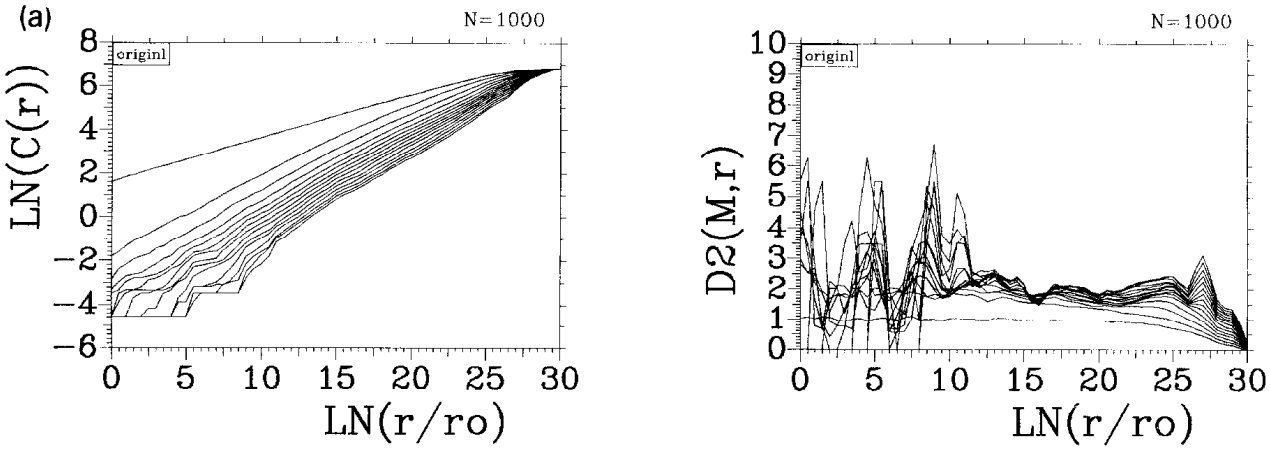

Fig. 1. (a) $\ln(C(M, r))$–$\ln(r)$ (left) and $D_2(M, r)$–$\ln(r)$ (right) derived from the original Grassberger–Procaccia algorithm (see section 2.1) for the first variable of the Lorenz model (see appendix). The number of data points $N = 1000$. The time delay $\tau = 2$ and the maximum embedding $M = 15$. The scaling region is indicated by the flat plateau of the $D_2(M, r)$–$\ln(r)$ diagram. (b) The same as (a), but obtained from the phase space of all the three variables (x, y, z). (c) The same as (a) but by using two variables (x, y) with maximum embedding $M = 10$. (d) The same as (a) but by using the re-embedding procedure (17).

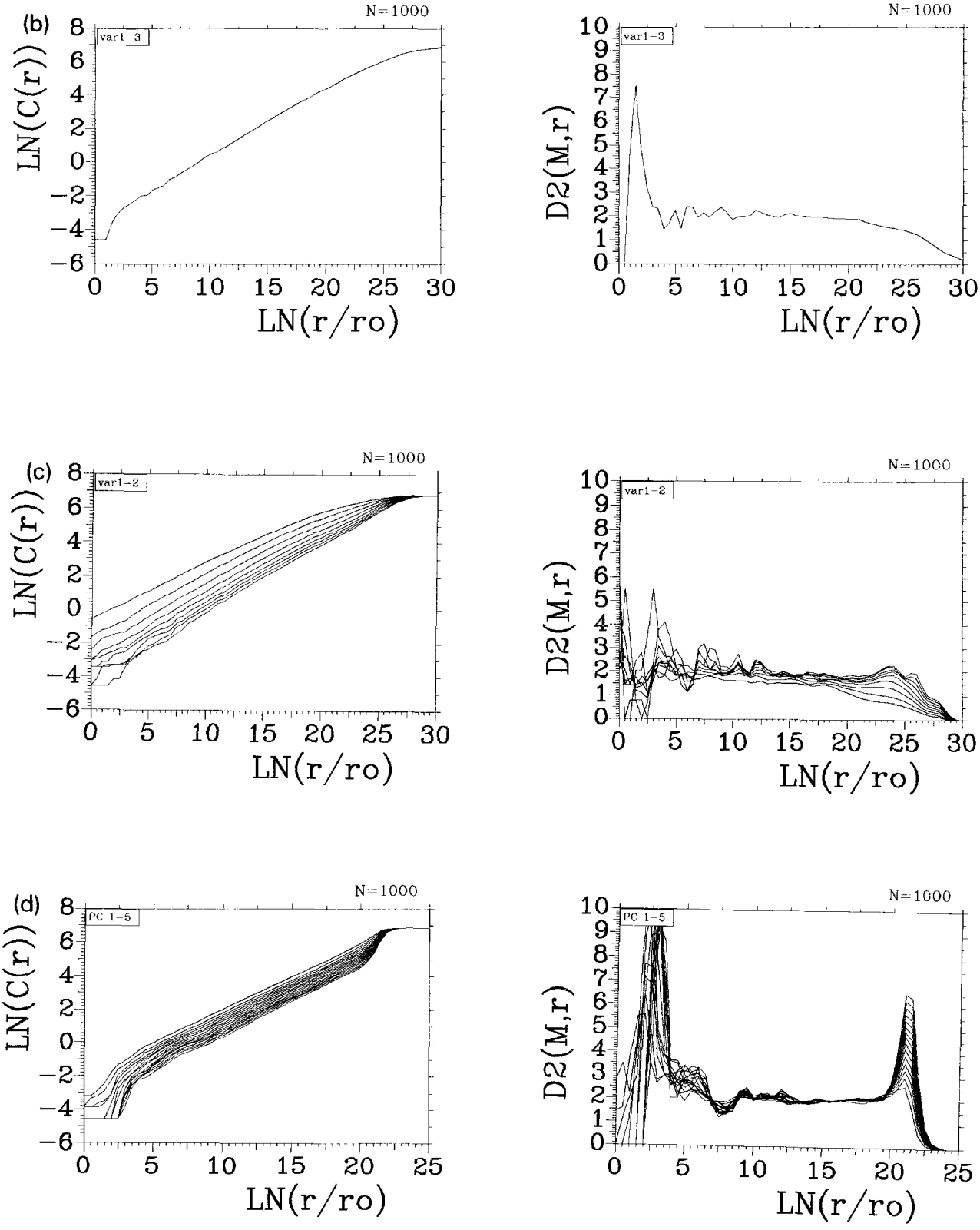

Fig. 1 (cont.).

where the $D_2(M, r)$–$\ln(r)$ diagram displays a flat "plateau", i.e. $D_2(M, r)$ does not change with r (see fig. 1 for example). If such a region cannot be detected in the $D_2(M, r)$–$\ln(r)$ diagram, it is impossible to give a reliable estimate of the dimension of the attractor or to conclude whether the scaling about the attractor exists. Furthermore, due to the limitation of the number of data points, M cannot be too large. One reasonable way to choose M is to plot $D_2(M)$, which is defined in the scaling region (r_1, r_2), against M and to find the minimum $M_{\min}$, beyond which $D_2(M)$ does not change with increasing M. For a random process, it was believed that there would be no saturation of D_2 with increasing M, which, however, is not the case for some coloured noises [45,51]. This needs not confuse good dimension calculations [65].

2.2. *Transformation of the embedding space and singular value decomposition*

The dimension of an attractor is a function of an invariant measure. The reconstruction of an attractor in embedding space is only needed to recover its invariant properties. If an invariant measure exists, the dimension will not be changed when the space is mapped into another space, given that the mapping is homeomorphic [56]. As each component of the embedding vector (2) is derived from the same time series, their variances are equally distributed among the time delay coordinates. On the other hand, chaotic attractors, unlike random processes, display large scale structures, although their dynamical behaviour is apparently unpredictable beyond certain time limits. Therefore the time delay coordinates (2), (3) are not orthogonal. It is possible to transform the embedding space into an equivalent space whose coordinates are linearly independent (orthogonal). Such an orthogonal basis can be conveniently generated by singular value decomposition [6,30,1]. As the singular value decomposition has become a standard method for time series analysis, we will not give a detailed treatment of its formalism, but a brief summary:

Let $\mathbf{A}$ be the covariance matrix of $\mathbf{Y}$ (see (3)):

$$\mathbf{A} = \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{Y} . \tag{7}$$

Thus $\mathbf{A}$ is a positive definite symmetric matrix and can be transformed into a diagonal matrix Λ by a series of rotations, which can be expressed as:

$$\Lambda = \mathbf{V}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{V} = \mathbf{V}^{\mathrm{T}} \cdot \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{Y} \cdot \mathbf{V} \tag{8}$$

where $\Lambda = \{\delta_{i,j}\lambda_i;\ i, j = 1, 2, \ldots, M\}$. In this way, a spectrum of non-negative values on the diagonal of Λ can be extracted, which are usually ordered by magnitude and called singular values or eigenvalues (in our case). The transformation matrix $\mathbf{V}$ is an $M \times M$ orthogonal matrix whose columns are called the eigenvectors:

$$\mathbf{V} = (V_1, V_2, \ldots, V_M) , \tag{9}$$

which form a natural orthogonal basis (or coordinates). The projection of $\mathbf{Y}$ onto this basis provides M linearly independent variables (given that all the eigenvalue are above 0, which is the case when the data is noisy.):

$$\mathbf{P} = \mathbf{Y} \cdot \mathbf{V} = (P_1(t), P_2(t), \ldots, P_M(t)) \quad \text{with } t = 1, 2, \ldots, N_M , \tag{10}$$

which are usually referred to as the principal components (PCs). From (8) it follows that

$$\sum_{t=1}^{N_M} P_i(t) \cdot P_j(t) = \delta_{i,j} \cdot \lambda_i . \tag{11}$$

That is, the variance of each of the components P_i is equal to the corresponding eigenvalue λ_i. For a time series that obeys some power-law scaling, λ_i is usually ordered by time scales. In most cases, only a few PCs are necessary to represent the main features of the observed dy-

namics. In this way, the relatively small scale fluctuations are cut off from the time series. This has some practical advantages, because the noise level usually increases with decreasing scale for observational data. However, it is still difficult to say how the number of significant principal components is related to the dimensionality of the dynamics underlying the time series, because the number of the principal modes above the "noise level" usually changes with the details of the embedding in (2). The modification of the Grassberger–Procaccia algorithm introduced in section 2.4 is an attempt to overcome this difficulty.

2.3. Multichannel embedding

The variables which completely describe the dynamics of a system are not mutually independent if the governing equations are nonlinear (see the Lorenz model in fig. 2a, for example). Suppose all these variables are known and measured, then the Grassberger–Procaccia algorithm can be directly applied to the real phase space formed by these variables to estimate the effective dimension (see fig. 1b for example). In this case, there is, of course, no need for embedding. Moreover, a practical way to estimate the number of linearly independent variables may be achieved by singular value decomposition applied directly to the trajectory in the real phase space. Accordingly, the three variables (x, y, z) involved in the Lorenz model can be projected onto three principal components (P_1, P_2, P_3) using (10). The variance of each component is represented by its associated eigenvalue (λ_i, $i = 1, 2, 3$). Fig . 2b shows the variance contribution of P_1, P_2 and P_3 changing with increasing number of data points (N). Note that the asymptotic dynamics of the Lorenz system, whose dimension is about 2.0, are dominated by the first two modes which account for 94% of the total variance.

However, when more than one of the variables of a dynamical system are measured, one should expect that the information contained in the subspace formed by these known observables is greater than that contained in a single variable for the same length of time series [11,44,61]. Therefore it is desirable, instead of embedding a single time series, to embed the subspace of these variables into a higher dimensional space to make full use of the information and to provide a better reconstruction of the attractor.

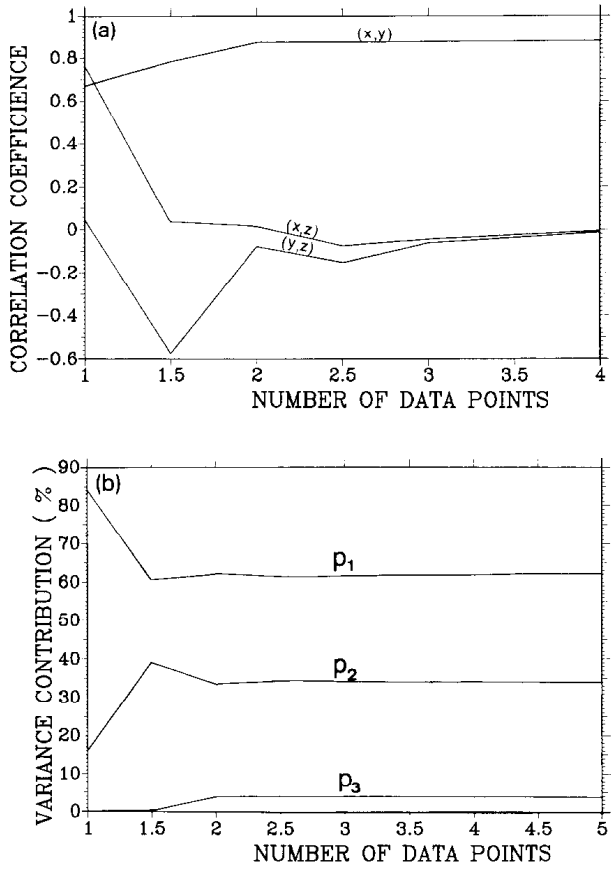

Fig. 2. (a) The correlation between different variables ((x, y), (x, z), (y, z)) of the Lorenz model (see appendix) with increasing number of data points ($\log_{10}$). (b) The distribution of variance among the three principal components (P1, P2 and P3) with increasing number of data points ($\log_{10}$).

Let $\boldsymbol{X}_t$ represent the L measured observables corresponding to the x_t of (1):

$$\boldsymbol{X}_t = (x_t^1, x_t^2, \ldots, x_t^L) \quad \text{and } L < K\,. \tag{12}$$

Thus all the measurements of $\{\boldsymbol{X}_t,\ t = 1, 2, \ldots, N\}$, define a subspace. In the same way as (2), this subspace will be embedded into an

M-dimensional pseudo-phase space using the time-delay technique:

$$\boldsymbol{Z}(M, t) = (\boldsymbol{X}_{t-(M-1)\tau}, \boldsymbol{X}_{t-(M-2)\tau}, \ldots, \boldsymbol{X}_t) , \quad (13)$$

which may be referred to as *multichannel embedding* as distinct from the embedding (2) using single time series. The effective embedding dimension of (13) is $L \times M$. The trajectory in the multichannel embedding space (13) can be expressed as an $(L \times M) \times N_n$ matrix $\mathbf{Z}$:

$$\mathbf{Z} = \{\boldsymbol{Z}(M, t); t = 1, 2, 3, \ldots, N_M\} . \quad (14)$$

Applying (4) to $\mathbf{Z}$ instead of $\mathbf{Y}$ (in (3)), an estimation of D_2 of the attractor can be achieved. This is shown to be plausible by an example using the Lorenz model where the first two variables (x and y) are used (see fig. 1c). Comparing fig. 1a and fig. 1c one observes a clearer scaling in the latter. In addition, if $\boldsymbol{X}_t$ in (12) is a spatial-temporal variable, then, application of the singular value decomposition to $\mathbf{X} = \{\boldsymbol{X}_t, t = 1, 2, \ldots, N\}$ and to $\mathbf{Z}$ leads to the standard empirical orthogonal function (EOF, see [35,33]) analysis and to an extended EOF analysis [69], respectively. The latter is also called multichannel singular spectral analysis [29] which is very useful for studying spatially extended systems (see for example [18]). However, the main concern of this paper is the problem associated with the estimation of D_2 based on a single variable time series. This leads to the introduction of the re-embedding discussed in the next subsection.

2.4. Re-embedding and the modified Grassberger–Procaccia algorithm

For a single variable time series, it is well known that singular value decomposition can provide a set of linearly independent variables (10). However, the cross(lag)-correlations of these variables are usually significant, which can be seen from the phase portraits of the first 5 principal components (fig. 3). Note that, in order to make full use of the data, we always keep $\tau = 1$. Let the main features of the dynamics corresponding to the observation be represented by the first m ($<M$) components (see (10)):

$$\boldsymbol{P}'(t) = (P_1(t), P_2(t), \ldots, P_m(t) , \quad t = 1, 2, \ldots, N_M) . \quad (15)$$

The essential point of our method is to treat these first m principal components (15) in the same way as the multivariate time series (12) and to embed the subspace formed by (15) into a higher dimensional space using a proper time delay (τ) (see the third point of the discussions in section 3.3) and embedding (M') in the same way as (2) and (13):

$$\boldsymbol{Y}'(M', t) = (\boldsymbol{P}'_{t-(M'-1)\tau}, \boldsymbol{P}'_{t-(M'-2)\tau}, \ldots, \boldsymbol{P}'_t) . \quad (16)$$

We call the embedding (16) the re-embedding as distinct from the embedding of the primary embedding of Takens [63], because it is the embedding of the embedding space (2). The theoretical justification for the concept of re-embedding can be found in [56]. Suppose that the first embedding (2) is an embedding of the underlying dynamics of the system, then the principal components (15) can be understood as a new set of variables obtained by measuring the trajectory (3) in the original embedding space (2) using a set of "instruments" (linear transformations (10)). The new "measurements " retain some information of the underline dynamics. Thus a proper embedding of these new variables (i.e. the re-embedding (16)) may be an embedding of the original embedding (2) and, therefore, an embedding of the underlying system. If the original embedding (2) is not a sufficient embedding, i.e. M is not large enough (due to some reason, we suggest to choose M moderately large. See the section 3.3 for discussion), then it is a projection of the embedding. By using the re-embed-

ding, the whole attractor is hopefully reconstructed based on this projection. In principle, (15) can be any combination of linear transformations, such as filters, instead of the principal components. However, the advantages of the singular value decomposition (see, for example, [6,68] and section 2.2) make it an optimal choice of linear transformation for making full use of the available data, since the main features of the underlying dynamics are captured by the first few PCs used for the re-embedding (16).

The correlation integral (4) on the re-embedding space (16) can then be written as:

$$C(M', r) = \frac{1}{N^2} \times \sum_{|i-j| \geq \tau_0} \Theta(r - \|\boldsymbol{Y}'(M', i) - \boldsymbol{Y}'(M', j)\|) \quad (17)$$

where τ_0 should not be less than the window M. D_2 is deduced accordingly through (5), (6), substituting $C(M, r)$ by (17). This leads to a modification of the Grassberger–Procaccia algorithm. An example derived from the Lorenz model is given in fig. 1d. Note that, like the multichannel embedding case in fig. 1c, the minimal embedding dimension of 1 does not exist in the re-embedding as long as $m > 1$, because the effective embedding dimension of the re-embedding (16) is $m \times M'$. Note also that the slope of the correlation integral often shows a high peak at large length scales before settling down to the plateau value (see fig. 1d and also fig. 5 and fig. 7). This can be understood to result from the macroscopic structure of the reconstruction (L. Smith, personal communication). For instance, a 1-dimensional limit cycle embedded in 3 dimensions is usually not confined to a 2-dimensional plane; at large scales the slope of the correlation integral reflects the extension in all 3 directions of this one dimensional curve, while at small scales its one dimensional nature would be seen. (Of course, this macroscopic structure is not invariant under transformation or changes in the measurement function.)

From fig. 1 it can be seen that all four analyses (see figs. 1a–d) produce a D_2 which is close to 2.0. However, the modified Grassberger–Procaccia algorithm (17) provides a more clear-cut scaling region compared with the original one (4) (compare figs. 1a and 1d). In the following section, we will discuss some of the practical advantages of (17) in estimating D_2 for noisy and relatively short time series. Additional tests concerning the effects of noise and the number of data points will be presented using the Lorenz model ($D_2 = 2.0$) as an example of a low dimensional system and the Mackey–Glass equation (whose parameter will be chosen such that $D_2 = 5.0$, see appendix) as an example of a somewhat higher dimensional case.

3. Tests of the algorithm: noisy and relatively short time series

One of the fundamental problems involved in the analysis of observational data is how to assess the influence of the noise contained in the data and how possibly to eliminate it. Various noise reduction procedures have been proposed in recent years (e.g. [31,57,25]), which are very useful for real data analysis. In addition, singular value decomposition can also be used to assess the noise level for a given time series [68]. We hope that the re-embedding method described in section 2.4 could also be used as a noise reduction procedure in the calculation of D_2, since only the dominant components (those above the noise level) are involved in the calculation (see sections 2.2 and 2.4).

Another problem with observational time series is the limitation of the number of data points [40]. There are theoretical considerations about the requirement for the minimal number of data points for estimating a dimension [60,55,41]. However, in the context of applications, one is often confronted with the question of how to get an optimal estimation using the *available* data sets, that is, how to extract the

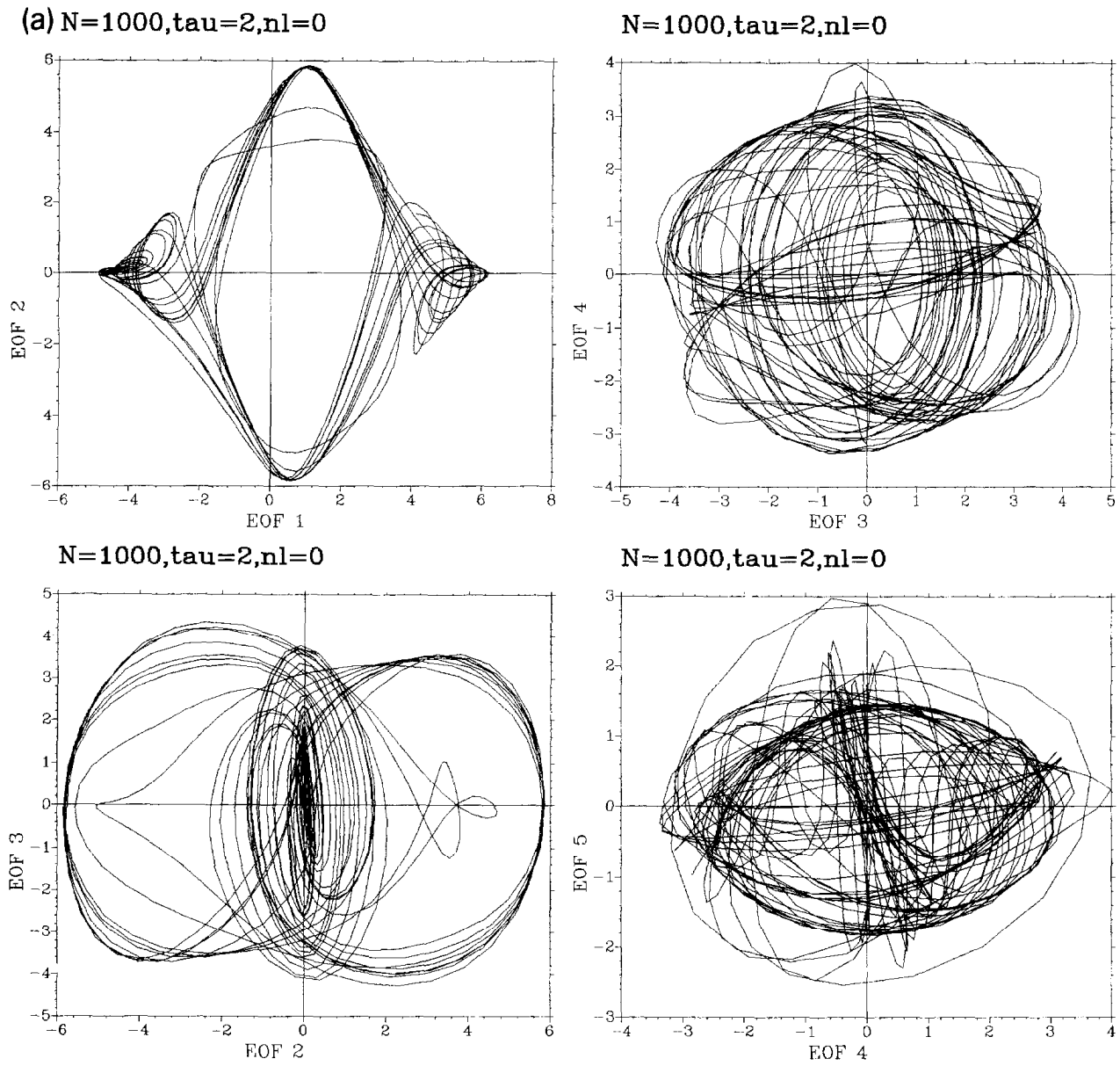

Fig. 3. Phase portraits of the first 5 principal components in the sub-space formed by the directions of the corresponding eigenvectors (a) and the same diagram with random noise in the original time series (b).

maximum amount of information from an available data set which is usually limited in size. Typically, how much information one can extract from a given time series depends on the way it is treated. For the calculation of D_2, the time delay τ in (2) has to be sufficiently large to guarantee roughly linear independence of the delay coordinates. This is widely regarded as necessary, but we consider it as inefficient, because, in this way, the resulting reconstruction is strongly influenced by noise [8,14,9] and much information about the small scale structure contained in the time series is not employed. This is one of the reasons for using the re-embedding procedure. Here we always keep the delay $\tau = 1$ for the first embedding in (2) without considering the sampling rate (see the section 3.3), since the singular value decomposition will automatically produce an orthogonal basis (9), (10) on which the attractor is reconstructed to estimate D_2. In this sense, the

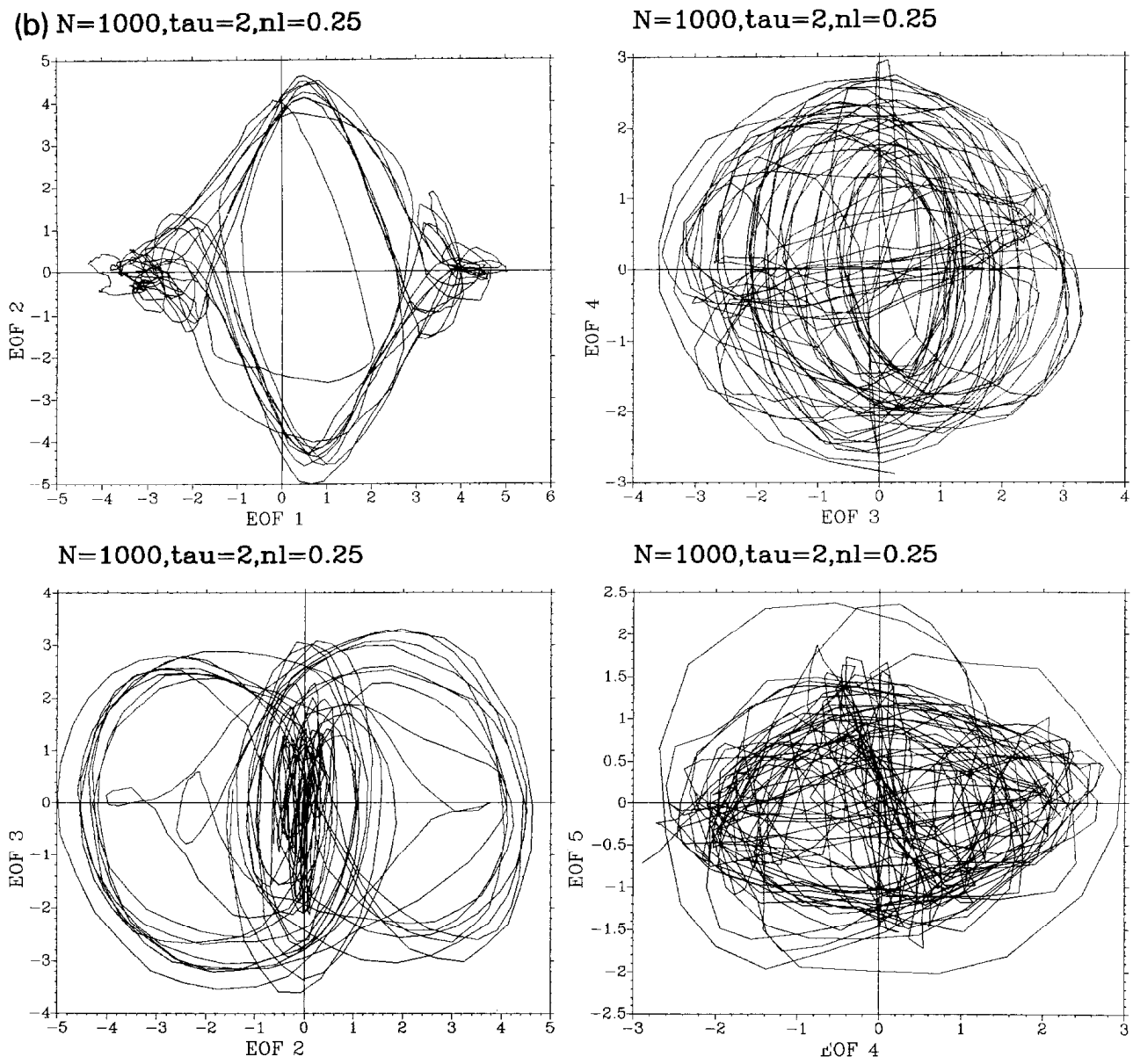

Fig. 3 (cont.).

information contained in the original time series can be more fully used and D_2 may be obtained with relatively short time series compared to the requirement of the original algorithm.

3.1. Noise reduction

We have used the Lorenz model [36] and Mackey–Glass equation plus white noise of various intensity to test the algorithms. The following shows the main results. For the sake of comparison, we have summarized the results for pure white noise (see appendix) in fig. 4 which shows the $D_2(M, r)$–$\ln(r)$ diagram of the original Grassberger–Procaccia algorithm and that of the modified version. For the latter, we used $M = 50$ in (2) and the first 5 PCs for the re-embedding (16). Both algorithms show no saturation of the slope (fig. 4a, 4b).

The Lorenz model is a typical example of a

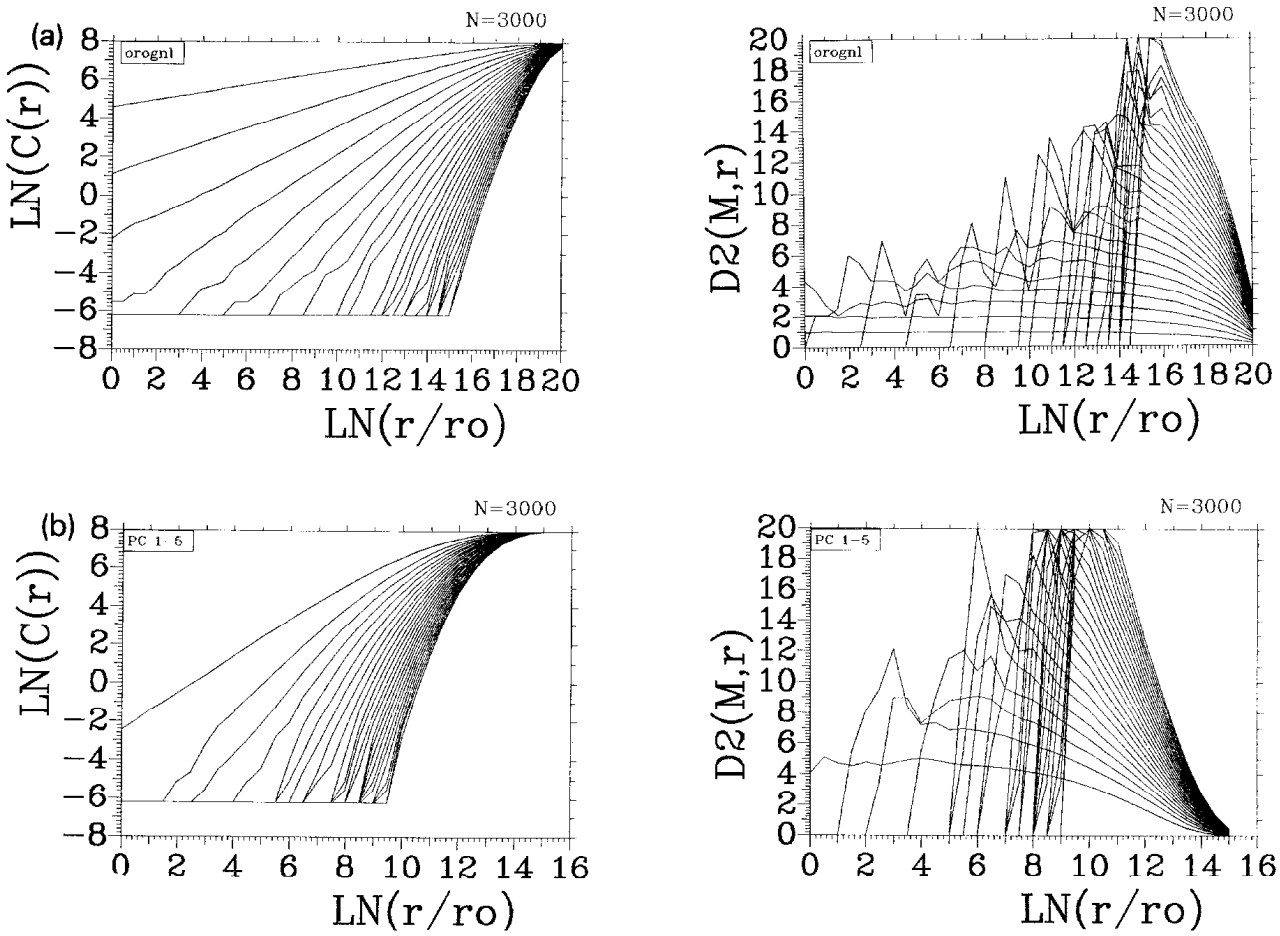

Fig. 4. The $D_2(M, r)$–ln(r) diagrams for white noise: (a) the Grassberger–Procaccia algorithm; (b) the modified version of the Grassberger–Procaccia algorithm.

low dimensional chaotic system. For the purpose of the test, the first variable plus noise (see appendix) was subjected to both the algorithms using $N = 1000$ data points. It turns out that both give a realistic estimation when the noise level is low. However, when the noise level exceeds a certain limit, the modified algorithm gives a clearer scaling on the $D_2(M, r)$–ln(r) diagram than the original one, as is shown in fig. 5 for example. From the phase portraits of the time series (see appendix and fig. A.2) which produces fig. 5b it can be seen that the attractor is severely blurred by the noise. When this noisy time series is subjected to the Grassberger–Procaccia algorithm, the scaling region is no longer clear and there is no clear sign of the slope becoming saturated (see fig. 5a). However, for the same data, the modified Grassberger–Procaccia algorithm (17) still provides a realistic estimate of D_2 with a clear scaling region (fig. 5b), which is independent of the details of the embedding and the number of principal components used, as long as the modes being chosen are above the noise level. This can be understood by comparing figs. 3a and 3b, which show the structure of the phase portraits of the first five principal components: the noisy one is almost the same as that derived from the noise free time series except that, for the noise free case, the trajectories are very smooth, whereas for the noisy data, the trajectories display some small wobbles which are reflected in the $D_2(M, r)$–ln(r) diagram by the divergence at small r.

In order to show that the present method

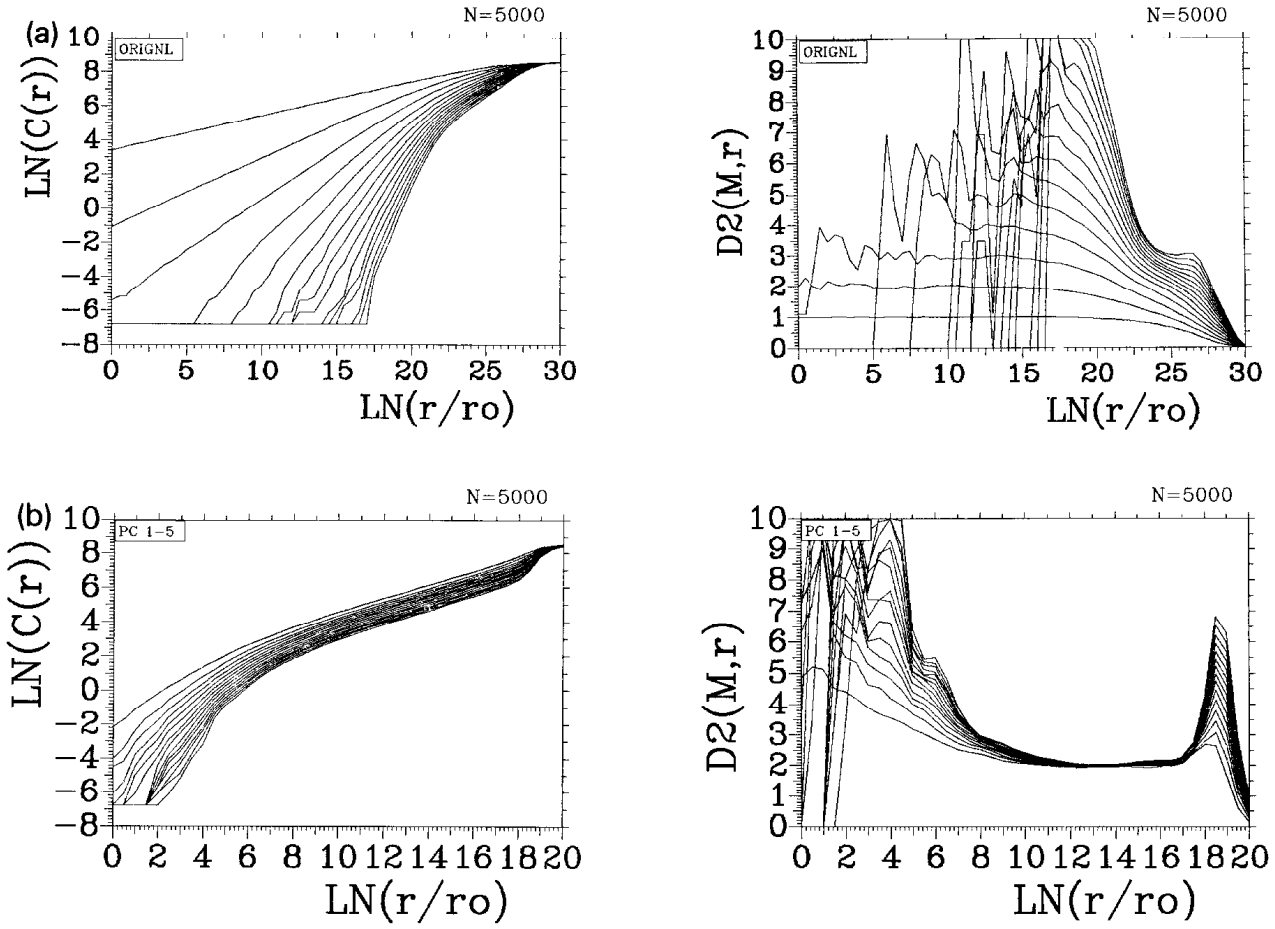

Fig. 5. (a) The same as fig. 1a for the first variable of the Lorenz model plus noise (see appendix). ($M = 1 - 15$, $\tau = 2$.) (b) The same as (a) but by using the re-embedding procedure.

works not only for a low dimensional model like the Lorenz model, but also for systems of higher dimension, we have used the Mackey–Glass equation, whose attractor dimension was designed to be about 5.0 (see appendix). Here $N = 3000$ data points were used, which is similar to typical situations in applications, though it is usually considered as far less than adequate for estimating a dimension about 5. For the noise free case, both the original Grassberger–Procaccia algorithm and the modified version are able to give an estimate of the dimension with a clearly defined scaling region (figures not shown). However, for noisy data, the latter appears to work better than the former (compare figs. 6a and 6b). For the sake of comparison, a similar calculation was carried out using $N =$ 10 000 data points and the results are given in figs. 6c and 6d. Note that the estimated D_2 values for both cases are very close. The only difference is that the scaling region is larger and clearer when N is larger (see next subsection for discussion).

3.2. D_2 from relatively small data sets

There are many arguments regarding the requirement of the minimal number data points N_{min} for estimating the dimension of an attractor of dimension D. L. Smith (1988) [60] gives a theoretical model which states clearly that N_{min} is related with the length of the scaling region (R,

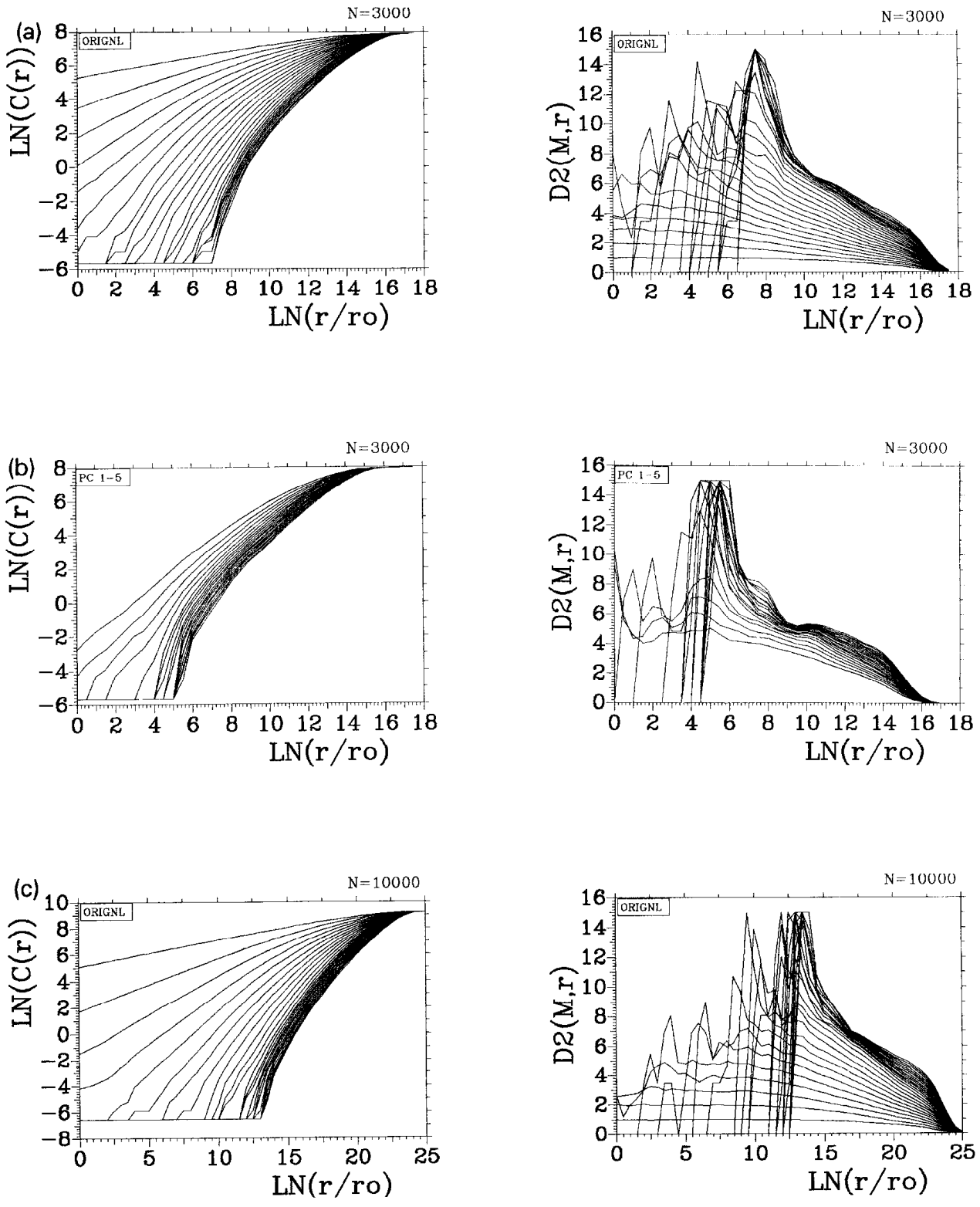

Fig. 6. The $C(M, r)$–$\ln(r)$ and the $D_2(M, r)$–$\ln(r)$ diagrams for the Mackey–Glass equation plus noise derived from the Grassberger–Procaccia algorithm and from the modified version of the algorithm, respectively, for $N = 3000$ (a), (b) and $N = 10\,000$ (c), (d).

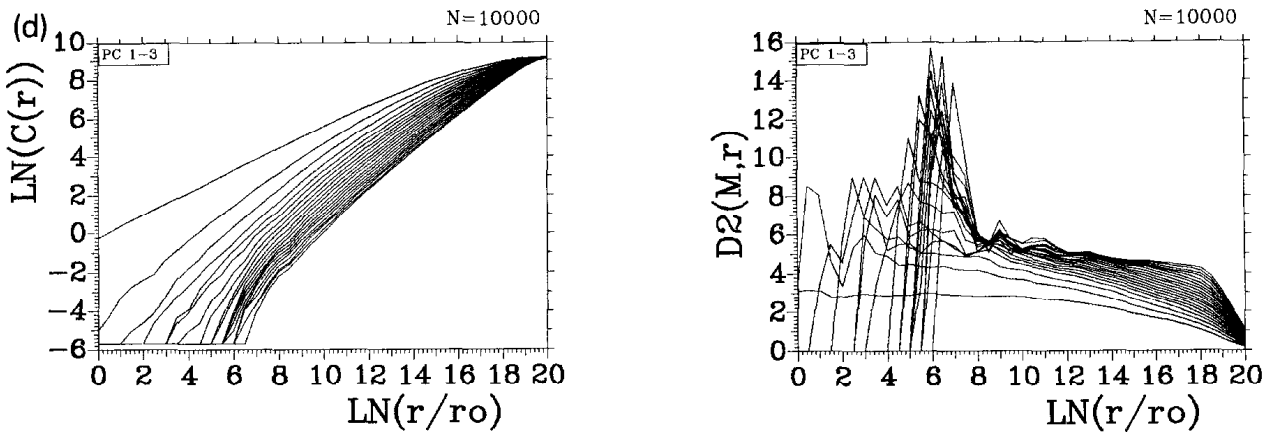

Fig. 6 (cont.).

represented by the base-2 log of the length of the plateau in the $D_2(M, r)$–$\ln(r)$ diagram in his paper [60]) and the required precision (Q). For a scaling length of 4.0 with $Q = 0.95$, according to [60], the minimal number of data points is 42^D. However, this number can be easily reduced to 5.5^D by setting the scaling length $R = 2.0$ and the quality factor $Q = 0.90$. Ruelle [55] suggests $N_{\min} \geq 10^{D/2}$. Nerenberger and Essex [41] given an estimation of N which is similar to Ruelle's. In reality, any attempt to give an absolute limit for the minimal requirement of the number of data points $N_{\min}$ would meet with difficulties, because $N_{\min}$ does not tell anything about the quality of the data with respect to, for example, the precision of measurements, the resolution δ in (1), and the coverage of the whole period of time. For convenience, let T represent the total span of time during the course of an experiment or computer simulation, and $T = N\delta$ where N is the number of data points in (1) and δ is the corresponding sampling interval. Thus, it is clear that N can be increased either by increasing T for a fixed sampling interval δ or by reducing δ for a fixed span of time T. During the course of an experiment within a fixed span of time T, one can easily increase the number of data points N by sampling densely (i.e. small δ in (1)) such that $N > N_{\min}$. In the case of the Lorenz model, we can manage to have a very large number of data points N within a limited length of the trajectory, say, one or two quasi-cycles such that N becomes larger than any estimated $N_{\min}$. However, it is still not adequate to estimate a dimension. In the case of the Lorenz model, for example, how many "quasi-cycles" are *identifiably* covered by the length of the data T appears to be more important than the absolute number of data points N. Moreover, the presence of the noise adds to another important factor that influences the quality of a given data set, especially when the data set is small. For almost noise-free data, one can get a quite reliable estimate of D_2 by both the original Grassberger–Procaccia algorithm and the modified version using a small data set, given that the sampling time is appropriate (i.e. not ill-sampled). As an example, fig. 7 gives the estimate of D_2 with 316 and 10 000 (noise free) data points (for the same sampling time δ), respectively, using both the original Grassberger–Procaccia algorithm and the modified one. In this example, the result did not improve drastically for either algorithms when increasing the number of data points from $10^{2.5}$ to 10^4 (see figs. 7a and 7b for the original Grassberger–Procaccia algorithm, figs. 7c and 7d for the modified version and also [41]), although there are indeed distinct differences in the length of scaling regions and the quality of the estimates which can be seen by the length of the plateau

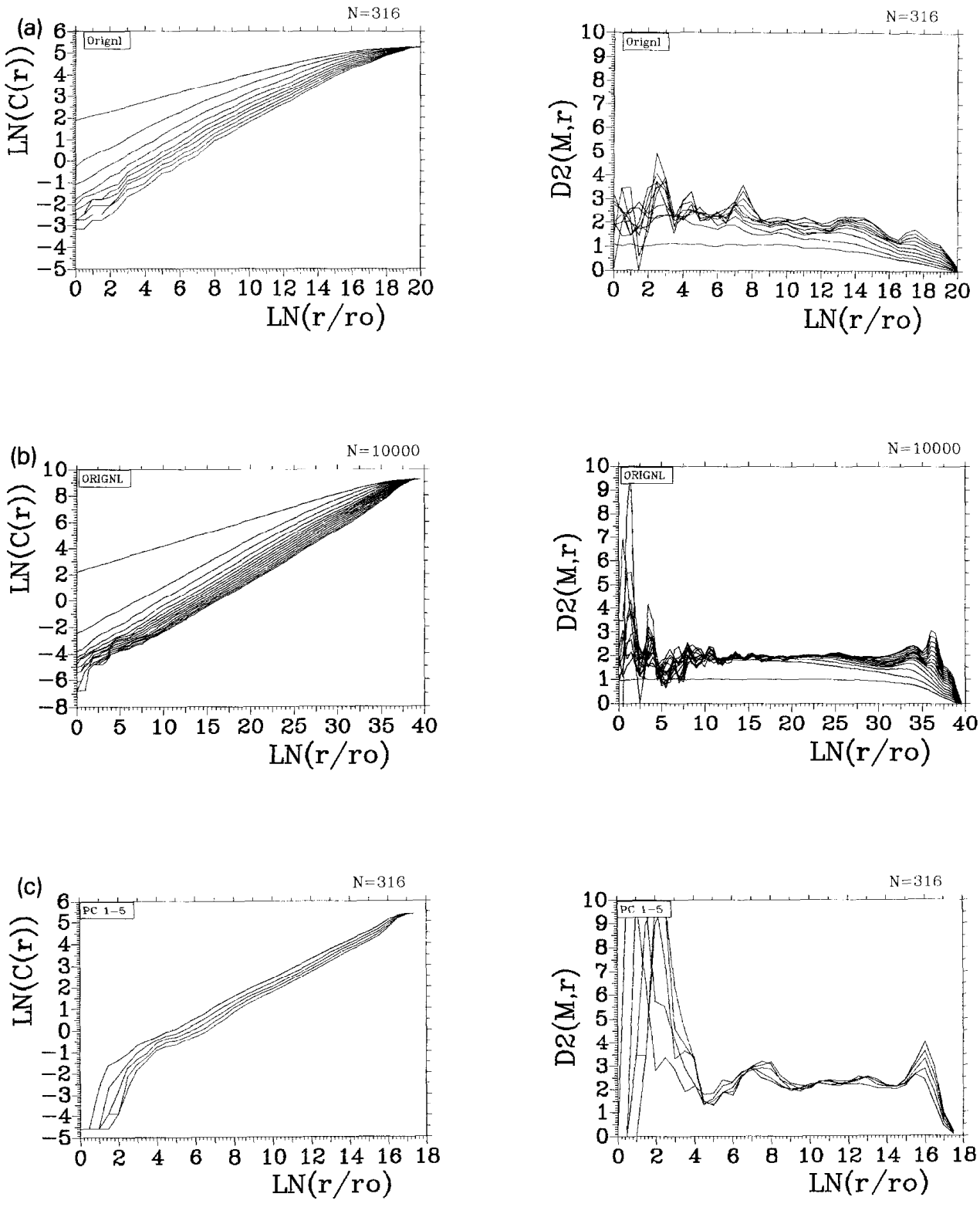

Fig. 7. Comparison of the estimation of D_2 for the Lorenz model with different number of data points (N) using the original Grassberger–Procaccia algorithm (4) and the modified version of the Grassberger–Procaccia algorithm (17): (a) $N = 316$; (b) $N = 10\,000$.

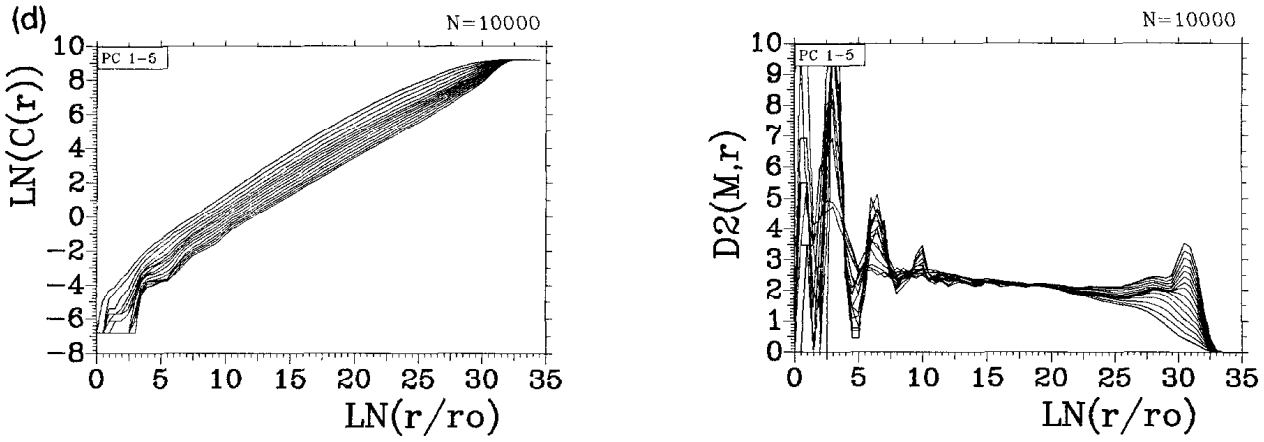

Fig. 7 (cont.).

and the smoothness of the slope curve (see fig. 6 and fig. 7). This indicates that, for fixed sampling rate and precision of measurements, large numbers of data points usually produce better estimates than small ones. On the contrary, it does not help much in the dimension calculation to increase the number of data points N while keeping the total time span T unchanged (figures omitted). Note that the situation would be different for fractal sets, given that the relative precision is the same on all scales. However, this is usually not the case for most experiments or field observations because of the noise associated with instrumental measurements.

According to the above discussion, it is difficult to give an absolute meaning to an estimate of $N_{\min}$. Note that Smith's estimate of $N_{\min} = 42^D$ is, in fact, a particular example of the relation derived as a function of the plateau length (R) and accuracy (Q). Generally speaking, for the same quality of data, large data sets, as discussed above, tend to produce a longer and clearer scaling region on the $D_2(M, r)$–$\ln(r)$ diagram and a better estimate than small data sets, which is roughly in agreement with the relation derived by Smith [60]. Typically, as far as the experimental (observational) data is concerned, the nature of the question becomes different. As we usually have no choice about the length of a time series, we have to try to get as realistic as possible an estimation using the available data by reducing the influence of noise or by making full use of the information contained in the data. As long as a plateau appears on the $D_2(M, r)$–$\ln(r)$ diagram, it is not hopeless even if one has a small data set. Therefore, in the present paper, we did not try to give a theoretical consideration of the minimal requirement of the number of data points ($N_{\min}$). Instead, we emphasize that the appearance of a *significant* plateau on the $D_2(M, r)$–$\ln(r)$ diagram is very necessary for a reliable dimension estimation. It is not surprising that some data sets will never produce such plateaus due to many reasons which could be, for example, the quality and quantity of the data, the effectiveness of the algorithm, or there is no scaling at all. In this case, of course, no conclusion about the dimension can be drawn based on the calculations. Since we cannot increase the data set in many applications, it is important to test the significance of an observed plateau on the $D_2(M, r)$–$\ln(r)$ diagram with the surrogate data as discussed, for example, in [51,61].

By comparing the results of dimension estimation using both the Grassberger–Procaccia algorithm and the modified algorithm, it appears that the latter produces a clearer scaling on the $D_2(M, r)$–$\ln(r)$ diagrams and a better estimate of the D_2 than the former for small data sets while both produce the same estimate when the data set is large (see fig. 7 for a noise free case). This

is particularly the case when the data sets are small and noisy (see figs. 5 and 6).

3.3. Comments and discussions

The effect of noise and the limitation of the data sets available are the main problems in estimating D_2 from experimental data. For a given time sequence of an observed signal, an optimal estimation of D_2 (if it exists) can be achieved when the noise level is minimized and, at the same time, the information contained in the data is fully utilized for reconstructing the attractor. For noise reduction, one can use some kind of filter designed to keep the dynamically relevant signals while reducing the noise level as much as possible. However, the use of linear filters is often problematical [40,57]. Various noise reduction procedures based on nonlinear filters have been proposed in recent years (see, for example, [31] and [57]), which provide a powerful tool for real data analysis. Our approach suggested in section 2.4 benefits from the features of the singular value decomposition, which picks up the main features of the dynamics in the first few principal components where the noise level is normally reduced to minimum. At the same time, it produces automatically several linearly independent variables from the original time-delay embedding (2), in which the time-delay τ is set to the sampling time in order to use all the data points available. The above tests using noisy and relatively short time series did show some advantages in the present algorithm for estimating the D_2. However, the following should be noted:

Firstly, as the basis of the present algorithm is derived directly from the Grassberger–Procaccia algorithm, some of the questions concerning the identification of the deterministic processes from stochastic ones [45,65,61,51] still remain open. We have been concerned only with white noise. The problems associated with some coloured noises may be solved using the procedures such as suggested for example in [65,51,61].

Secondly, our approach may be identified as a modified version of the Grassberger–Procaccia algorithm because it differs from the original one by the re-embedding method. However, this is not intended to replace the original one. It is clear that the modified version is not necessary whenever the Grassberger–Procaccia algorithm works, because the latter is easier in its implementation than the former.

Thirdly, the choice of window for the first embedding is also important for a good estimation. For a given time series, a suitable window (M) in (2) should be chosen such that the dynamics contained in the time series could be best represented by the reconstructed attractor. For example, if we choose M too small, there will be only one significant eigenvalue (which is especially the case when the sampling time is small and the time series is highly auto-correlated). In this case, one can use the first eigenvector as a filter, based on which it is also possible to estimate D_2, given that the embedding dimension for the re-embedding (16) is sufficiently large. On the other hand, M should not be chosen too large, otherwise the first few eigenvectors tend to smooth the time series too much, which could give rise to some other problems (see for example [64,40]). There are no general rules for choosing M when the dimension of the attractor is not known a priori. From our experience, however, it appears to be sufficient to choose M (for a given time series) such that at least 3 dominant eigenvalues are obviously above the noise level. Indeed, one of the advantages of the present method is that the calculation of D_2 is not sensitively dependent on the choice of M for (2) or the delay time τ for the re-embedding (16). That is, the choice for these parameters is not unique. If there exists some scaling invariance about the attractor, the resulting dimension calculation using (17) neither depends on the length of the window (M) in (2) for a certain range of M, nor on the value of the τ in (16) for a certain range of τ, where the range of M or τ depends on the resolution and the length of the

time series. It can also be shown that the result does not depend on the number of principal components (PCs) involved in calculating D_2 (as long as the PCs are all significantly above the noise level), although the number of PCs changes with the length of the window (M) in (2).

Finally, before application of an algorithm to a data set, one should be well aware of the scale range reliably represented by the data, which is determined by the sampling time (resolution in time) and the total length of the time series. As far as the observational data is concerned, the precision of the data (which is related to the instrument) also plays an important role in determining the scale range of the dynamics. A useful guideline can mostly be obtained from power spectral analysis by examining the scaling behaviour of the time series in frequency domain. For example, small scale (high frequency) fluctuations are usually corrupted by noise while extra-low frequency fluctuations are poorly represented due to the limitation of data. Thus, the noise contained in data and the limitation of data points form the major problems in real application. In many cases, it is difficult to give an absolute value for the minimal requirement of data points (see section 3.2 for details). However, if the signals that fall in the scale range can be effectively extracted and, thereby, the noise level reduced to a very low level, then it is possible to estimate the D_2 from a relatively small data set using both the Grassberger–Procaccia algorithm and the modified one. (It is clear that an estimation of any quantity with a certain reliability often requires more data points when the data is corrupted by noise.) This is the philosophy for the introduction of the re-embedding in this paper. As the re-embedding procedure is, in some cases, able to extract the relevant signals and, at the same time, to reduce noise of the observed time series, it is quite natural that the modified version, which is simply the combination of re-embedding and the classical Grassberger–Procaccia algorithm, demonstrates some effectiveness in extracting D_2 for *relatively* small and noisy data sets. However, by "small dataset" is meant *relatively* small, as is indicated above.

4. Application to meteorological time series

Detecting a weather and climate attractor from univariate time series has been of some concern in recent years as it holds the promise of further understanding the climate system. The dynamics of the climate system, governed by partial differential equations, involves an infinite number of free variables. However, the large scale dynamics of the system displays remarkable regularities, i.e. large scale patterns or space–time structures, which are typically recurrent but never exactly cyclic. This could be an indication of the existence of lower dimensional chaotic attractors which are commonly characterized by the correlation dimension (D_2). Many studies suggest different D_2's ranging from 3 to 7, depending on the data being used (see for example [42,17,12,66,27,28,32]). However, there have also been criticisms of these findings [22,25]. Here we would like to point out two important aspects. One is the scale range of the dynamics; for example, the planetary scale circulation might not necessarily have the same static properties as fully developed turbulence on small scales. The other is the scaling region in the Grassberger–Procaccia algorithm, i.e. the flat "plateau" in the $D_2(M, r)$–$\ln(r)$ diagram. Examining the $\ln(C(m, r))$–$\ln(r)$ diagram alone is not enough for determining the scaling region, because linear regression or a least square fit might give rise to a false saturation of D_2, especially when the scaling region being chosen is very narrow and only a few points are involved in the regression or the least square fit. As a result, the estimate may be quite dependent on the region used for regression. Therefore we suggest that it is much more important to display the $D_2(M, r)$–$\ln(r)$ diagram than the $C(M, r)$–$\ln(r)$ diagram for dimension estimation (see sec-

tion 2). In this way, any apparent saturation in the $C(M, r)$–$\ln(r)$ diagram can be under exact examination in the $D_2(M, r)$–$\ln(r)$ diagram. The importance of looking at the slope as a function of length scale was stressed by many authors (for example, see [11,60,62]).

As an example of an application, we have used the single station daily surface pressure (Potsdam) for about 94 years (from 1st January 1893 to 31st December 1986, $N = 31\,040$) to examine the existence of lower dimensional weather attractors. First we apply the original Grassberger algorithm to the time series using a delay of 2 days and a maximum embedding dimension of 30. Fig. 8a shows the plots of $\ln(C(r))$–$\ln(r)$ which seems to have a kind of saturation with increasing embedding. However, the $D_2(M, r)$–$\ln(r)$ diagram (fig. 8b) shows clearly that there is no obvious saturation of the slope with increasing embedding dimension, nor is there any indication of scaling regions in $\ln(r)$. The same holds for different τ values in (2). This should not be surprising, because the daily data include the synoptic scale fluctuations (2 to 7 days) which, due to sampling and measurement errors, increase the apparent noise level of the time series. On the other hand, this daily series of a hundred year length covers a wide spectrum of fluctua-

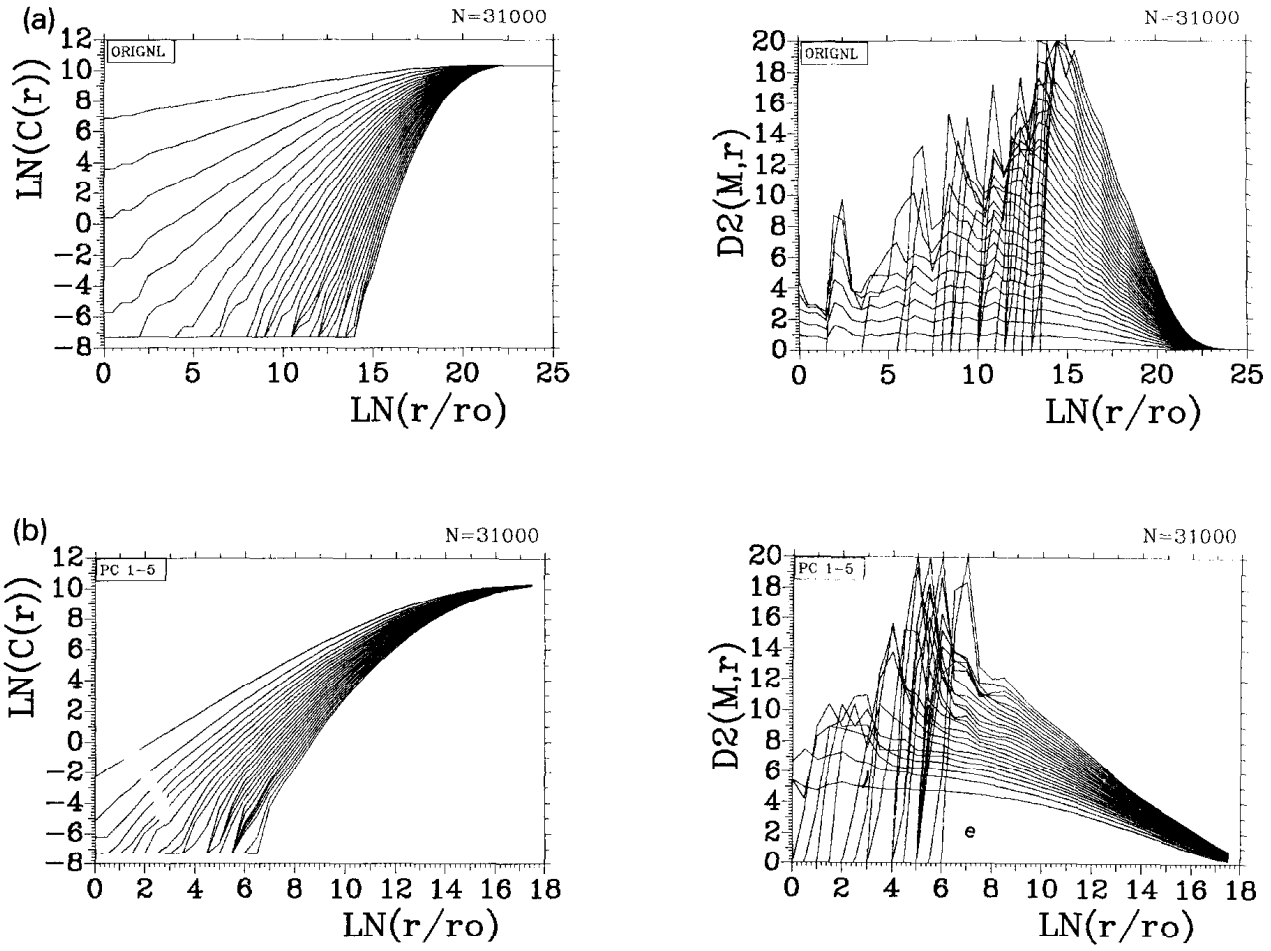

Fig. 8. (a) The $\ln(C(M, r))$–$\ln(r)$ diagram derived by direct application of the Grassberger–Procaccia algorithm to daily pressure data with total number of data points $N = 31040$ (see appendix) and the corresponding $D_2(M, r)$–$\ln(r)$ diagram. (b) The same as in (a) but using the re-embedding procedure with the first 5 principal components.

tions whose dynamics might be governed by different physical processes. What we observe in the time series is a mixture of different scale dynamics plus noise. This shows the difficulties in estimating the dimensionality of a system which is spatially extended and extremely complicated, using only a single observable at a single station (see also (37)).

The same data was subjected to the modified Grassberger–Procaccia algorithm (17), using a window $M = 50$ and a time delay $\tau = 1$ for the first embedding (2). Fig. 9 shows the eigenvalues (a), the first 5 eigenvectors (b) and the associated PCs (c). The decreasing of time scale with the order of the modes can be clearly observed. Considering that the PCs are the projection of the original time series onto the eigenvectors which work, in a sense, like a filter, the first 5 PCs represent the relatively large scale fluctuations (larger than synoptic scale, see fig. 9b). D_2 was calculated with an increasing number of PCs (m in (15)). However, there is no obvious convergence of D_2 even using the first few dominant PCs (see figs. 8c, d for example). A spurious plateau could occur if τ_0 (see (17)) is taken less than the window M. Therefore, based on the above calculations, it is difficult to draw a conclusion about the existence of a weather attractor.

The existence of a weather or climate attractor

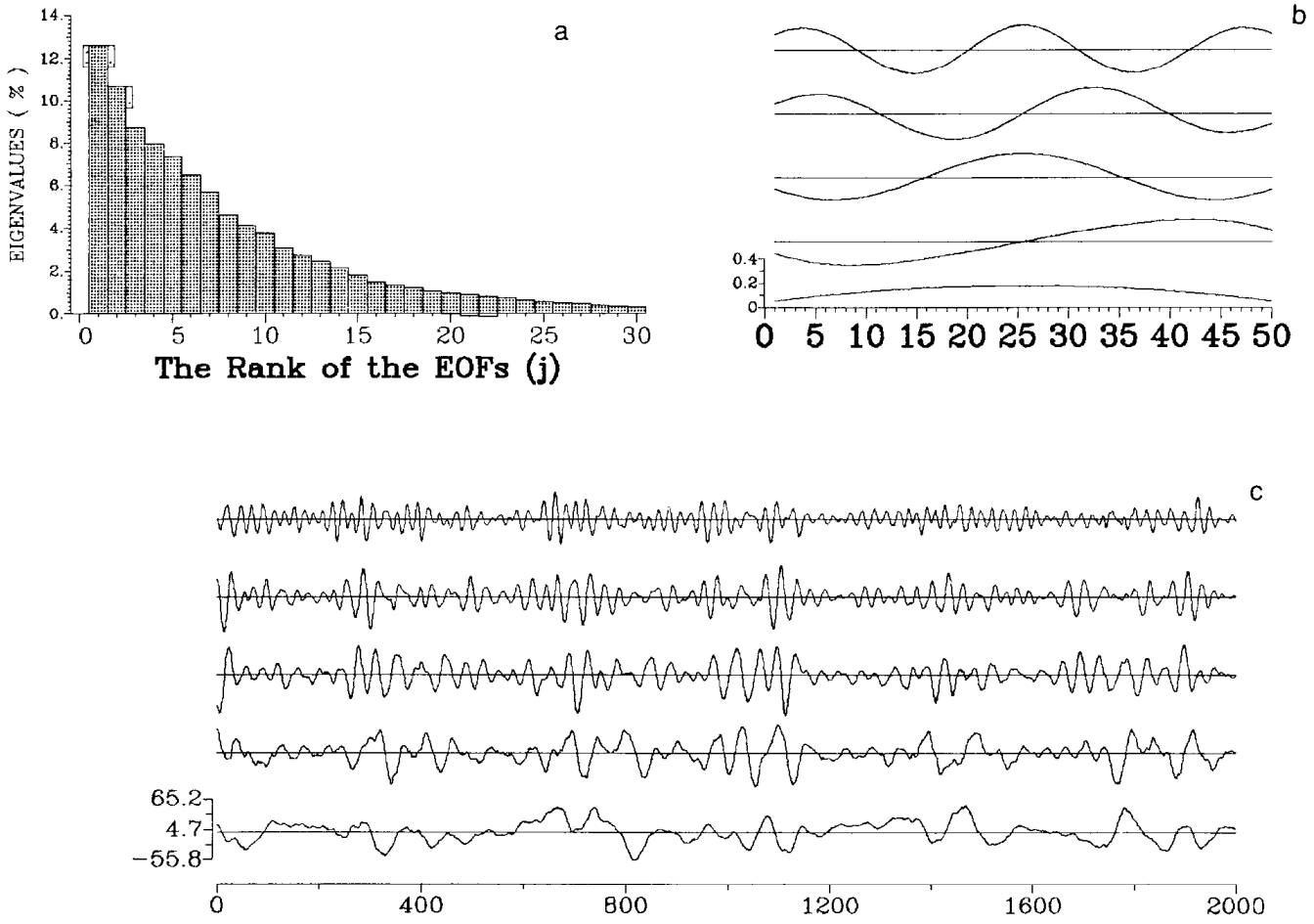

Fig. 9. The eigenvalues (a), the first 5 eigenvectors (b) and the associated principal components (c) derived from the daily pressure series (see appendix). The window used is 50 days.

and of the scaling detected from the data may be a subtle question. One can argue whether it is possible to measure the weather or climate attractor based on the data available, because the quality and quantity of data are still very limited. However, it is hard to believe that the weather patterns observed every day are some kind of stochastic signals. Here it may not be much help to talk about the weather or climate attractor as a whole, because the whole system includes many strongly and weakly coupled subsystems, each of which may follow its own physical processes and develop relatively independently within certain space and time scales. Therefore it would be more useful to specify the attractor of a particular mechanism under consideration, such as the interaction of the synoptic eddies and the basic flow in the mid-latitude circulation, and the quasi-biannual oscillation (QBO) of the tropical atmosphere, etc. As the weather system is spatially extended, it may be necessary to use multichannel measurements and GCMs (general circulation models) to specify a weather attractor that is dynamically relevant to a certain mechanism.

5. Conclusions

The Grassberger–Procaccia algorithm has been modified using a re-embedding procedure described in section 2. It has been shown to be effective for extracting the correlation dimension (D_2) from noisy and relatively short time series by tests based on the Lorenz model and the Mackey–Glass equation plus random noise. Comparing the results between the original and the modified Grassberger–Procaccia algorithm, the latter generally provides a clearer scaling region in the $D_2(M, r)$–$\ln(r)$ diagram and therefore yields a better estimation with some tolerance to the details of the embedding. As only the prominent principal components derived from the singular value decomposition are involved in estimating D_2, this method can be regarded as a natural way of reducing noise. Since all signals contained in the principal components (those above the noise floor) are simultaneously and independently used for re-embedding (16), D_2 can be derived from a *relatively* short time series by using the information contained in the observational data more fully.

An application of the present method to weather observations (daily surface pressure data for about 100 years) have revealed some problems involved in estimating the dimensionality of a possible weather attractor. It appears to be difficult to draw a conclusion of a weather attractor in terms of D_2 using single station measurements. We pose a question of whether one can generalize the weather attractor as a whole without focussing on a specific mechanism or a subsystem (see section 4 for discussion). Further investigation of this may need to include multichannel measurements and the GCMs (general circulation models) output.

Acknowledgements

We wish to thank Drs. M. Ghil, H. Herzel, A. Provencale, Dr. L. Smith for stimulating discussions and valuable comments, Dr. K.-C. Mo for exchanging ideas and the program for SSA, and Dr. P.M. James for reading the manuscript very carefully during the preparation of this paper and giving valuable comments. We also wish to thank the anonymous reviewers and Dr. A.V. Holden whose valuable comments and suggestions have helped improve the paper. R. Wang is indebted to the Foreign Affairs Office of the Freie Universität Berlin (Dr. H. Hartwich, Dr. J. Plähn) for the hospitality, and the support by a cooperation project between the above mentioned University and the Peking University, P.R. China. The present work is supported by a grant in the BMFT project "Analyse, numerische Simulation und Vorhersage natürlicher and anthropogener Klima-Änderungen".

Appendix

1. Lorenz model [36]:

$$\frac{dx}{dt} = \sigma(y - x) ,$$

$$\frac{dy}{dt} = \gamma x - y - xz ,$$

$$\frac{dz}{dt} = -bz + xy ,$$

$$\sigma = 10 , \quad \gamma = 28 , \quad b = \tfrac{8}{3} .$$

The equations were integrated by a fourth order Runge–Kutta routine using a step size 0.004. The values on every 10th step were recorded for use in this paper.

2. Mackey–Glass equation:

$$\dot{X} = \frac{aX_\theta}{1 + X_\theta^c} - bX$$

is a delay differential equation of Mackey and Glass [39], where X_θ is the variable with $X_\theta = X(t - \theta)$. Chaotic attractors are obtained for most parameter values (given $\theta > 16.8$) and the correlation dimension D_2 is approximately pro-

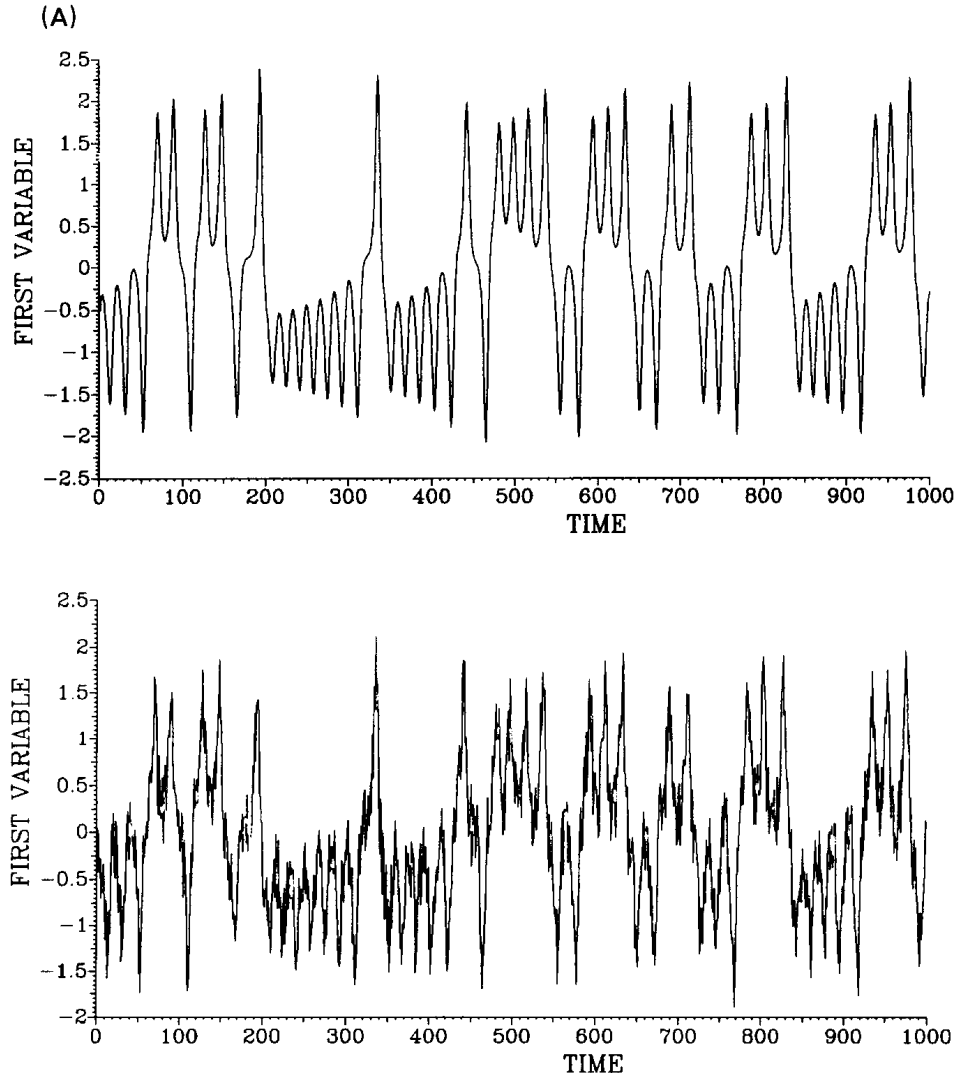

Fig. A.1. First variable of the Lorenz model with (bottom) amd without noise (top).

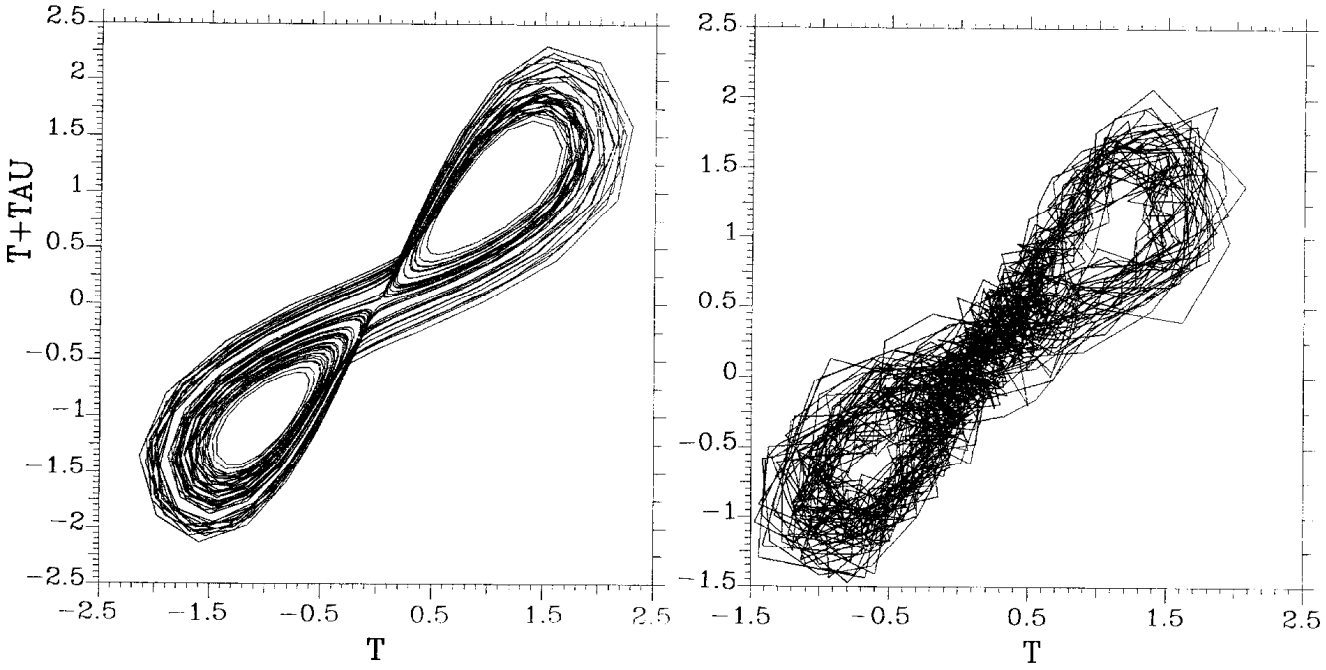

Fig. A.2. The phase portraits corresponding to fig. A.1.

portional to the delay time θ. The parameters a, b and c are fixed at 0.2, 0.1 and 10, respectively. For $\theta = 64$, D_2 is close to 5.0 [7].

3. The random noise is produced by a random number generator and added to the model output. Fig. A.1 shows the time series of the first variable of Lorenz model plus noise (bottom) and that without noise (top). Fig. A.2 shows the corresponding phase portraits in the first two time delay coordinates.

References

[1] A.M. Albano, J. Muench and C. Schwartz, Singular-value decomposition and the Grassberger–Procaccia algorithm, Phys. Rev. A 38 (1988) 3017.

[2] N.B. Abraham, A.M. Albano, B. Das, G. De Guzman, S. Yong, R.S. Gioggia, G.P. Puccioni and J.R. Tredicce, Calculating the dimension of attractors from small data sets, Phys. Lett. A 114 (1986) 217.

[3] H. Atmanspacher, H. Scheingraber and W. Voges, Global scaling properties of a chaotic attractor reconstructed from experimental data, Phys. Rev. A 37 (1988) 1314.

[4] R. Badii, Conservation laws and thermodynamic formalism for dissipative dynamical systems, Rivista del Nuovo Cimento 12 (1989) 3.

[5] R. Badii and A. Politi, Statistical description of chaotic attractors, J. Stat. Phys. 40 (1989) 725.

[6] D.S. Broomhead and G.P. King, Extracting qualitative dynamics from experimental data, Physica D 20 (1986) 217.

[7] R. Cerf and M.L. Ben Maati, Trans-embedding-scaled dynamics, Phys. Lett. A 158 (1991) 119.

[8] J.P. Crutchfield and B.S. McNamara, Equations of motion from a data series, Compl. Syst. 1 (1987) 417.

[9] W. Ebeling, H. Engel-Herbert and H. Herzel, Selbstorganisation in der Zeit (Akademie Verlag, Berlin, 1990).

[10] J.P. Eckmann, Roads to turbulence in dissipative dynamical systems, Rev. Mod. Phys. 53 (1981) 643.

[11] J.P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. 57 (1985) 617; Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems, Physica D 56 (1992) 185.

[12] C. Essex, T. Lookman and M.A.H. Nerenberg, The climate attractor over short time scales, Nature 326 (1987) 64.

[13] J.D. Farmer, Spectral broadening of period-doubling bifurcation sequences, Phys. Rev. Lett. 47 (1981) 179.

[14] J.D. Farmer and J.J. Sidorovich, Predicting chaotic time series, Phys. Rev. Lett. 59 (1987) 845.

[15] J.D. Farmer and J.J. Sidorovich, Optimal shadowing and noise reduction, Physica D 47 (1991) 373.

[16] K. Fraedrich, Estimating the dimensions of weather and climate attractors, J. Atmos. Sci. 43 (1986) 419.
[17] K. Fraedrich, Estimating weather and climate predictability on attractors, J. Atmos. Sci. 44 (1987) 722.
[18] K. Fraedrich, S. Pawson and R. Wang, An EOF analysis of the vertical-time-delay structure of the quasibiannual oscillation, J. Atmos. Sci. (1993) 50, in press.
[19] A.M. Fraser and H.L. Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A 33 (1986) 1134.
[20] M. Ghil, M. Kimoto and J.D. Neelin, Nonlinear Dynamics and predictability in the atmospheric sciences, Rev. Geophys. Suppl. (April, 1991) 46.
[21] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, Physica D 9 (1983) 189.
[22] P. Grassberger, Do climate attractors exist?, Nature 323 (1986) 609.
[23] P. Grassberger, R. Badii and A. Politi, Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors, J. Stat. Phys. 51 (1988) 135.
[24] P. Grassberger, An optimal box-assisted algorithm for fractal dimensions, Phys. Lett. A 148 (1990) 63.
[25] P. Grassberger, T. Schreiber and C. Schaffrath, Nonlinear time series analysis, Preprint (1991).
[26] Bai-lin Hao, ed, Directions in Chaos (World Scientific, Singapore, 1987).
[27] H.W. Henderson and R. Wells, Obtaining attractor dimensions from meteorological time series, Adv. Geophys. 30 (1998) 205.
[28] C.L. Keppenne and C. Nicolis, Global properties and local structure of the weather attractor over Western Europe, J. Atmos. Sci. 46 (1989) 2356.
[29] M. Kimoto, M. Ghil and K.-C. Mo, The 40-day oscillation in the extra-tropical atmosphere as identified by multi-channel singular spectrum analysis, in: The Conference Abstracts of the IUTAM Symposium and Advanced Research Workshop on Interpretation of Time Series from Nonlinear Mechanical Systems, August 1991, University of Warwick, Coventry, England.
[30] G. King, R. Jones and D.S. Broomhead, Phase portraits from a time series: a singular system approach, Nucl. Phys. B (Proc. Suppl.) 2 (1987) 379.
[31] E.J. Kostelich and J.A. Yorke, Physica D 41 (1990) 183.
[32] T.R. Krishna Mohan, J. Subba Rao and R. Ramaswamy, Dimension analysis of climatic data, J. Climate 2 (1989) 217.
[33] J.E. Kutzbach, Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over North America, J. Appl. Meteor. 6 (1967) 791.
[34] W. Liebert and H.G. Schuster, Proper choice of the time delays for the analysis of chaotic time series, Phys. Lett. A 142 (1989) 107.
[35] E.N. Lorenz, Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1: Statistical Forecasting Project, Cambridge, Massachusetts (1956).
[36] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (1963) 130.
[37] E.N. Lorenz, Dimension of weather and climate attractors, Nature 353 (1991) 241.
[38] B. Mandelbrot, Fractals, Form, Chance and Dimension (Freeman, San Francisco, 1977).
[39] M.C. Mackey and L. Glass, Science 197 (1977) 287.
[40] G. Mayer-Kress, ed., Dimensions and Entropies in Chaotic Systems (Springer, Berlin, 1986).
[41] Nerenberger and Essex, Correlation dimension and systematic geometric effects, Phys. Rev. A 42 (1990) 7065.
[42] C. Nicolis and G. Nicolis, Is there a climatic attractor?, Nature 323 (1984) 609.
[43] A.R. Osborne, A.D. Kirwan, Jr., A. Provenzale and L. Bergamasco, A search for chaotic behaviour in large and meoscale motions in the Pacific Ocean, Physica D 23 (1986) 75; Tellus A 41 (1986) 416.
[44] A.R. Osborne, A.D. Kirwan, Jr., A. Provenzale and L. Bergamasco, Fractal drifter trajectories in the Kuroshio extension, Tellus A 41 (1989) 416.
[45] A.R. Osborne and A. Provenzale, Finite correlation dimension for stochastic systems with power-law aspectra, Physica D 20 (1989) 357–381.
[46] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, Geometry from a time series, Phys. Rev. Lett. 47 (1980) 712.
[47] M. Paluš and I. Dvořak, Singular-value decomposition in attractor reconstruction: pitfalls and precautions, Physica D 55 (1992) 221.
[48] K.W. Pettis, T.A. Baily, A.K. Jain and R.C. Dubes, An intrinsic dimensionality estimator from near neighbour information, IEEE Trans. Pattern Anal. Machine Intell. PAMI 1 (1979) 25.
[49] A. Provenzale, A.R. Osborne and R. Soj, Convergence of K_2 entropy for random noises with power law spectra, Physica D 47 (1991) 361–372.
[50] A. Provenzale, A.R. Osborne, A.D. Kirwan and L. Bergamasco, The study of fluid parcel trajectories in large-scale ocean flows, in: Nonlinear Topics in Ocean Physics, A.R. Osborne, ed. (Elsevier, Amsterdam, 1991).
[51] A. Provenzale, L.A. Smith, R. Vio and G. Murante, Distinguishing between low-dimensional dynamics and randomness in measured time series, Physica D 58 (1992) 31.
[52] E.M. Rasmusson, X. Wang and C.F. Ropelewski, J. Marine Syst. 1 (1990) 71.
[53] D. Ruelle and F. Takens, On the nature of turbulence, Commun. Math. Phys. 20 (1971) 167–192; 23 (1971) 343.
[54] D. Ruelle, Strange attractors. Math. Intell. 2 (1980) 126–137.
[55] D. Ruelle, Proc. R. Soc. Lond. A 427 (1990) 241.
[56] T. Sauer, J.A. Yorke and M. Casdagli, Embedology, J. Stat. Phys. 65 (1991) 579.
[57] T. Schreiber and P. Grassberger, A simple noise reduction method for real data, preprint (1991).

[58] H.G. Schuster, Deterministic Chaos (VCH, Weinheim, 1989).
[59] R.L. Somorjai, Methods for estimating the intrinsic dimensionality of high dimensional point sets, Dimensions and Entropies in Chaotic Systems, G. Mayer-Kress, ed. (Springer, Berlin, 1986).
[60] L.A. Smith, Intrinsic limits on dimension calculations, Phys. Lett. A 133 (1988) 283.
[61] L.A. Smith, Identification and prediction of deterministic dynamical systems, Physica D 58 (1992) 50.
[62] H.L. Swinney, Characterization of hydrodynamic strange attractors, Physica D 18 (1986) 448.
[63] F. Takens, Detecting strange attractors in turbulence. Proc. Warwick Symp. 1980, D.A. Rand and L.-S. Young, eds., Lecture notes in Mathematics, vol. 898 (Springer, Berlin, 1981) p. 366.
[64] J. Theiler, Spurious dimension from correlation algorithms applied to limited time-series data, Phys. Rev. A 34 (1986) 2427.
[65] J. Theiler, Some comments on the correlation dimension of $1/f^{\alpha}$ noise, Phys. Lett. A 155 (1991) 480.
[66] A.A. Tsonis and J.B. Elsner, The weather attractor over very short timescales, Nature 333 (1988) 545.
[67] A.A. Tsonis and J.B. Elsner, Chaos, strange attractors, and weather, Bull. Amer. Meteor. Soc. 70 (1989) 14.
[68] R. Vautard and M. Ghil, Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series, Physica D 35 (1989) 395.
[69] R. Wang, Propagating waves in the EOF-decomposition of space-time-delay diagrams, Ann. Geophysicae 9 (Supplement: Abstracts of the EGS XVI General Assembly, Wiesbaden, April 10–15 1991) C588–C599.
[70] H. Whitney, Differentiable Manifolds, Ann. Math. 37 (1936) 645.
[71] Lai-Sang Young, Dimension, entropy and Lyapunov exponents in differentiable dynamical systems, Physica A 124 (1984) 639.