Non-Linear Time Sequence Analysis

Article *in* International Journal of Bifurcation and Chaos · January 1992

DOI: 10.1142/S0218127491000403

CITATIONS
438

READS
112

3 authors, including:

Peter Grassberger
Forschungszentrum Jülich
69 PUBLICATIONS **1,728** CITATIONS

International Journal of Bifurcation and Chaos (IJBC)
Vol.1, No.3 (September 1991)

International Journal of Bifurcation and Chaos, Vol. 1, No. 3 (1991) 521-547

NONLINEAR TIME SEQUENCE ANALYSIS

PETER GRASSBERGER, THOMAS SCHREIBER,* and CARSTEN SCHAFFRATH
Physics Department, University of Wuppertal, D-5600 Wuppertal 1, Gauss-Strasse 20, Germany

Received April 16, 1991; Revised April 26, 1991

We review several aspects of the analysis of time sequences, and concentrate on recent methods using concepts from the theory of nonlinear dynamical systems. In particular, we discuss problems in estimating attractor dimensions, entropies, and Lyapunov exponents, in reducing noise and in forecasting. For completeness and since we want to stress connections to more traditional (mostly spectrum-based) methods, we also give a short review of spectral methods.

1. Introduction

During the last decades, we have seen a dramatic increase in the understanding of chaotic deterministic behavior. This was to a large part due to the wider and easier access to digital computers, which helped to simulate simple model systems.

Apart from applications to very well-controlled and somewhat artificial laboratory experiments (which in some cases even come closer to analog computers than 'experiments'), success in applying the resulting concepts and methods to natural phenomena has been much slower. It is true that there have been claims that low-dimensional deterministic chaos has been found in numerous time sequences, coming from astrophysical, meteorological, biological and even economical sources, but in our view most (if not all) of these claims have to be taken with very much caution.

One reason often seems to be that the algorithms provided by modern nonlinear system analysis have not always been implemented with the greatest care and efficiency. We hope that reviews like the present might help in this respect.

But there also seems to be a deeper problem. All physical systems are corrupted by noise. While in well-controlled laboratory experiments a clear separation between signal and noise can be possible, such a separation is in general not possible in field observations. Indeed, in such cases not only is the actual separation infeasible, it is often not clear whether such a separation should be expected in principle.

An example where no such separation should be possible is velocity measurements in fully-developed turbulence. There, according to Kolmogorov's theory, eddies of all sizes and with all velocity amplitudes should come from the same mechanism. Thus, a typical time sequence will involve few large-scale excursions which might look rather simple, superimposed by eddies which become more and more complex with decreasing amplitude. Attempts to separate these into 'signal' and 'noise' would lead to very wrong conclusions.

In spite of this, most researchers in the past tried to follow the lead set by noiseless computer experiments and by nearly noise-free laboratory experiments, and tried to find scaling laws holding in the limit of infinitely fine resolution (attractor dimensions, Lyapunov exponents, dynamical entropies). A notable exception happened in economics, where Sheinkman and Le Baron [1989] (see also Brock & Dechert [1988, 1991]) observed that even if no such scaling laws can be found, the information obtained by applying an algorithm usually used for estimating dimension can be very useful. More precisely, estimations of phase space correlation integrals helped to show quite unambiguously that there is *some* determinism in data which had passed other tests for randomness. Thus, the emphasis should not be whether the strongest form of determinism (absence of all random noise) is a

*Present address: Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen ‡

good model for the underlying process, but rather whether the weakest form of determinism consistent with other tests can be ruled out.

A similar situation was recently found by Pawelzik [1991] in the cross correlations of signals from single neurons in the visual cortex. In both cases, the interesting structures would not have been found by conventional methods, and since they are not in the form of scaling laws, they would have been overlooked if one had used the new methods only to find scaling laws.

In the following, we will also discuss the scaling laws expected in purely deterministic chaos. But we hope that the above remarks will help in applying the algorithms with a wider perspective in mind.

In addition to this, it might not be useless to stress the need for common sense in nonlinear data analysis. Though many of the algorithms (e.g., the use of delay coordinates) are justifiable by rigorous theorems, the practitioner should never forget that these theorems hold only under conditions which are most likely not guaranteed in his case. In most cases, a few minutes of thought, a plot of the observable against time, or a phase portrait (plot of one observable against another, resp. plot of x_t against $x_{t-\tau}$) will be more revealing than a dimension analysis or an estimation of Lyapunov exponents taking hours of CPU time. Interesting features can often be seen in recurrence plots as suggested in Eckmann & Ruelle [1987] and Babloyantz [1989]. In this context, an interactive general-purpose program for viewing data sets and selected subsets of these data (e.g., Brosa [1990]) might also be very useful.

In the next section, we shall recall traditional (mostly linear and spectrum-based) methods. In Sec. 3, some general concepts relevant for nonlinear time sequence analysis are reviewed, like Takens embedding, Poincaré sections, and Lorenz's "method of analogs" [Lorenz, 1969]. In Sec. 4, tests for nonlinearity based on density correlations in delay-reconstructed phase space ('correlation integrals') are discussed. In particular, we stress that such tests can show regularities not visible from any spectral analysis. More formal aspects of chaotic systems are discussed in Sec. 5, where in particular concepts like Lyapunov exponents, invariant manifolds, topological and metric entropies, Renyi dimensions, partial Renyi dimensions, and the connection with unstable periodic orbits are shortly reviewed. Dimension estimates and the typical pitfalls encountered thereby are discussed in more detail in Sec. 6, while estimates of Lyapunov exponents are treated in Sec. 7. The latter typically involve either global or — and this seems easier in more complex systems, and thus more interesting — local fits to the dynamics. Similar fits can then also be used for forecasting and noise reduction. These are discussed in Secs. 8 and 9.

Two subjects, namely information theory and box-assisted methods of neighbor searching and sorting, are treated in appendices.

An excellent review dealing in more detail with the topics of Secs. 3, 5, 6, and 7 is that by Eckmann and Ruelle [1985]. More phenomenologically oriented is Eubank & Farmer [1990]. Standard references for the more traditional methods discussed in Sec. 2 are Box & Jenkins [1976], Priestley [1981], Hamming [1977], and Press *et al.*, [1988]. Nonlinear time sequence analysis as seen by a statistician is reviewed in Tong [1990].

Throughout the following, we assume that the time sequence consists of measurements $x(t), x(t+\tau), x(t+2\tau), \ldots$ equally spaced in time. Alternatively, we shall denote such a sequence as $\mathbf{x} = (x_0, x_1, x_2, \ldots)$. Gaps in this sequence represent few problems, while observations at unequally spaced times are much harder to analyze. In most cases, x_t will consist of one single number, but sequences of vector-valued data can be treated by very much the same methods.

2. Spectral Methods and Generalizations

2.1. *Generalities*

The first tests which should be done with any time sequence concern obvious periodicities or trends. If a time sequence is not stationary, then it can often be rendered so by differentiating. Unfortunately, this has the effect of enhancing noise and of eliminating possibly interesting structures. A method which avoids this consists in subtracting a linearly growing term (or dividing by a suitable factor), though this has the danger of introducing new and unwanted structures.

Similarly, periodicities due to seasonal effects, for example, can be largely eliminated either by differentiating or by subtracting (or respectively dividing) by suitable terms.

After this has been done, the standard analysis is to look at the spectrum, typically by using a fast Fourier transform (FFT). In this way we can discriminate between discrete, singular continuous, and absolutely continuous spectra, corresponding to periodic, quasiperiodic, and aperiodic motions. We should however warn the reader that this is not always as trivial as it seems, in particular since the finite window of observations produces side bands of the spectral lines (see Press *et al.* [1988]).

2.2 *ARMA models*

A method avoiding the latter problem is Burg's maximum entropy method [Press *et al.*, 1988]. There are a number of ways of motivating and interpreting it. Burg's observation was that it maximizes the entropy

of the spectral distribution (see appendix B), given the observed data as constraints. Alternatively, it can be viewed as a fit to the Fourier transform by an inverse polynomial in $e^{i\omega\tau}$ ("all poles method"), or as a fit by an autoregressive (AR) model. We shall discuss the latter in some detail, as we shall use local versions of it later in Secs. 8 and 9.

In an autoregressive model of order k (called AR(k)) we assume that the process can be modeled as

$$x_n = a_0 + \sum_{i=1}^{k} a_i x_{n-i} + \epsilon_n \tag{1}$$

where $a_1, a_2, \ldots, a_k$ are constants and $\epsilon_1, \epsilon_2, \ldots$, are independent Gaussian random variables with zero mean and with variance σ,

$$\langle \epsilon_i \rangle = 0, \quad \langle \epsilon_i \epsilon_k \rangle = \delta_{ik}\sigma \quad . \tag{2}$$

Notice that Eq. (1) assumes a linear relationship (up to additive noise) between successive measurements, and claims that all nonlinearities have effects indistinguishable from pure noise. One consequence is that each x_n is also a Gaussian random variable, and has, in particular, zero skewness. If this does not hold for the observed sequence, then Eq. (1) should be applied only after a suitable nonlinear transformation has rendered its skewness to zero.

Calling B the backward time shift operator, $Bx_n = x_{n-1}$, we can write Eq. (1) as

$$x_n = a_0 + \phi(B)x_n + \epsilon_n \quad , \tag{3}$$

with $\phi(B) = \sum_{i=1}^{k} a_i B^i$. Multiplying this with $[1 - \phi(B)]^{-1}$, we can write x_n formally as a function of all previous ϵ_i,

$$x_n = b_0 + \psi(B)\epsilon_n \quad , \tag{4}$$

where $\psi(B) = [1 - \phi(B)]^{-1}$ and $b_0 = a_0/(1 - \sum_{i=1}^{k} a_i)$. If $\psi(B)$ were a polynomial of order k (instead of an inverse polynomial as in our case), the last ansatz would be called a moving average model of order k, MA(k). An ARMA(n, m) model finally would be characterized by $\psi(B) = P(B)/Q(B)$, where $P(B)$ and $Q(B)$ are polynomials of order m resp. n. The spectrum of an ARMA model is given by

$$S(\omega) \propto |\psi(e^{i\omega\tau})|^2 \quad . \tag{5}$$

For an acceptable ARMA model, all poles of $\psi(B)$ have to be outside the unit circle. Otherwise, it would not be guaranteed that the effect of noise in the distant past damps out, and that the trajectory does not diverge.

Forecasting a time sequence is particularly simple for AR models, since the best forecasts of x_{n+j} from $x_1, \ldots, x_n$ are just

$$z_m = a_0 + \sum_{i=1}^{k} a_i z_{m-i}, \quad m > n \tag{6}$$

with $z_m = x_m$ for $m \leqslant n$.

Once the order is chosen, fitting the parameters is also particularly easy and elegant for pure AR models. In this case, one essentially minimizes the function

$$L(a_0, \ldots a_k | \mathbf{x}) = \sum_n \left(a_0 + \sum_{i=1}^{k} a_i x_{n-i} - x_n \right)^2 \tag{7}$$

with respect to the a_i, since according to the model this is supposed to be $\sum_n \epsilon_n^2$. The equations $\partial L/\partial a_i = 0$ form a set of linear equations for the a's, with the coefficients being just the autocovariances

$$C_{ij} = \sum_n x_{n-i}\, x_{n-j} \quad . \tag{8}$$

For stationary processes, C_{ij} is a function of $|i - j|$ only (whose Fourier transform is the spectrum). Due to this special structure, solving for the coefficients a_i then becomes particularly easy (see Press *et al.* [1988]).

In general there exists no algorithm for finding the best orders n and m in an ARMA(n, m) model, though a number of criteria and recipes have been given [Box & Jenkins, 1976; Akaike, 1974; Rissanen, 1986]. If one is to take the fit seriously as a model for prediction, one has to be careful not to over-fit with too many degrees of freedom. It seems that Rissanen's minimum description length criterion [Rissanen, 1986] is in this sense preferable over Akaike's criterion [Akaike, 1974].

For not too long time sequences, a good heuristics to find the best orders is to use what some authors call "take-one-out statistics" [Efron, 1982]. In this method (which we discuss here only for AR models) one first modifies Eq. (7) by deleting from the right-hand sum the jth term, and calls the resulting fit $\mathbf{a}^{(j)}$. This is done for every $j \in [1, N]$ (N is the length of the time sequence). Then, one uses $\mathbf{a}^{(j)}$ to "predict" exactly the point not used in its fit,

$$z_j^{(j)} = a_0^{(j)} + \sum_{i=1}^{k} a_i^{(j)} x_{j-i}, \tag{9}$$

and computes the sum of squares

$$\sigma = \sum_j \left(z_j^{(j)} - x \right)^2 \quad . \tag{10}$$

The best choice of k is that which minimizes σ. Overfitting (i.e., too large k) would not further suppress σ since it is a sum of "out-of-sample" errors only.

For very long time sequences, take-one-out statistics becomes infeasible due to the large amount of CPU time needed.

We have discussed take-one-out statistics in some detail since it bears a strong resemblance to the forecasting and noise-reduction methods to be discussed in Secs. 8 and 9.

We should finally point out that ARMA models can be used as linear low-pass filters if the ϵ_n's in Eqs. (1-4) are not noise but the signal to be filtered.

2.3. *Nonlinear and threshold AR models*

A straightforward generalization of AR models are the threshold AR ("TAR") models proposed by Tong [1983]. Here, phase space is divided into disjoint regions $\mathcal{G}_m$, $m = 1, 2, \ldots$, by defining thresholds for some x_{n-i} (notice that the x_{n-i} are just delay coordinates as studied more systematically in later sections), and a different AR model is fitted in each $\mathcal{G}_m$. When applied to standard sequences like yearly sunspot numbers or canadian lynx furs [Tong, 1983], this gives definitely better results than global linear models.

Notice that for a TAR model it is no longer necessary that the polynomials $\phi_m(B)$ for all regions $\mathcal{G}_m$ have only poles outside the unit circle. A blow-up of noise in some region $\mathcal{G}_m$ is acceptable provided it is compensated by damping once the trajectory has left that region. In this way, TAR models are also able to produce noisy limit cycles and chaos.

Another straightforward generalization of Eq. (1) is a quadratic ansatz

$$x_n = a_0 + \sum_{i=1}^{k} a_i x_{n-i} + \sum_{i,j=1}^{k} a_{ij} x_{n-i} x_{n-j} + \epsilon_n \quad . \quad (11)$$

This has been tried in Cox [1977] for the sunspot numbers, with a success comparable to that of Tong's TAR model. Various other nonlinear models of the type $x_n = f(x_{n-1}, \ldots, x_{n-k}, \epsilon_n)$ are discussed in Tong [1990].

Similar models with higher-order polynomials have been tried more recently, and with different motivations [Cremers & Hübler, 1987; Crutchfield & McNamara, 1987]. Provided the system is sufficiently simple (i.e., the order k is small) and the data are sufficiently clean and of high statistics, one can achieve very good results. In particular, one can model in this way even noise-free strange attractors. It seems, however, that in such cases other parametrizations than polynomials — e.g., neural networks [Lapedes & Farber, 1987; Stockbro *et al.*, 1990], radial basis functions [Casdagli, 1989], etc., — will in general be more economic. We will discuss these methods further in Sec. 8.

2.4. *Wavelets*

For completeness, we should mention a new technique for time-dependent (and thus, by Heisenberg's uncertainty relation, finite-resolution) Fourier transform. For this, one has to use wave packets which decay at $t \pm \infty$. Traditional methods have in general not been very systematic in choosing these wave packets. In a number of recent papers [Grossman & Morlet, 1985, Arneodo *et al.*, 1989, Combes *et al.*, 1989] it is claimed that the situation can be drastically improved by choosing proper functions called wavelets. We refer the reader to those references.

3. Elementary Concepts for Nonlinear Deterministic Systems

Let us now take the opposite point of view that noise is not essential, and a better approach might be a completely deterministic model showing chaotic behavior (we should, however, stress again that in our opinion most systems are situated in between these two paradigms).

The basic assumption thus is that the state of the system can be described by a small number of variables (at least when it has reached its attractor — the number of variables during the initial transient period might be large), and that the evolution is given by some differential equation. Since we measure only one coordinate $x(t)$ (and even this only at discrete intervals), it might not be obvious how this assumption should be used.

3.1. *Poincaré sections, phase portraits*

The first observation is that in a deterministic system the evolution over finite time steps is still deterministic. Thus, if $\vec{x}_n$ denotes the state at time $t_n = t + n\tau$, then the map

$$\vec{x}_{n+1} = F(\vec{x}_n) \quad (12)$$

might look awkward analytically, but it will be smooth and well-behaved. An alternative is Poincaré maps, where the time t_{n+1} is defined implicitly by demanding that $\vec{x}_{n+1}$ is on some fixed (hyper-)plane in phase space. The resulting time sequence will in general be awkward for Fourier analysis, but the advantage is that the set $\{\vec{x}_n, n = 0, 1, 2, \ldots\}$, being a cut of the attractor with a plane, has a dimension which is reduced by one unit. Thus, in low-dimensional systems, plotting x_n against x_{n-1} ("phase portraits"; x_n is just one component of $\vec{x}_n$) can show more details when done with Poincaré cuts than with fixed delays. A case where Poincaré maps (called "stroboscopic" maps there) are particularly useful are periodically-forced systems. Considering the force $A \cos \omega t$ as one of the variables, the Poincaré cut consists simply of points at times $t_n = 2\pi n/\omega$.

3.2. *Takens embeddings*

The choice of coordinates for the state vectors $\vec{x}_n$ is largely arbitrary. It was recognized in Takens [1981] and Packard *et al.* [1980] that a very convenient choice are *delay coordinates*, $\vec{x}_n = (x_{n-m+1}, \ldots, x_n)$. The integer m is called the embedding dimension. Assume that the attractor is a differentiable manifold of dimension D. Using Whitney's embedding theorem for manifolds, Takens showed that embeddings with $m \geqslant 2D + 1$ will be faithful generically, and lead to deterministic evolutions of the type of Eq. (12). Thus one cannot exclude special cases where delay coordinates are bad, but they will be atypical and rare.

Taken's theorem does not give any clue to what are the "best" delay coordinates, in particular for noisy data. It is obvious that not every delay τ will give, in practice, the same results. Too small τ will give coordinates which are too strongly correlated, while x_{n-m+1} and x_n will show practically no causal connection for too large τ. Also, in determining m one has to know already the dimension D, which is in general only known *a posteriori*.

In the presence of measurements with finite precision ϵ, a slightly different interpretation of delay coordinates seems advantageous. Assume that we want to characterize the state $\vec{x}(t)$ by measurements of the single coordinate x taken within a time window $[t - T, t]$. If we had infinitely precise measurements, then any $T > 0$ would do, but with errors of size $\epsilon \neq 0$ the choice of T is limited from below and from above. On the one hand, it does not make much sense to use a value of T so large that the first and last measurements are practically uncorrelated, which happens when $\epsilon e^{|\lambda_\pm T|} = \mathcal{O}(1)$ ($\lambda_\pm$ are the largest positive (resp. negative) Lyapunov exponents, see Sec. 5). On the other hand, T should be sufficiently large to cover the dominant frequency of the motion, and to have $|x(t) - x(t - T)| \gg \epsilon$. Independent of the choice of T is the choice of the number m of points in $[t - T, t]$. This has to be sufficiently large that the chance can be neglected that a trajectory leaves the "ϵ-tube" $\mathfrak{M}_{\epsilon,T}(\vec{x}, t) = \{\vec{y}: |y(t') - x(t')| < \epsilon \ \forall t' \in [t - T, t]\}$ between these points. Thus in principle the embedding dimension should be chosen as large as possible. Practical considerations will of course put an upper limit on m.

A number of authors have devised algorithms for estimating lower limits on the embedding dimension m, for either a given T or a given τ [Cenys & Pyrageas, 1988; Liebert *et al.*, 1991]. The most sensitive method seems to be that of Liebert *et al.*, [1991]. There, it was observed that for a good embedding the map $F(\vec{x})$ should be smooth, implying in particular that it should preserve the ranks of the k-nearest neighbors for small k. Details are found in Liebert *et al.*, [1991].

For spatially-extended systems, measurements at one fixed point should in principle give enough information to reconstruct the full attractor (provided, of course the extension is finite). In practice, things are in general more complicated. For instance, in Ciliberto [1987] measurements were performed simultaneously in different points in a Bénard cell with aspect ratio 4. Within the feasible resolution, each sequence looked as if the attractor were low-dimensional, but different sequences gave significantly different dimensions. This shows clearly that the various parts of the Bénard cell were coupled on an amplitude level too small to be seen in the experiment. Similar effective decoupling is also seen sometimes in coupled map lattices (Kaneko [1989], Grassberger [1986a]) and in very large aspect ratio Bénard cells, [Ciliberto & Bigazzi, 1988]. In such cases, multichannel measurements are necessary to characterize the state of the system, and mutual information in the data from different points can show interesting structures [Pawelzik, 1991]. Differences in results of Takens embedding and multichannel measurements for EEG data are discussed in Dvořak [1990].

3.3. *Choice of delay: Mutual information*

In the present subsection we shall discuss carefully a criterion for the optimal delay $\tau = t_{n+1} - t_n$ suggested in Fraser & Swinney [1986]. There, it was claimed that the best choice is given by the first minimum of the mutual information $I(\mathbf{X}_n | \mathbf{X}_{n-1})$ (see Appendix B). The motivation for this is that the information in two successive delay coordinates should be as independent as possible, without making the delay however too long.

If the aim of the embedding is to use only two-dimensional vectors (x_{n-1}, x_n), then this is clearly a good criterion. But it lacks rationale if higher-dimensional embeddings are used. Assume that two delay coordinates $x(t)$ and $x(t - \tau)$ have already been measured and we want to choose the third delay coordinate $x(t - \tau - \tau')$ in an optimal way. The second delay τ' will in general be different from τ. Indeed, in general we expect $\tau' < \tau$. The reason is that the choice of τ implied a compromise between representing well low frequencies (large τ) and high frequencies (small τ). If we would take $\tau' = \tau$, we would improve on low frequencies, but not on high ones. As a compromise, we should thus take $\tau' < \tau$. As an illustration, we show in Figs. 1 and 2 coarse grained mutual informations for the Mackey-Glass delay equation [Mackey & Glass, 1977] with delay $t_{\text{delay}} = 30$. In Fig. 1, the k first delays are fixed at the minima of

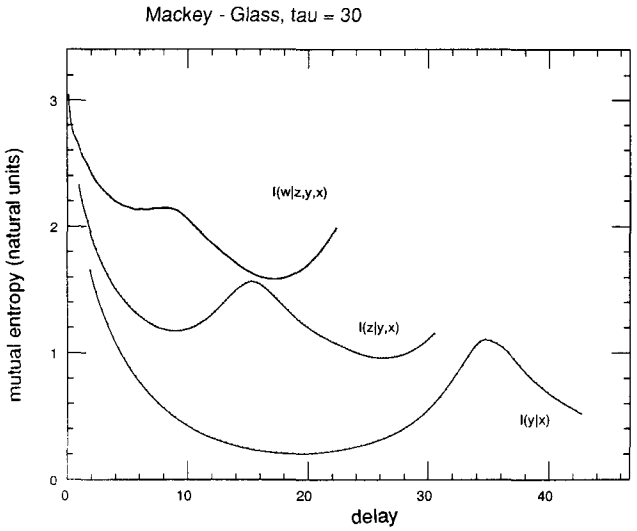

Fig. 1. Mutual entropies for the kth delay coordinate $I(\mathbf{X}_{n-k} | \mathbf{X}_n \ldots, \mathbf{X}_{n-k+1})$ for $k = 1,2,3$, plotted against the delay $\tau_k = t_{n-k+1} - t_{n-k}$, for the Mackey-Glass equation with delay $t_{\text{delay}} = 30$. The entropies (measured in natural units) were estimated using meshes with 32^k boxes and 9×10^6 iterates. The dimension of the attractor for this system lies between 3 and 4. In computing the kth delay, the previous delays were kept fixed at the minima of their respective mutual informations: $\tau_1 = 19.2$ for $k = 2$ and $k = 3$, and $\tau_2 = 8.2$ for $k = 3$.

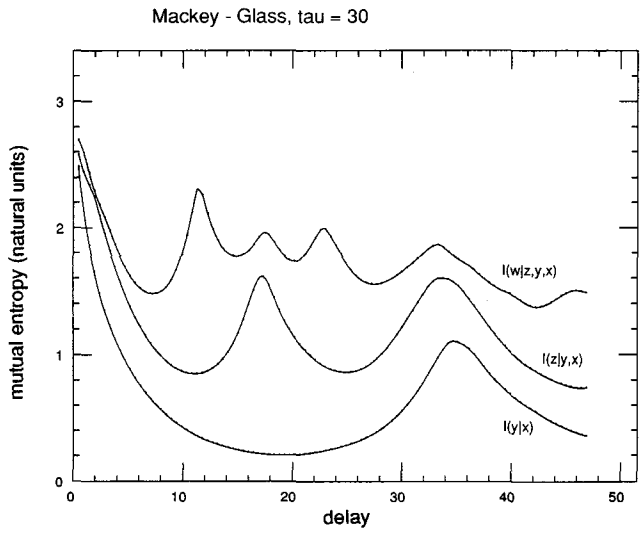

Fig. 2. Same as Fig. 1, but with all delays taken equal.

$I(\mathbf{X}_{n-k} | \mathbf{X}_n, \ldots, \mathbf{X}_{n-k+1})$ when minimizing $I(\mathbf{X}_{n-k-1} | \mathbf{X}_n, \ldots, \mathbf{X}_{n-k})$ in order to get the $(k + 1)$th delay. In Fig. 2, all delays were chosen the same, i.e., the times t_{n-j} are equidistant. In both figures, we see the same qualitative behavior that the minimum decreases with increasing embedding dimension.

It was suggested in Liebert & Schuster [1989] to measure the mutual information between m delay coordinates (for equidistant times), and choose the delay according to its first minimum. In particular, it was proposed to use this delay in subsequent dimension estimates. We should point out that this is not needed, as the correlation "integral" itself is a measure of the mutual (Renyi-) information, as discussed in Secs. 5 and 6.

Before concluding this section, we should stress again that we do not believe that there exists a unique optimal choice of delay. Instead, the best choice depends on the computational tools available (shorter delays obviously imply larger amounts of data) and on the task to be accomplished. We see no reason why delays which are, e.g., best suited for dimension estimates should still be optimal for forecasting. The choice of the best delay will also depend on the effort one is willing to invest. When taking no precautions, short delays (and correspondingly high embedding dimensions) will e.g., enhance the effect of noise on dimension estimates. But the same short delays improve the efficiency of noise reduction algorithms. Thus, if one is willing to apply noise reduction before estimating the dimension, short delays will be better.

3.4. *Singular-value decomposition*

Broomhead and King [1986] proposed to transform from the original embedding coordinates to new coordinates by singular-value decomposition (SVD).

From the set of points $\{\vec{x}_n\}$ in the embedding space we first form the $m \times m$ matrix $\mathbf{C}$ with

$$C_{ik} = \sum_n (\vec{x}_n)_i (\vec{x}_n)_k \tag{13}$$

$$= \sum_n x_{n-i} x_{n-k}, \quad i, k = 0, \ldots, m-1 \ , \tag{14}$$

Notice that $\mathbf{C}$ is just the autocovariance matrix encountered already in Eq. (8). We then make a rotation $\vec{x} \to \vec{\xi} = R\vec{x}$ to bring $\mathbf{C}$ into diagonal form. The eigenvalues μ_i of $\mathbf{C}$ give us the rms extensions of the attractor in the new coordinates, $\mu_i = \langle \xi_i^2 \rangle$.

If the attractor in the original delay coordinates were flat (i.e., confined to some m' dimensional subspace, $m' < m$), then $m - m'$ new coordinates were superfluous, $\vec{\xi} = (\xi_0, \ldots, \xi_{m'-1}, 0, \ldots, 0)$ (here we have assumed that the coordinates are ordered such that $\mu_0 \geqslant \mu_1 \geqslant \ldots$). For a generic strange attractor this will not be exactly true, but some μ_i will be very small. In Broomhead & King [1986] it was proposed to just drop those components for which μ_i is less than some arbitrary threshold, and keep only the m' most important ones.

Provided that m' is sufficiently large, this should still give a faithful reconstruction of phase space. This suggests the following strategy: take first an oversampled time sequence and a rather large embedding dimension m, and reduce this later to the wanted dimension m' by

discarding the coordinates in the least important directions. The advantage is that in this way we have kept only the most relevant information. A bad original choice of m and τ will manifest itself in a too fast or too slow decrease of the eigenvalues μ_i. We see a potential problem with this method only in so far as it mixes measurements at different times. If one wants to use the embedding for forecasting, one wants to put more weight on more recent measurements than on earlier ones. With coordinates obtained by SVD this seems not so easy.

Since the matrix **C** diagonalized by SVD is just the same as that used for AR model spectral estimates, one should not be surprised that the coordinates in SVD become essentially the Fourier coefficients when the embedding dimension is high [King *et al.*, 1987]). This shows also that disregarding small SVD coordinates would not give a good general-purpose noise reduction scheme, as it just corresponds to linear filtering.

A method similar to that of Broomhead and King [1986] was proposed in Landa & Rosenblum [1989, 1991].

3.5. *A simple application: Lorenz' "method of analogs"*

As a simple application of the embedding idea, we consider a method of forecasting suggested by E. Lorenz [1969]. Assume you want to forecast the value x_{n+1}, given all x_j up to $j = n$. Lorenz suggests just to look for the nearest neighbor of $\vec{x}_n$, *called* $\vec{z}$ in the following, and to use $F(\vec{z})$ as a forecast. This method has also been used successfully for the prediction of river run-offs after heavy rainfall in Karlson & Yakowitz [1987] and Yakowitz [1987]. It can be considered as a zeroth-order version of the local prediction methods discussed in Sec. 8.

An improvement of this method was suggested in Pikovsky [1986]. There, the cloud of k nearest neighbors was used, not the nearest neighbor alone. The forecasted value is just the average of their iterates. The advantage is that, in this way, fluctuations are damped out by averaging.

At this time we should stress again that we have to assume what is sometimes called 'strong stationarity.' This demands that all probabilities $P(i_1, \ldots, i_m)$ for the system to find itself in an element of a partition of m-dimensional delay phase space exist and are time-independent. In contrast to this is 'weak stationarity' which concerns only the average value of x_n and the covariance matrix. It is obvious that for a weakly stationary sequence the analog method could fail to work.

Stationarity means that there exists a probability measure invariant under time translation. Indeed, for chaotic deterministic systems, in general an infinity of such invariant measures exists. Most of them (such as delta measures concentrated on unstable periodic orbits) are unstable under a small change either of the initial conditions and/or of the system. If there is only one stable invariant measure, it is called the *natural measure* μ, and the system is called ergodic. We shall in the following assume ergodicity. This guarantees that time averages are equal to space averages. If we have only one time sequence at hand, and if the system is not ergodic, we obviously cannot make any statement about the behavior for arbitrary initial conditions. This would not prevent us, however, from making forecasts for the same time sequence our data are from.

4. Testing for Nonlinearities

Let us assume that the sequence we are interested in has a non-flat Fourier spectrum. We thus know already that it has a certain minimum of structure which can, e.g., be used for forecasting. Let us now discuss methods which can show whether there is indeed more determinism in the sequence than that implied by the spectrum, leaving aside for the moment the question whether it might be even purely deterministic or not. This problem arises very often in economics, where the spectra often are compatible with low-order ARMA models, and other traditional methods are unable to discriminate against them too.

A clear-cut indication that a delay coordinate representation can provide such a discrimination was a study of stock returns in Sheinkman & LeBaron [1989]. Let us denote by $c_m(\vec{x}, \vec{y})$ the density-density correlation in m-dimensional delay-reconstructed phase space. An estimate for the integral

$$C_m(r) = \int\int dxdy\ \Theta(r - \|\vec{x} - \vec{y}\|)\, c_m(\vec{x}, \vec{y}) \quad , \tag{15}$$

called *correlation integral* in Grassberger & Procaccia [1983], is

$$C_m(r) \approx \frac{2}{N(N-1)} \times \{\text{number of pairs } (i, k) \text{ with } \|\vec{x}_i - \vec{x}_k\| < r\} \quad . \tag{16}$$

Let us take in particular the maximum norm,

$$\|\vec{x}_i - \vec{x}_k\| = \max_{0 \leqslant j < m} |x_{i-j} - x_{k-j}| \quad . \tag{17}$$

If there are no correlations between successive x_i, then we have

$$C_m(r) = [C_1(r)]^m \quad . \tag{18}$$

Any sort of correlations, however, would imply $C_m(r)/[C_1(r)]^m > 1$.

When testing this, we should not use the original time sequence x_n since we know already from the spectrum that it has correlations. If, however, the AR model fit which reproduces the spectrum is indeed correct, then the *residues*

$$\epsilon_n = x_n - a_0 - \sum_{i=1}^{k} a_i x_{n-i} \tag{19}$$

should be independent, and the correlation integrals formed with the residues should satisfy the null hypothesis, Eq. (18).

Not only the data studied in Sheinkman & LeBaron [1989] failed to pass this test. Many other economic sequences previously described by AR models also showed statistically significant signals [Brock, 1991]. Notice that this test (called Brock-Dechert-Sheinkman test) is sufficiently simple that exact significance levels can be computed.

Though in economics the BDS test is usually applied to short and noisy sequences whose spectra can be fitted by very low-order AR(k) models (typically $k \sim 1 - 3$), it can for longer time sequences also be used to discriminate against AR models with very rich spectra. We should remind the reader again that AR models fits are indeed often considered as the most convenient and reliable spectrum estimators (see Press *et al.* [1988]).

There exist a number of variants of the BDS test. For instance, one can replace the actual AR residues by Gaussian random numbers, and check whether the correlation integrals for the fake time sequence obtained by inserting them into Eq. (1) differs from the true one. Or one can keep the actual AR residues, but destroy all determinism in their sequence by "shuffling" them (this was originally done in Sheinkman & LeBaron [1989]), and compare the correlation integrals of the resulting x_n with the original one. A different construction of a fake time sequence with the same spectrum for a comparison of correlation integrals was proposed in Theiler [1987a]. He first made a conventional FFT, multiplied $\tilde{x}(\omega)$ with a random frequency-dependent phase, and transformed it back. It seems that all these variants need similar CPU times and give similar results.

5. Dynamical Systems: Theoretical Concepts

In this section we shall continue with our review of theoretical concepts started in Sec. 3. We shall only give a heuristic description, referring to Eckmann & Ruelle [1985] for exact definitions and for further references.

5.1. *Lyapunov exponents and invariant manifolds*

The first such concept is that of Lyapunov exponents which measure the divergence of nearby trajectories [Oseledec, 1968]. Let B_ϵ be an ϵ-ball around $\vec{x}$. If ϵ is infinitesimal, then this ball will be transformed after a time t into an ellipsoid with semi-axes ϵ_i. We assume them to be ordered by magnitude, $\epsilon_1 \geqslant \epsilon_2 \geqslant \ldots$, and define

$$\lambda_i = \lim_{t\to\infty} \left\langle \frac{1}{t} \log \epsilon_i \right\rangle \ . \tag{20}$$

In other words, the λ_i's measure the geometric average growth of vectors in tangent space.

Closely related is the concept of stable and unstable manifolds. The stable manifold $W^s(\vec{x})$ of a point $\vec{x}$ is the set of all points whose iterates approach the iterates of $\vec{x}$ for $t \to +\infty$. Similarly, the unstable manifold $W^u(\vec{x})$ is the set of all points approaching $\vec{x}$ for $t \to -\infty$, i.e., backward in time. For large t, the above ellipsoid will have its growing axes tangential to W^u, while the ellipsoid would be along W^s if we would iterate backward. Chaotic systems are easiest to treat mathematically if they are *hyperbolic*, i.e., if W^s and W^u are everywhere transverse to each other, and span everywhere the space transverse to the direction of the flow $d\vec{x}/dt$ [Eckmann & Ruelle, 1985]. Unfortunately, most realistic systems seem not to be of this type.

In principle, λ_i can still depend on the center $\vec{x}$ of the ϵ-ball. For ergodic systems, the points where $\lambda_i(\vec{x})$ differs from the average will however be of measure zero. Still, the spread of possible values of λ_i will influence the fluctuations of estimates based on randomly-chosen trajectory pieces of finite length.

A major advantage of Lyapunov exponents is that they are invariant under a smooth change of coordinates. Thus they are also independent — at least in the limit of infinite precision and in generic cases — of the chosen embedding. They share this with attractor dimensions and entropies, information theoretic concepts to be discussed now.

5.2. *Entropies and dimensions*

Let us consider again an ϵ-tube $\mathfrak{M}_{\epsilon,T}(\vec{x}, t)$ of length T around a reference trajectory $\vec{x}(t)$, and let us call

$$P_{\epsilon,T}(\vec{x}, t) = \int_{\mathfrak{M}} d\mu \tag{21}$$

its weight according to the natural invariant measure μ. Thus, $P_{\epsilon,T}(\vec{x}, t)$ is the probability that in the stationary

state a randomly-picked trajectory $\vec{y}(t')$ will have a distance $< \epsilon$ from $\vec{x}(t')$ for all $t - T < t' < t$. According to Appendix B,

$$\left\langle \log \frac{1}{P_{\epsilon,T}} \right\rangle = -\int d\mu(x) \log P_{\epsilon,T}(\vec{x}, 0) \qquad (22)$$

is the average information needed for the description of a trajectory, if we want it to stay with an ϵ-tube of length T around a typical reference trajectory.

It is heuristically very plausible that this should have a term $\propto T$, and another term $\propto 1/\log \epsilon$. We thus make the ansatz

$$\left\langle \log \frac{1}{P_{\epsilon,T}} \right\rangle \approx hT - D \log \epsilon + \text{const}, \quad T \to \infty, \quad \epsilon \to 0 \ . \qquad (23)$$

h is called the Kolmogorov-Sinai metric entropy, while D is the information dimension of the attractor.

If we would introduce "symbolic dynamics" by partitioning phase space, we could compute the Shannon entropy of the resulting symbol sequence from Eq. (65) of Appendix B. It would indeed agree with the KS entropy if the partition is *generating* (i.e., it is sufficiently fine not to confuse different trajectories), and if τ is chosen as time unit.

For an attractor, we have *Pesin's theorem* [Pesin, 1977; Eckmann & Ruelle, 1985] saying that h is equal to the sum of positive Lyapunov exponents,

$$h = \sum_{i:\lambda_i > 0} \lambda_i \ . \qquad (24)$$

For generic attractors, the natural invariant measure will be continuous along the unstable manifold W^u of any point, while it will in general be Cantor-like along the stable manifold. Both manifolds can be further split into submanifolds, each of which is characterized by one of the Lyapunov exponents (up to possible degeneracies in the Lyapunov spectrum, see Eckmann & Ruelle [1985]). More precisely, for $\pm\lambda_i > 0$, $W^{(i)}(\vec{x})$ is the set of all points $\vec{y}$ with $\|\vec{x}(t) - \vec{y}(t)\| \leqslant e^{-|\lambda_i t|}$ for $t \to \mp\infty$. We can now define *partial dimensions* $D^{(i)}$ by $D^{(1)} = \dim W^{(1)}$ and

$$D^{(i)} = \dim W^{(i)} - \dim W^{(i-1)}, \quad i > 1 \ . \qquad (25)$$

Obviously, for a generic attractor we have $D^{(i)} = 1$ for all directions with $\lambda_i \geqslant 0$, while the $D^{(i)}$ for directions with $\lambda_i < 0$ will be between 0 and 1.

A generalization of Pesin's equality, holding also for repellers where $D^{(i)} < 1$ along some unstable directions [Kantz & Grassberger, 1985], was given in Ledrappier & Young [1985]:

$$h = \sum_{i:\lambda_i > 0} \lambda_i D^{(i)} = -\sum_{i:\lambda_i < 0} \lambda_i D^{(i)} \ . \qquad (26)$$

Under some assumptions which seem to be true typically, one obtains from this the Kaplan-Yorke relation

$$D = j + \frac{\sum_{i=1}^{j} \lambda_i}{|\lambda_{j+1}|} \ , \qquad (27)$$

where j is the smallest integer with $\sum_{i=1}^{j} \lambda_i > 0$.

5.3. *Generalized entropies and dimensions, multifractals*

In the above, all entropies were Shannon entropies. We can also try to build up an analogous formalism for Renyi entropies. For this, we replace the average over logarithms (i.e., the geometric average over probabilities) in Eq. (22) by an average over powers, and define generalized KS entropies h_q and dimensions D_q via a *generalized correlation integral*:

$$C^{(q)}(\epsilon, T) \equiv \langle [P_{\epsilon,T}]^{q-1} \rangle \approx \text{const.}\ e^{(1-q)h_q T} \epsilon^{(q-1)D_q}, \quad T \to \infty, \ \epsilon \to 0 \ . \qquad (28)$$

In the limit $q \to 1$ we recover the Shannon case. As we shall see in the next section, the case $q = 2$ is by far the easiest to evaluate from a time sequence, since there the generalized correlation integral coincides with the correlation integral studied in Sec. 4. Thus, often only the "correlation dimension" D_2 and the entropy h_2 are estimated, in the hope that $D \approx D_2$ and $h \approx h_2$. We should, however, point out that D_q and h_q are both monotonically decreasing, whence D_2 can in principle seriously underestimate D.

If D_q does really depend on q, we call the attractor (or rather its natural measure) a multifractal [Paladin & Vulpiani, 1987]. To describe multifractals, there exists a beautiful formalism formally analogous to that of thermodynamics. In particular, one can obtain from q and $\tau(q) \equiv (q-1)D_q$ an alternative description of the natural measure by a Legendre transform

$$\alpha = d\tau/dq, \quad f(\alpha) = \alpha q - \tau \ . \qquad (29)$$

Here, $f(\alpha)$ is the Hausdorff dimension of the set of points with pointwise dimension $D(\vec{x}) = \alpha$. The *pointwise dimension* of a measure at a point $\vec{x}$ is defined as

$$D(\vec{x}) = \lim_{\epsilon \to 0} \frac{\log P_{\epsilon,T}(\vec{x})}{\log \epsilon} \ . \qquad (30)$$

The analogy with thermodynamics includes even the possibility of "phase transitions," where, however, the above interpretation of $f(\alpha)$ can break down, see Politi *et al.*, [1988].

In the last years, a large number of papers on multifractals have appeared, and many attempts have been made to measure $f(\alpha)$ of experimental time series. We should, however, point out that claims that $f(\alpha)$ determines an attractor "completely" are misleading. On the other hand it is very hard to obtain reliably more than a rough shape which could have been guessed already, e.g., from D_0, $D_1 \equiv D$, and D_2. The most important aspect of multifractal analyses for experimental sequences seems to be that in this way we can estimate intermittency effects, and can in particular test the assumption that D_2 is a good measure of the information dimension D.

5.4. *Symbolic dynamics*

We have already mentioned the possibility of coding a trajectory by a discrete "alphabet." It is supposed that such codings which in particular preserve the entropy are always possible in generic cases, but they are often not easy to find. Exceptions are, e.g., the famous Lorenz equations [Sparrow, 1982] or the logistic map $x \to a - x^2$, where the obvious binary left/right codings are sufficient.

Experiments or field observations where good symbolic dynamics seemed obvious are described in Honerkamp [1983], Argoul *et al.* [1989], and Hennequin [1991]. There, approximate grammars for the symbol sequences are also given. The results of Honerkamp [1983] seem particularly interesting, since he shows that even a system as complex as the human heart can — under certain conditions, namely if it is in a state of Wenckebach oscillations — have a strange attractor with well understood symbolic dynamics.

In systems without such obvious structure, it is often not easy to find good partitions. Examples where such partitions were found nonetheless are the Hénon map [Grassberger & Kantz, 1985; D'Alessandro *et al.*, 1990] and the Duffing equation [Giovannini & Politi, 1991].

In cases where a detailed study of symbolic dynamics is possible, it can of course provide an extremely sensitive test of models. Also, a number of different definitions of complexities of time sequences have been proposed [Grassberger, 1989; Badii, 1989; D'Alessandro & Politi, 1989; Crutchfield & Young, 1989] which all are tied to symbolic dynamics. Unfortunately, at present all this seems applicable only in relatively simple cases.

5.5. *Unstable periodic orbits*

It was observed already quite early that unstable periodic orbits are dense on typical attractors and that a formalism based on them can be mathematically very convenient. In particular, the natural measure can be approximated by a sum over delta functions concentrated on periodic orbits [Grebogi *et al.*, 1988]. If one uses a "grand canonical" formalism by averaging over different periods (called zeta function formalism due to its analogy with the Riemann ζ function), the formalism becomes particularly elegant, see Bowen & Ruelle [1975]. In this way not only attractor invariants like (generalized) dimensions, entropies and Lyapunov exponents can be obtained, but also autocorrelation functions (see Ruelle [1986, 1986a], Baladi *et al.*, [1989], and Christiansen *et al.* [1990]).

The observation that unstable periodic orbits could also be easy to find in numerical and experimental studies, and that they thus could be numerically useful is much more recent [Auerbach *et al.*, 1987; Cvitanovic, 1988; Cvitanovic *et al.* 1988]. In particular, in Cvitanovic *et al.* [1988], it has been shown that periodic orbits can also be useful for finding good symbolic dynamics. This is easily understood once one realizes that no two different periodic orbits with the same encoding can exist.

In strictly hyperbolic systems (as, e.g., in most repellers, see Cvitanovic [1988]. Gunaratne *et al.* [1988], and Grassberger [1990a]), the resulting expansions can be made to converge extremely fast, see Artuso *et al.* [1990], and Christiansen *et al.* [1990]. This is related to the fact that in such systems the grammar of the symbolic dynamics is usually simple. But nonhyperbolic systems like the Hénon map $(x, y) \to (1.4 + 0.3y - x^2, x)$ [Hénon, 1976] seem experimentally very common. There it was shown that analyses based on periodic orbits are less efficient than either straightforward use of randomly-chosen chaotic orbits [Grassberger *et al.*, 1989], or detailed use of stable and invariant manifolds [D'Alessandro *et al.*, 1990] — even though periodic orbits in the Hénon map can be found very easily by the method given in Biham & Wenzel [1989].

As regards real-world applications, periodic orbits have been extracted in simple systems [Lathrop & Kostelich, 1989; Flepp *et al.*, 1991], and in these examples symbolic dynamics have also been deduced from them. Yet it has still to be shown that reliable extraction and organization of periodic orbits is feasible in more complex situations. Indeed, the results of Lathrop & Kostelich [1989] are somewhat strange as in this reference (see their Table 1) exactly one periodic orbit was found for each period, up to period 8. This suggests that either

the entropy in this system was extremely low (the number of periodic orbits per period p should increase like $p^{h_{\text{top}}}$, with $h_{\text{top}} \geqslant h$), or that each periodic "orbit" of Lathrop & Kostelich [1989] is indeed a cluster of orbits.

6. Phase Space Correlations and Dimension Estimates

Before discussing recent papers in this field, we should point out that it is essentially an application of density estimates. This is an old subject with a considerable literature predating the modern interest in nonlinear systems. A good account of work not primarily concerned with dynamical systems is found in Silverman [1986].

6.1. *Correlation dimension*

As we have already pointed out, the easiest generalized dimension to estimate is D_2 since there we have just to count pairs which remain closer than a distance r during a time interval T (resp. during m successive delay coordinates). Together with D_2, one obtains of course also an estimate of h_2.

Concerning practical implementations we have the following remarks (see also Theiler [1990a]):

(1) As most of the effort in naive implementations is spent on searching for close neighbors, CPU times can be cut by several orders of magnitude by fast neighbor search algorithms such as described in Appendix A or in Grassberger [1990]. (Similar algorithms are described in Theiler [1987], and Bingham & Kot [1989]). When this is done, dimension analyses of sequences with $\approx 10^6$ points (which take *all* close pairs into account!) are done easily on work stations.

(2) As stressed earlier, we do not advise to look *only* at scaling behavior. Even if no scaling is found, qualitative features of the correlation integral might be interesting.

(3) Comparing with white noise (as in Nicolis & Nicolis [1984], Essex *et al.* [1987]) can be very misleading unless the observed spectra are indeed flat. If they are not, significance tests should be done which take the spectra into account as discussed in Sec. 4, see, e.g., Grassberger [1986].

(4) In the sum over pairs $(\vec{x}_i, \vec{x}_k)$, diagonal terms $i = k$ should of course not be included. Where this seems to have been done [Cohen & Procaccia, 1985; Fraedrich, 1986; Holzfuss & Mayer-Kress, 1986; Layne *et al.*, 1986], the results should be taken with great care.

(5) But leaving out the diagonal terms is in general not sufficient. Remember that the correlation sum should reflect the clustering of points in phase space due to purely geometric effects, not due to dynamical correlations. Thus, all pairs should be discarded whose distance in time is not much larger than the correlation time t_{correl}, as pointed out in Theiler [1986]. It seems that a number of spuriously small observed attractor dimensions are obsolete due to a neglect of this. A popular alternative to excluding just pairs with $|t_i - t_k| < t_{\text{correl}}$, is to take only a subset of the time sequence where the delay $|t_n - t_{n-1}|$ between successive points is $> t_{\text{correl}}$. While this does indeed eliminate the above problem, it has the drawback of reducing the statistics substantially, and is thus *not recommended* except when the time sequence is extremely long.

(6) The time sequence has of course to be long enough to sample the attractor reasonably, and the system has to be stationary. This remark is indeed related to the preceding remark: if the attractor is not yet sampled enough, *all* points in the time sequence are dynamically correlated. The neglect of this obvious requirement is probably the most common reason why many authors found low-dimensional chaos in real-world phenomena. The small dimensions found in these analyses are dimensions of individual trajectories, but not of invariant measures.

In particular, it is well known that Brownian diffusion leads to paths with dimension $D = 2$ (in spaces with $\geqslant 2$ dimensions, of course), and Levy flights can have any dimension < 2 (see Mandelbrot [1985]). Thus observations of low dimensions in real-life diffusion [Osborne *et al.*, 1986] and in simulations [Osborne & Provenzale, 1989] should not surprise. Such observations should, however, not be mistaken as indications of small *attractor* dimensions (indeed, for a free particle undergoing diffusion, there does not exist a normalizable invariant measure, and if one makes the invariant measure normalizable by, e.g., adding a potential, its dimension becomes infinite). In contrast to claims in Osborne & Provenzale [1989], they do not contradict the fact that stochastic noise leads to infinite attractor dimensions.

The same holds for the claim of Provenzale *et al.* [1991] that Levy flights can lead to small entropies h_2. As illustrated by Pesin's theorem, the KS entropy (and h_2) measures the divergence of *nearby* trajectories. For a Brownian or Levy flight, however, one does not have enough nearby pairs, due to the lack of stationarity. The entropy in

Provenzale *et al.* [1991] was thus measured from the divergence of distant pairs. While the power law divergence in a Brownian or Levy flight is faster than exponential for small times (leading thus to the correct result $h = \infty$), an estimate based on a power law divergence at large times has to give spuriously $h = 0$.

We hope that the discussion triggered by Osborne *et al.* [1986], Osborne & Provenzale, and Provenzale *et al.* [1991] will help future authors to be more careful as regards stationarity.

(7) Small measured correlation dimensions can be misleading in systems with strong intermittency, or with two different dynamics acting at different times. Take for instance wind speed measurements near a coast with strong winds in the afternoon and weak winds at night and in the morning. Most pairs (i, k) with $(\vec{v}_i \approx \vec{v}_k)$ will come from night or morning measurements, representing a phase space region which on the global "attractor" scale is essentially a single point. A small measured dimension would then result trivially. Such a measurement was indeed done in Tsonis & Elsner [1988]. Though the night data were taken out in this analysis, the morning and evening data were still retained. They might easily account for the observed effect.

Notice that the correlation dimension is particularly sensitive to this problem. If one suspects it to be present in one's data, it is essential to estimate the information dimension D_1 instead, since it is less affected.

A cheaper way to suppress spurious effects due to very inhomogeneous densities in phase space is to transform the values x_n first so that they are uniformly distributed [Politi *et al.*, 1990]. Indeed, in Politi *et al.* [1990] this was applied before measuring the information dimension in a coupled map lattice, and gave a definite improvement. Once a uniform density in every single coordinate is reached, further boundary effects lead to tolerable underestimates of the dimension.

(8) We do not want to dwell too much on the much discussed minimal length needed for a dimension estimate (see Ruelle [1990], Nerenberg & Essex [1990], and Theiler [1990]). The very pessimistic estimate of Smith [1988] is based on the wrong assumption that each point $\vec{x}_i$ has to have many neighbors with $|\vec{x}_i - \vec{x}_k| < \epsilon$ if the correlation integral has be trusted down to ϵ. This would be correct for other generalized correlation integrals, but not the case $q = 2$, as already pointed out in Grassberger [1985].

Often (e.g., in physiological measurements) it is difficult to obtain long time sequences since it is hard to keep the system stationary over long times. But it is often possible to repeat the measurement. In this case, it is advisable to concatenate several short sequences, i.e., to use points from different sequences in Eq. (16). In this way, statistics can be substantially enhanced, and spuriously small dimensions avoided.

(9) While most dimension analyses have been done directly on the unmodified time sequences, at least two other possibilities have been suggested in the literature. In Casdagli [1989], it was suggested first to use the data for fitting the map $\vec{x}_{n+1} = F(\vec{x}_n)$ (see Sec. 8), and then to produce an artificial time sequence from this fit. In low-dimensional models, this gave very good results.

Alternatively, in Kostelich & Yorke [1988] and in Schreiber & Grassberger [1991] time sequences were first cleaned by a noise reduction algorithm described in Sec. 9, and then passed through a dimension algorithm. Results from Schreiber & Grassberger [1991] for a time sequence from a Taylor-Couette experiment (Buzug *et al.*, [1990]) are shown in Figs. 3 and 4. While the raw data did not give any indication of low dimension, the cleaned data show a small but clearly visible scaling region.

(10) We have already discussed the optimal choice of the embedding in Sec. 3. We just want to stress again that the main problem appears to be not so much the optimal embedding but the optimal choice of metric in the embedding space. The most obvious arbitrariness here concerns the choice between the Euclidean and the maximum norm. We have already seen in Sec. 4 that the max norm has some advantages, and it was indeed first proposed in Takens [1983]. We propose that it is in general preferable over the Euclidean norm, in particular if one has noise-free data or if one has applied noise reduction before. The reason is that with the Euclidean norm $C(\epsilon, T)$ does not become independent of τ in the limit of small τ, whence convergence in Eq. (28) is delayed.

A less obvious possibility (not yet studied, to our knowledge) would be to let the metric be anisotropic in embedding space, e.g.,

$$\|\vec{x}_n - \vec{x}_p\| = \max_{0 \leqslant i < m} |x_{n-i} - x_{p-i}|/\delta_i \quad . \quad (31)$$

The condition $\|\vec{x}_n - \vec{x}_p\| < \epsilon$ would then correspond to $|x_{n-i} - x_{p-i}| < \epsilon\delta_i \ \forall i$. A natural

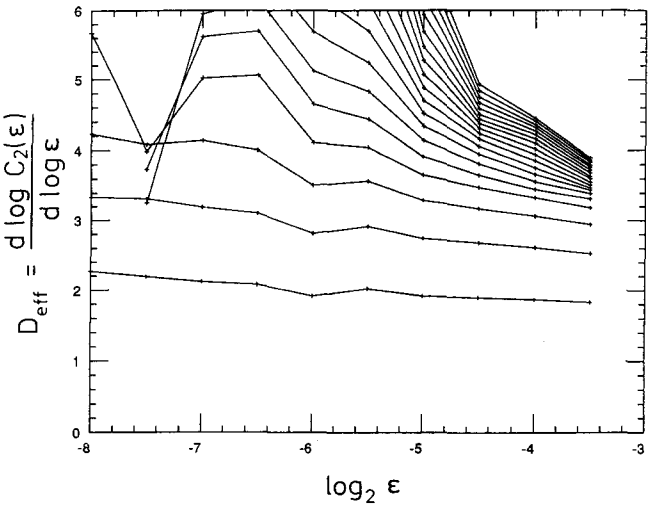

Fig. 3. Local slopes $d \log C/d \log \epsilon$ for data from a Taylor–Couette experiment by Buzug *et al.*, [1990]. The data consisted of 32 768 measurements sampled at 20 Hz, and taken with 10-bit precision. Embedding dimensions range from 2 to 20, and maximum norm was used. No scaling region is visible.

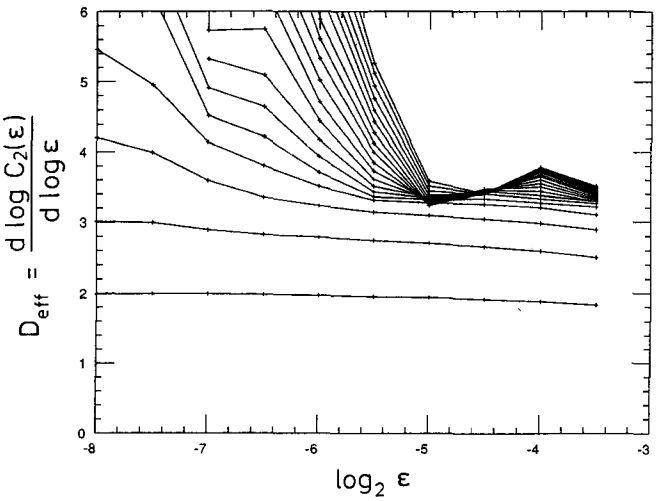

Fig. 4. Same data as in Fig. 3, but after cleaning with the noise reduction algorithm described in Sec. 9.

choice would be one where δ_i increases exponentially both in the forward and in the backward direction. We do not claim that such a choice is preferable, but we believe it illustrates that one has more freedom than often taken for granted.

6.2. *Other generalized dimensions*

As we have already pointed out, the correlation dimension is not very useful in the case of intermittent signals. At least in this case one should try to estimate one of the other generalized dimensions, preferably the information dimension D (and with it, using Eq. (22), the KS entropy h).

There are two natural ways of proceeding: one can either keep the size ϵ of the tubes fixed (as indicated in Eq. (22)), or one can keep the measure fixed by demanding the tubes to contain a given number of trajectories. The latter fixes the tube size of course only up to statistical fluctuations which vanish only in the limit of thick tubes.

(1) Keeping ϵ fixed [Grassberger, 1983], one has first to choose a set of N reference points and another set of M trial points. Then, for each reference point $\vec{x}_i$ one has to count the number $M_i(\epsilon, T)$ trial points closer than ϵ, using an embedding time T. At the end one has to take the averages, over i, of $[M_i(\epsilon, T)]^{q-1}$.

An obvious problem arises in the case $q < 0$ if one of the reference points has no close neighbor. But even when $q \geqslant 0$ and all $M_i \neq 0$, statistical fluctuations imply deviations from the result obtained with the naive estimate $p_i \approx M_i/M$, see Grassberger [1988]. The dominant correction for the Shannon case $q = 1$ is given in Eq. (6) of Appendix B. Since these corrections are only asymptotic, one usually needs very large amounts of data for reliable D_q or $f(\alpha)$ spectra.

In Fig. 5 we show D_q for the Rössler attractor [Rössler, 1976] with $a = 0.15$, $b = 0.2$, $c = 10.0$. We compare several low statistics ($N = M = 9000$) estimates from Liebert & Schuster [1989] with results obtained with very high statistics ($N = 50\,000$, $M =$

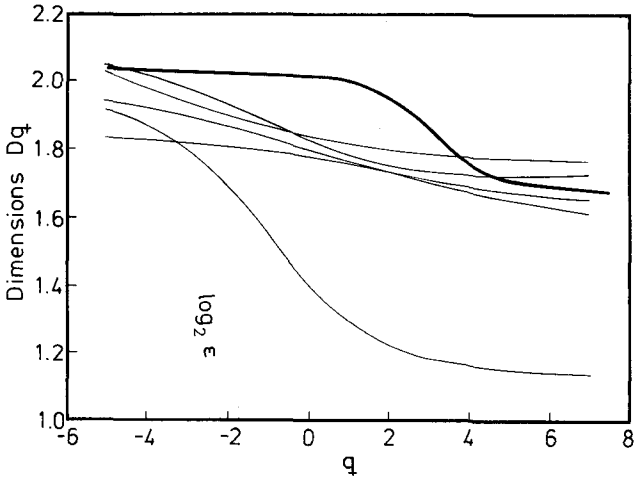

Fig. 5. Generalized dimensions of Rössler attractor with $a = 0.15$, $b = 0.2$, $c = 10.0$. The light curves are low statistics ($N = M = 9000$) data from Liebert & Schuster [1989], and are mainly distinguished by different values of τ. Our own high statistics data ($N = 50\,000$, $M = 10^8$; heavy line) are independent of m and τ within a wide range.

10^8) and with the corrections of Grassberger [1988]. We see essentially no agreement with both sets. The spread in the low statistics data comes mainly from varying τ. With sufficiently-high statistics (which becomes feasible due to the searching algorithm sketched in Appendix A), the dependence on τ disappears. Our results for all $\tau \in [\pi/8, \pi/2]$ and $m\tau \in [\pi, 3\pi]$ agree within the thickness of the line (we might add that D_2 for which we have particularly high statistics seems to be smaller than 2. Thus, the partial measure along the unstable manifold seems to be not smooth; the information dimension is $D = 2.00 \pm 0.01$, in agreement with the Kaplan-Yorke estimate of Wolf *et al.* [1985]).

(2) For estimating generalized dimensions and entropies from tubes with fixed "mass" p, one has to find the distance of k-nearest neighbors among M trial points with $k = pM$. Again a fast searching algorithm is essential. Instead of Eq. (28), we now have [Grassberger, 1985; Badii & Politi, 1985]

$$\frac{1}{N}\sum_{i=1}^{N}[r_i^{(k)}]^{(1-q)D_q} \sim \left(\frac{ke^{h_q T}}{M}\right)^{1-q}, \qquad (32)$$

where $r_i^{(k)}$ is the distance to the k-nearest neighbor of the reference point $\vec{x}_i$. We can use Eq. (32) either by measuring the decrease of the distances with M, or by measuring their increase with k. In the latter case, we have systematic corrections for small statistics (see Somorjai [1986] and Grassberger [1985]).

The advantage in using Eq. (32) is that every reference point has enough points in its neighborhood. Thus we can probe scaling down to very small distances. If the data are noisy, however, this might prove disadvantageous: If we use sufficiently small values of k to be nowhere outside the scaling region, we can be in some places already inside the noise region.

Most of the remarks made in connection with the correlation dimension apply also for generalized dimensions, and will not be repeated.

6.3. *Pointwise dimensions and local information production*

While the generalized dimensions and entropies are averaged quantities, one can also be interested in local behavior. The pointwise dimension $\alpha(\vec{x})$ can be measured directly by counting the number of points within a distance ϵ from the reference point [Mayer-Kress *et al.*, 1988; Havstad & Ehlers, 1989]. Particularly interesting are pointwise entropies, see Drepper [1987, 1988]: They measure the phase space expansion rates during finite times starting from specific points, and hence the predictability or the information "production" by the system. If different regions in phase space have different expansion rates, it might be not worthwhile to put much effort into predicting during times of fast expansion, and one should concentrate the efforts rather on situations in which forecasts are possible.

An interesting method to find such predictable phases is described in Meyer & Packard [1990]. There, a genetic algorithm is used in the evolution of a population of forecasters, selecting the best forecasters for a specific dynamical system and for regions in phase space which are as large as possible.

7. Estimates of Lyapunov Exponents

While the metric entropy measures only the overall phase space expansion in all expanding directions (the sum over all Lyapunov exponents), the spectrum of Lyapunov exponents gives more detailed information on the expansion rates along different directions.

For simplicity, we assume that the system is defined via a map $\vec{x}_{n+1} = \vec{F}(\vec{x}_n)$. Then the divergence between nearby trajectories is governed by the *tangent map*

$$\delta\vec{x}_{n+1} = T^{(n)}\delta\vec{x}_n \qquad (33)$$

with

$$(\mathbf{T}^{(n)})_{ik} = \frac{\partial F_i(\vec{x}_n)}{\partial x_{n,k}} . \qquad (34)$$

The products of tangent maps along a trajectory can be written as a product of two rotations and one diagonal matrix,

$$\prod_{n=1}^{N}\mathbf{T}^{(n)} = \mathbf{R}_a\mathbf{T}_{\text{diag}}\mathbf{R}_b , \qquad (35)$$

and the Lyapunov exponents are just

$$\lambda_i = \lim_{N\to\infty}\frac{1}{N}\log(\mathbf{T}_{\text{diag}})_{ii} . \qquad (36)$$

Thus we need the tangent map if we want to compute Lyapunov exponents.

If the map $\vec{F}$ is defined analytically, the evaluation of the Lyapunov spectrum is easy. One has just to take precautions since multiplying the $\mathbf{T}^{(n)}$ straightforwardly would lead to an ill-conditioned matrix and would produce overflow. In general, one is not interested in all Lyapunov exponents but only in the k largest ones. Then the easiest way to proceed is to iterate a k-bein of vectors $\vec{u}^{(i)}$, $i = 1, \ldots, k$,

$$\vec{u}_{n+1}^{(i)} = \mathbf{T}^{(n)}\vec{u}_n^{(i)} , \qquad (37)$$

and reorthogonalize it from time to time by means of a Gram–Schmidt procedure.

If the system is not given explicitly, but only implicitly through a time sequence, then life is much harder. Early algorithms [Wolf *et al.*, 1985] did not use tangent maps explicitly. The fact that Lyapunov exponent estimates can be made more systematic by fitting tangent maps explicitly was observed in Eckmann *et al.* [1986] and Sano & Sawada [1985].

Assume we have a system with $\leqslant m$ degrees of freedom embedded in m-dimensional delay space, and we want to estimate the tangent map in a point $\vec{x}_n$ of the time series. Then, we must first find $M \geqslant m$ points $\vec{x}_i$ of the time series in a neighborhood of x_n. Let us call this neighborhood $\mathfrak{U}_n$, and let us denote the distance vectors $\vec{x}_i - \vec{x}_n$ by $\vec{u}_n^{(i)}$ (see Fig. 6). Then $\mathbf{T}^{(n)}$ is obtained by a least square fit

$$\sum_{i:\vec{x}_i \in \mathfrak{U}_n} \|\vec{u}_{n+1}^{(i)} - \mathbf{T}^{(n)} \vec{u}_n^{(i)}\|^2 \stackrel{!}{=} \min \quad . \tag{38}$$

There are a few peculiarities when this is applied to delay coordinates. Since $x_{n,i} = x_{n+1,i-1}$, the tangent matrix has the form

$$\begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots \\ T_{m1}^{(n)} & T_{m2}^{(n)} & \dots & T_{mm}^{(n)} \end{pmatrix} \tag{39}$$

and Eq. (38) becomes simply

$$L^{(n)} \equiv \sum_{i:\vec{x}_i \in \mathfrak{U}_n} |u_{n+1}^{(i)} - \vec{a}^{(n)} \vec{u}_n^{(i)}|^2 \stackrel{!}{=} \min \tag{40}$$

where the vector $\vec{a}^{(n)}$ is given by

$$a_i^{(n)} = T_{mi}^{(n)} \quad . \tag{41}$$

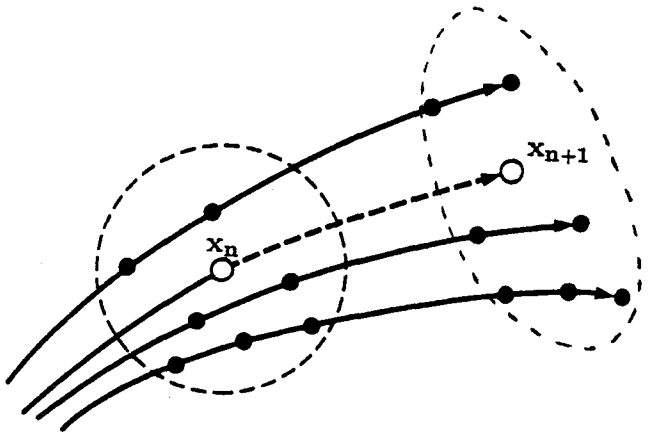

Fig. 6. Evolution of a neighborhood of point $\vec{x}_n$ during one time step. If the neighborhood is sufficiently small, the evolution is governed by the tangent map which is just the linearized version of the map (adapted from Farmer & Sidorowich [1988]).

Thus the map $\vec{x}_n \to x_{n+1}$ is just modeled by a local autoregressive model of order m, as could have been seen also immediately by approximating $x_{n+1} = F(\vec{x}_n)$ by a linear function. It can be considered as an extreme version of a threshold AR where each point $\vec{x}_n$ has its own linear model.

Problems arise if the embedding dimension m is larger than necessary, such that the distance vectors between points in $\mathfrak{U}_n$ are not linearly independent. This implies linear relations

$$\vec{b}^{(n)} \cdot \vec{u}_n^{(i)} = 0 \quad \forall i : \vec{x}_i \in \mathfrak{U}_n \tag{42}$$

and Eq. (40) is not able to determine the vector $\vec{a}^{(n)}$ uniquely. Straightforward minimization of $L^{(n)}$ leads then to spurious Lyapunov exponents [Eckmann *et al.*, 1986]. Notice that this problem does not improve with better data: quite to the contrary, it is worst with noiseless high-statistics data, since then Eq. (42) is best fulfilled.

Explicitly, this is seen, e.g., if the logistic map $x' = a - x^2$ is embedded in two dimensions. We should find one positive Lyapunov exponent λ, but we find actually two positive exponents with values λ and 2λ. Both measure indeed the spreading of the covariance matrix of a cloud of nearby but not infinitesimally nearby points. Since such a cloud is a slightly bent curve in $\mathbb{R}^2$, its covariance matrix is highly anisotropic. The exponent λ measures the spreading along the long axis, while the exponent 2λ describes the "thickening" due to the increasing effect of curvature.

Notice that in practice the minimal embedding dimension is unknown, and it would not give optimal results in the presence of noise anyhow. Thus, with noisy data, in general only the largest Lyapunov exponents can be estimated well by the methods of Wolf *et al.* [1985], Eckmann *et al.* [1986], and Sano & Sawada [1985], while the smaller ones are not accessible [Vastano & Kostelich, 1986; Sato *et al.*, 1987; Holzfuss, 1987; Abarbanel *et al.*, 1990].

There exist several possible modifications to improve the situation. One, suggested in Stoop *et al.* [1991], is to perform SVDs for both clouds $\mathfrak{U}_n$ and $\mathfrak{U}_{n+1}$, and to keep only those rotated coordinates which are deemed essential.

Another modification is suggested by the fact that the density of points in each neighborhood $\mathfrak{U}_n$ will be very anisotropic, being continuous along the unstable directions and Cantor- or delta-like along the most stable directions. With noise and/or curvature effects, the latter is no longer strictly true. But still the distribution along stable directions will be sparse, implying that with points in a spherical neighborhood the mapping along this direction is badly defined. This is of course just the

problem discussed above, but the present formulation suggests a way out: use nonisotropic neighborhoods which are more elongated along the stable directions. Preliminary investigations along these lines gave indeed some improvement for the Hénon map.

Finally, Brown *et al.* [1990] suggested not to fit with linear local maps but with higher-order polynomials. In simple models, quadratic maps indeed gave impressive results. But in more complex system fitting a quadratic polynomial becomes increasingly difficult (the number of parameters to be fitted for each neighborhood is $(m + 1)^2$), and the method might be less efficient.

8. Forecasting and Modeling

One method of (short-time) forecasting is immediately suggested by the preceding paragraph. There, we have always assumed that x_{n+1} is known. If it is not, then the local map constructed from the neighborhood $\mathfrak{U}_n$ can be used to forecast $x_n \to x_{n+1}$ [Farmer & Sidorowich, 1987, 1988; Grassberger, 1986a; Sugihara & May, 1990].

Practically, one first defines $\mathfrak{U}_n$ either by defining its radius or by fixing the number of points it should contain, and searches all points $\vec{x}_i$ in it. Then one fits the linear model, in general weighting each point with some function w which decreases with the distance from $\vec{x}_n$,

$$L^{(n)} \equiv \sum_{i:\vec{x}_i \in \mathfrak{U}_n} \|x_{i+1} - \vec{a}^{(n)}\vec{x}_i - b^{(n)}\|^2 \times \omega(\vec{x}_n, \vec{x}_i; n - i) \stackrel{!}{=} \min \quad . \tag{43}$$

We have here included the possibility that ω also depends on the time difference $t_n - t_i$, in order to allow for nonstationarity. Proper choice of ω is essential and nontrivial for optimal performance. In particular, we should not expect isotropic w's (in Euclidean norm) to be optimal since they put the same weight on neighbors which were close in recent delay coordinates and those which were close earlier. Mees [1989] proposed to choose the neighborhoods such that the points $\vec{x}_i$ form a simplex containing $\vec{x}_n$ inside. In this way, it is guaranteed that $\vec{x}_{n+1}$ is also inside the simplex formed by the iterates $\vec{x}_{i+1}$, and one avoids runaway forecasts.

On one hand, such a local linear model can be considered as a fine-grained limit of threshold AR models; on the other hand it is a straightforward generalization of Pikovsky's [1986] method mentioned in Sec. 3.5. Compared to the latter, it needs more CPU time and it is less robust (forecasts can be occasionally very bad, if the tangent map is ill-conditioned), but with sufficient data it can lead to much better results on average.

A seemingly very different local method was proposed in Linsay [1991]. There it was pointed out that if $\vec{x}_n$ is a weighted average of the neighbor points $\vec{x}_i$,

$$\vec{x}_n = \sum_{i:\vec{x}_i \in \mathfrak{U}_n} p_i^{(n)}\vec{x}_i, \qquad \sum_{i:\vec{x}_i \in \mathfrak{U}_n} p_i^{(n)} = 1 \tag{44}$$

with not necessarily positive weights $p_i^{(n)}$ then

$$x_{n+1} = \sum_{i:\vec{x}_i \in \mathfrak{U}_n} p_i^{(n)} x_{i+1} \tag{45}$$

for any map F which is linear in $\mathfrak{U}_n$. In Linsay [1991], it was claimed that the main advantage of using this as a forecast is that one has to compute only M coefficients $p_i^{(n)}$ while in a linear fit one has to compute $b^{(n)}$ and an entire $m \times m$ matrix $\mathbf{T}^{(n)}$. As we have seen in the last section, the latter is not true if one uses delay coordinates. Indeed, with delay coordinates and with exactly m linearly-independent points $\vec{x}_i$ per neighborhood, Eq. (45) yields *exactly* the same forecast as a local linear fit, with essentially the same numerical effort. In spite of this, Linsay's observation is very interesting in providing an alternative interpretation for locally-linear methods.

When forecasting k time steps, it seems that k iterated one-time-step forecasts give in general better results than forecasts using a single map $\mathfrak{U}_n \to \mathfrak{U}_{n+k}$ (see Farmer & Sidorowich [1988]).

The use of local polynomial fits was also considered in Farmer & Sidorowich [1988] and Crutchfield & McNamara [1987]. As we should expect, it gives better results for low-dimensional cases but seems to become impractical at higher dimensions.

In using local fits, two strategies are to be distinguished. In the last section, we have used for each point a different "model" (i.e., a different vector $\vec{a}^{(n)}$). These "models" were of course not fitted by using only the central point, but they were obtained from larger neighborhoods. Thus one builds each model by using also information from outside its domain of validity. In contrast to these, in the simplicial approximation of Varosi *et al.* [1987], the 'atlas method' of Crutchfield & McNamara [1987] and in TAR's the local model is assumed to hold throughout the region which is used in its fit. This means that these regions have to be larger, and the fit will be worse in general. On the other hand, the latter methods can more readily be considered as modeling approaches.

From local fits with high polynomials it is only a small step to global polynomial fits as considered in Cox [1977], Cremers & Hübler [1987], and Crutchfield & McNamara [1987]. In making global fits, we have definitely passed to the domain of modeling.

Though polynomial global fits can be useful in simple systems, in general we must expect nonpolynomial ansatzes to lead to better results. This was indeed seen

in several investigations. Radial basis functions were used in Casdagli [1989], neural networks in Lapedes & Farber [1987], Broomhead & Lowe [1988], Moody & Darken [1989], Stockbro *et al.* [1990, 1990], and Admoaitis *et al.* [1990]. We do not want to enter into too many details here, since making efficient function fits is a vast field, progress is tied to very technical aspects, and results are expected to improve considerably during the next years.

Before leaving this section, we should again stress that some of these forecasting methods should work also in cases where a separation of noise and signal is not possible or even not meaningful. This is obvious for the method of analogs and its modification by Pikovsky, but it is also true for more powerful and less simple methods.

9. Noise Reduction

9.1. *Linear filters*

The simplest noise reduction schemes are linear low-pass filters. If the signal is highly oversampled, then linear filters can be useful even for nonlinear systems. But in reasonably clean systems one often avoids oversampling in order to reduce the amount of data, and then linear filters can distort the signal badly. In particular, there seems no way how a linear filter can selectively eliminate noise without suppressing also the signal, in not highly oversampled sequences.

While moving-average filters at least leave the dimension of an attractor unchanged (they are just linear transformations), AR filters can even enhance the dimension of an attractor (see Badii & Politi [1986] and Badii *et al.* [1988]) if the damping is too weak. Formally this is seen by means of the Kaplan-Yorke relationship. Intuitively, it is understood by the fact that an AR filter stores the information coming in with the signal during a time inversely proportional to the damping rate. Since the (information) dimension is just the stored information divided by $\log 1/\epsilon$, it is plausible that for too small damping the dimension will finally rise.

An AR filter corresponds to dividing the Fourier transform $x(\omega)$ by a polynomial in $e^{i\omega\tau}$. As shown in Mitschke [1990], the dimension does not increase if one divides $x(\omega)$ by a real function of ω instead.

9.2. *Non-linear methods: Two-step procedures*

It was pointed out in Kostelich & Yorke [1988, 1990], Farmer & Sidorowich [1988, 1991] and Hammel [1990] that noise can be eliminated selectively — without throwing away part of the signal — by using local fits to the dynamics as discussed in the last two sections.

Let us denote the original noisy series by $(x_n, n = 1, \ldots, N)$. In the first step, linear or nonlinear local AR models $f^{(n)}$ are fitted in a neighborhood $\mathfrak{U}_n$ of each point $\vec{x}_n$.

$$y' = f^{(n)}(\vec{y}), \quad \vec{y} \in \mathfrak{U}_n \quad . \tag{46}$$

In the second step, this dynamics is then used to fit a cleaned trajectory $(y_n, n = 1, \ldots, N)$ by demanding that it both satisfies Eq. (46) and resembles the original trajectory x_n. Finally, the whole procedure can be repeated after having replaced x_n by y_n. In this way, one obtains an iteration scheme where each step consists of two parts, a dynamics fitting and a trajectory fitting.

In Farmer & Sidorowich [1991], it is demanded that y_n satisfies Eq. (46) as exact constraints, while the closeness to x_n is measured via the Euclidean norm. Introducing Lagrange multipliers μ_n for the constraints, we end up with a "Lagrangian" L to minimize,

$$L = \sum_{n=1}^{N} \left([y_{n+1} - f^{(n)}(\vec{y}_n)]\mu_n + (x_n - y_n)^2 \right) \stackrel{!}{=} \min \quad . \tag{47}$$

For solving the problems encountered in doing the minimization (essentially inverting ill-conditioned matrices), we refer to Farmer & Sidorowich [1991].

In contrast to this, in Kostelich & Yorke [1988] also the deviation from Eq. (46) was minimized according to the L^2 norm. Also, the trajectory was not cleaned as a whole but only parts of length p were cleaned in one step,

$$L = \sum_{i=n+1}^{n+p} \Big([y_{n+i+1} - f^{(n+i)}(\vec{y}_{n+i})]^2 + [y_{n+i} - f^{(n+i-1)}(\vec{y}_{n+i-1})]^2 + (x_{n+i} - y_{n+i})^2 \Big) \quad . \tag{48}$$

If p were large, both methods would give the same for y_{n+i}, $1 \ll i \ll p$ provided the system is strictly hyperbolic. But if p is finite (in the example given in Kostelich & Yorke [1988], $p = 25$) the ill-conditioning is avoided and the minimization can be done straightforwardly.

9.3. *Nonlinear methods: One-step procedure*

A simplified algorithm which eliminates one of the two steps, with at least as good performance as the above algorithms, is given in Schreiber & Grassberger [1991]. There, local linear fits similar to Eq. (43) are made, except that the vectors $\vec{x}^{(n)}$ are constructed not only

from *delayed* coordinates, but both from delayed and advanced coordinates,

$$\vec{a}^{(n)} \cdot \vec{x}_n = \sum_{i=-l}^{k} a_i^{(n)} x_{n-i} \quad . \tag{49}$$

The corrected value is then simply

$$y_n = \vec{a}^{(n)} \cdot \vec{x}_n + b^{(n)} \quad . \tag{50}$$

Minimizing Eq. (43) with respect to all $a_i^{(n)}$ would give of course simply $a_i^{(n)} = \delta_{i,0}$, $b^{(n)} = 0$, and thus $y_n = x_n$. Therefore, $a_0^{(n)}$ must not be varied but is chosen fixed instead. Maximal correction is obtained with $a_0^{(n)} = 0$. Better and more stable results are obtained if one chooses

$$a_0^{(n)} = 1 - \delta, \quad 0 < \delta \ll 1 \quad , \tag{51}$$

and iterates the procedure just as in the previous two-step algorithms.

Notice that it is essential that both past and future values are used in the fit. Only in this way, the corrected value y_n can be fixed in both the stable and unstable directions. The algorithm performs badly in points where stable and unstable manifolds are tangent. We found that it is sufficient just to leave out the correction in such points (which are recognized by bad linear fits and by large would-be corrections).

The performance of this noise reduction is seen in its effect on dimension estimates shown in Figs. 3 and 4. Similar improvement was found for the Mackey-Glass and Hénon systems in Schreiber & Grassberger [1991]. If x_n was obtained by adding noise to a "true" trajectory obeying the correct model,

$$x_n = z_n + r_n, \quad z_{n+1} = F(\vec{z}_n) \tag{52}$$

("measurement noise", in contrast to "dynamic noise" where $x_{n+1} = F(\vec{x}_n) + r_{n+1}$), then we can distinguish between the mean deviation from determinism.

$$E_{\text{dyn}}(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} [x_{n+1} - F(\vec{x}_n)]^2} \quad , \tag{53}$$

and the deviation from the true orbit

$$E_{\text{true}}(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} [x_n - z_n]^2} \quad . \tag{54}$$

The corresponding correction factors are r_{dyn} (resp. r_{true}), with $r = E(\mathbf{x})/E(\mathbf{y})$. Here, $\mathbf{y} = (y_1, y_2, \ldots)$ is the final corrected orbit. While in Kostelich & Yorke [1988] a value $r_{\text{dyn}} = 4.76$ was obtained for the Hénon map (using a time series of 64K points; no value of r_{true} was quoted there), the above one-step procedure gave $r_{\text{dyn}} = 9.65$ and $r_{\text{true}} \approx 4$ under the same conditions.

Our last example are the Mackey-Glass equations [Mackey & Glass, 1977] with $t_{\text{delay}} = 30$. To an orbit of 32 000 points we added 1% Gaussian noise. Though this is hardly visible on a phase portrait (Figs. 7a, b), it had a very big effect on dimension estimates (see Fig. 8a). In Figs. 7c,d we show the effects which an MA filter $y_n = (x_{n-1} + 2x_n + x_{n+1})/4$ and a *global* linear fit with $k = l = 5$ would have. Both distort the data badly. Our method, finally, improves dimension estimates impressively (Fig. 8b; our best estimate of the correlation dimension as obtained from noise-free data is 3.15), without distorting the phase portrait (Fig. 7e).

10. Conclusions

We have discussed various aspects of nonlinear time sequence analysis. We tried to emphasize aspects which might also be relevant in field observations and in experiments which cannot be perfectly controlled. In particular, we have stressed that phase space correlations are not only useful for measuring attractor dimensions. As seen in Sec. 4, they also provide very sensitive tests for any "determinism beyond the spectrum." They should also be very useful in deciding whether elaborate noise reduction, forecasting, or even global modeling are feasible. Forecasting one time step ahead, e.g., is only possible if the ratio of correlation integrals

$$\frac{C_{m+1}(\epsilon)}{C_m(\epsilon) C_1(\epsilon)} \tag{55}$$

is significantly larger than one. Optimizing embedding dimension m, distance ϵ, choice of norm, etc., by maximizing this ratio could improve locally-linear forecast considerably. Still more useful for this purpose could be generalized correlation sums as discussed in Sec. 6.2 (in particularly the Shannon case $q = 1$), but they are also more difficult to estimate. In Sec. 6, we have discussed in some detail the pitfalls in estimating these correlation sums.

We have also tried to point out similarities between the modern theory of dynamical systems and more classical methods of time series analysis. Very often, both are presented as having hardly anything in common, which certainly is not true. We believe that much progress will result if these similarities are appreciated.

Finally, we tried to avoid too much formalism, and to replace it by more common-sense arguments. Again we hope that it will be fruitful to amalgamate dynamical systems theory with common sense and with the

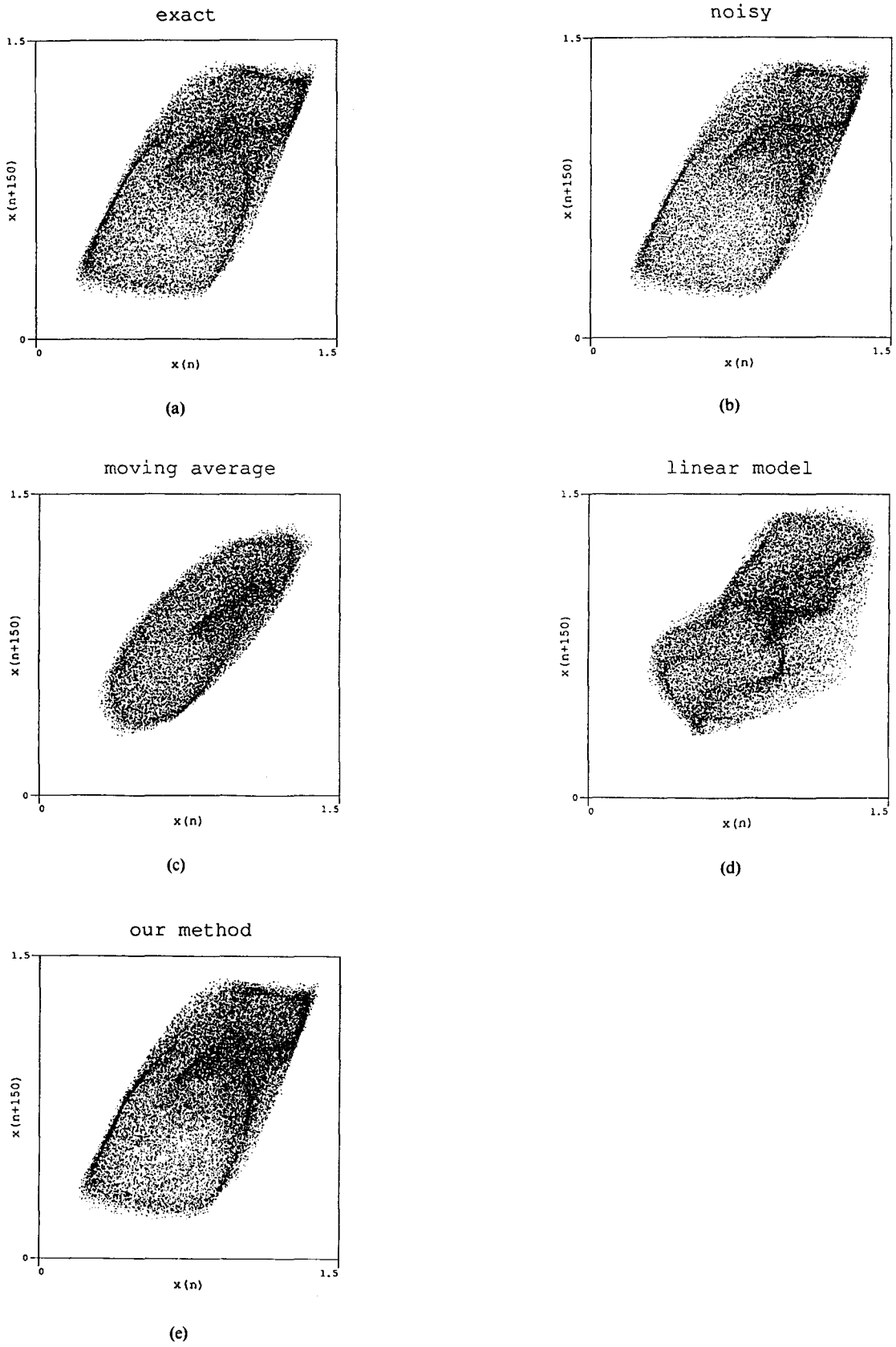

Fig. 7. Each panel shows 32 768 points on an orbit of the Mackey–Glass differential delay equation (approximated by a 500-dimensional map) with $t_{\text{delay}} = 30$ and embedding delay $\tau = 12$. Panel (a) original data; (b) 1% Gaussian noise added; (c) noisy data passed through MA filter; (d) noisy data passed through global linear filter; (e) noisy data cleaned with the method of Schreiber & Grassberger [1991]. Notice that the last essentially agrees with the noise-free data.

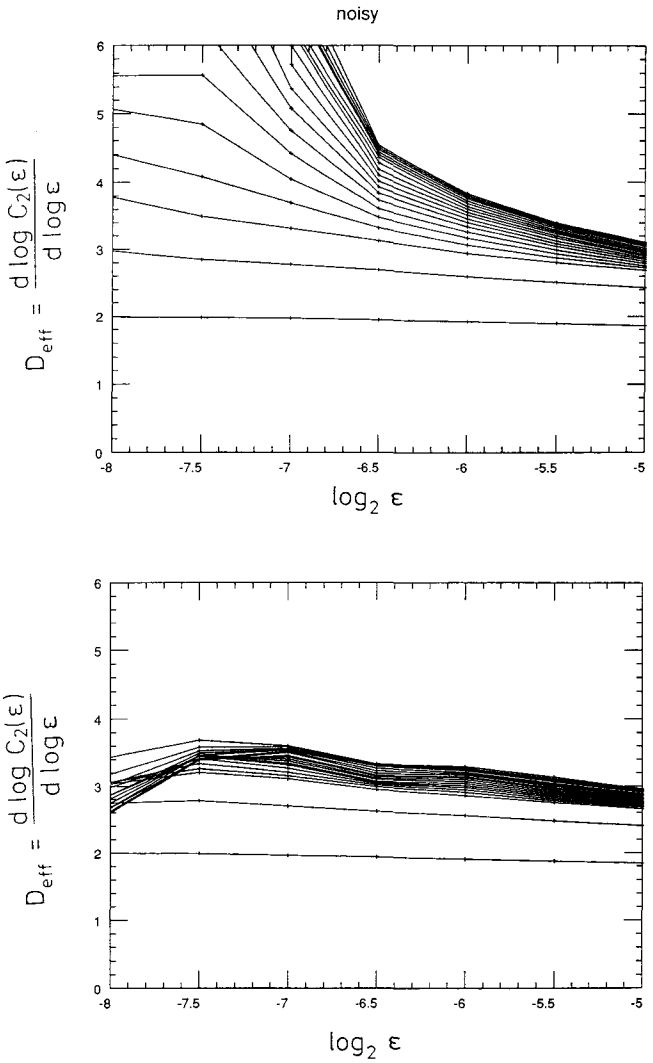

Fig. 8. Effective correlation dimension from the noisy Mackey-Glass orbit shown in Fig. 7. Embedding dimensions are 2-20. Panel (a): uncleaned noisy data; (b): cleaned data [Schreiber & Grassberger, 1991]. Notice the difference in spite of the fact that hardly any difference is seen in the corresponding phase portraits.

practical experience accumulated in various fields such as economics, EEG and ECG analysis, and meteorology. While forgetting old prejudices might be needed for new ideas to take off, one has to remember some of the old expertise if one finally wants to do practical work.

Acknowledgements

We would like to thank the Institute for Scientific Interchange (Torino, Italy) for their kind hospitality during a stay where this work was finished. This work is supported by Deutsche Forschungsgeminschaft SFB 237. One of us (T. S.) receives a grant within the SCIENCE plan of the European Community.

References

Abarbanel, H. D. I., Brown, R. & Kadtke, J. B. [1990] "Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra," *Phys. Rev.* **A41**, 1742.

Admoaitis, R. A., Farber, S. M., Hudson, J. L., Kevrekidis, I. G., Kube, M. & Lapedes, A. S. [1990] "Application of neural nets to system identification and bifurcation analysis of real world experimental data," Los Alamos preprint LA-UR-90-515.

Akaike, H. [1974] "A new look at the statistical model identification," *IEEE Trans. Automat. Control* **AC-19**, 716.

Argoul, F, Arneodo, A. & Richetti, P. [1989] "Symbolic dynamics in the Belousov-Zhabotinski reaction: From Rössler's intuition to experimental evidence for Shil'nikov's. homoclinic chaos," preprint.

Arneodo, A., Argoul, F. & Grasseau, G. [1989] "Transformation en ondelettes et renormalisation," preprint.

Artuso, R., Aurell, E. & Cvitanović, P. [1990] "Recycling of strange sets," (parts I and II) *Nonlinearity* **3**, 325, 361.

Ash, R. [1965] *Information Theory* (Interscience Publishers, New York).

Auerbach, D., Cvitanović, P., Eckmann, J. P., Gunaratne, G. & Procaccia, I. [1987] "Exploring chaotic motion through periodic orbits," *Phys. Rev Lett.* **58**, 2387.

Babloyantz, A. [1989] "Some remarks on nonlinear data analysis of physiological time series," in *Quantitative Measures of Dynamical Complexity in Nonlinear Systems,* ed. N. Abraham, *et al.* (Plenum).

Badii, R. [1989] "Measures of complexity and chaos," ed. N. B. Abraham *et al.* (Plenum, N.Y.) p. 313.

Badii, R., Broggi, G., Derighetti, B., Ravani, M., Ciliberto, S., Politi, A. & Rubio, M. A. [1988] "Dimension increase in filtering chaotic signals," *Phys. Rev. Lett.* **60**, 979.

Badii, R. & Politi, A. [1985] "Statistical description of chaotic attractors: the dimension function," *J. Stat. Phys.* **40**, 725.

Badii, R. & Politi, A. [1986] "On the fractal dimension of filtered chaotic signals," in *Dimensions and Entropies in Chaotic Systems,* ed. Mayer-Kress, G. (Springer) p. 67.

Baladi, V., Eckmann, J.-P. & Ruelle, D. [1989] "Resonances for intermittent systems," *Nonlinearity* **2**, 119.

Biham, O. & Wenzel, W. [1989] "Characterization of unstable periodic orbits in chaotic attractors and repellers," *Phys. Rev. Lett.* **63**, 819.

Bingham, S. & Kot, M. [1989] "Multidimensional trees, range searching, and correlation dimension algorithms of reduced complexity," *Phys. Lett.* **A140**, 327.

Bowen, R. & Ruelle, D. [1975] "The ergodic theory of axiom A flows," *Invent. Math.* **29**, 181.

Box, G. E. P. & Jenkins, G. M. [1976] *Time Series Analysis* (Holden-Day, San Francisco).

Brock, W. A. [1991] "Causality, chaos, explanation and prediction in economics and finance," in *Beyond Belief: Randomness, Prediction and Explanation in Science*, J. Casti *et al.*, ed. (CRC Press, Boca Raton, FL).

Brock, W. & Dechert, W. [1988] "A general class of speciofication tests: The scalar case," *Proc. Business and Economic Statistics Section of the Amer. Statist. Assoc.*, 70.

Broomhead, D. & King, G. P. [1986] "Extracting qualitative dynamics from experimental data," *Physica* **D20**, 217.

Broomhead, D. & Lowe, D. [1988] "Mulitvariable functional interpolation and adaptive networks," *Complex Systems* **2**, 321.

Brosa, U. [1990].

Brown, R., Bryant, P. & Abarbanel, H. D. I. [1990] "Computing the Lyapunov spectrum of a dynamical system from observed time series," La Jolla preprint.

Bulmer, M. G. [1979] *Principles of Statistics* (Dover, New York).

Buzug, Th., Reimers, T. & Pfister, G. [1990] private communication.

Casdagli, M. [1989] "Nonlinear prediction of chaotic time series," *Physica* **D35**, 335.

Cenys, A. and Pyragas, K. [1988] "Estimation of the number of degrees of freedom from chaotic time series," *Phys. Lett.* **A129**, 227.

Chaitin, G. J. [1987] *Algorithmic Information Theory* (Cambridge Univ. Press).

Christiansen, F., Paladin, G. & Rugh, H. H. [1990] "Determination of correlation spectra in chaotic systems," *Phys. Rev. Lett.* **65**, 2087.

Ciliberto, S. [1987] "Fractal dimension and metric entropy in extended systems," *Europhys. Lett.* **4**, 685.

Ciliberto, S. & Bigazzi, P. [1988] "Spatio-temporal intermittency in Raileigh-Bénard convection," *Phys. Rev. Lett.* **60**, 286.

Cohen, A. & Procaccia, I. [1985] "Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems," *Phys. Rev.* **A31**, 1872.

Combes, J. M., Grossmann, A. & Tchamitchian, P. (eds.) [1989] *Wavelets, Time-Frequency Methods and Phase Space* (Springer).

Cox, D. R. [1977] Discussion of papers by Campbell, Walker, Tong & Morris, *J. Roy. Stat. Soc.* **A140**, 453.

Cremers, J. & Hübler, A. [1987] "Construction of differential equations from experimental data," *Z. Naturforsch.* **42a**, 797.

Crutchfield, J. P. & McNamara, B. S. [1987] "Equations of Motion from a Data Series," *Complex Systems* **1**, 417.

Crutchfield, J. P. & Young, K. [1989] "Inferring statistical complexity," *Phys. Rev. Lett.* **63**, 105.

Cvitanović, P. [1988] "Invariant measures of strange sets in terms of cycles," *Phys. Rev. Lett.* **61**, 2729.

Cvitanović, P., Gunaratne, G. & Procaccia, I. [1988] "Topological and metric properties of Hénon-type strange attractors," *Phys. Rev.* **A38**, 1503.

D'Alessandro, G., Grassberger, P., Isola, S. & Politi, A. [1990] "On the topology of the Hénon map," *J. Phys.* **A23**, 5285.

D'Alessandro, G. & Politi, A. [1989] "Hierarchical approach to complexity with applications to dynamical systems," *Phys. Rev. Lett.* **64**, 1609.

Drepper, F. [1987] "Determinism in the information production profile of a stock market index," KFA Jülich preprint.

Drepper, F. [1988] "Unstable determinism in the information production profile of an epidemiological time series," KFA Jülich preprint.

Dvořak, I. [1990] "Takens versus multichannel reconstruction in EEG correlation exponent estimates," *Phys. Lett.* **151**, 225.

Eckmann, J. P., Oliffson Kamphorst, S., Ruelle, D. & Ciliberto, S. [1986] "Lyapunov Exponents from a Time Series," *Phys. Rev.* **A34**, 4971.

Eckmann, J. P. & Ruelle, D. [1985] "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.* **57**, 617.

Eckmann, J. P., Oliffson Kamphorst, S. & Ruelle, D. [1987], "Recurrence plots of dynamical systems," *Europhys. Lett.* **4**, 973.

Efron, B. [1982] *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, Penns.).

Essex, C., Lookman, T. & Nerenberg, M. A. H. [1987] "The climate attractor over short timescales," *Nature* **326**, 64.

Eubank, S. & Farmer, D. [1990] "An introduction to chaos and randomness," Los Alamos preprint, to appear in *1989 Santa Fe Institute Lectures in Complex Systems*.

Farmer, J. D. & Sidorowich, J. J. [1987] "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845.

Farmer, J. D. & Sidorowich, J. J. [1988] "Exploiting chaos to predict the future and reduce noise," in *Evolution, Learning and Cognition*, ed. Lee, Y. C. (World Scientific).

Farmer, J. D. & Sidorowich, J. J. [1991] "Optimal shadowing and noise reduction," *Physica* **D47**, 373.

Flepp, L., Finardi, M., Brun, E. & Badii, R. [1991] to be published.

Fraedrich, K. [1986] "Estimating the dimension of weather and climate attractors," *J. Atmos. Sci.* **43**, 419.

Fraser, A. M. & Swinney, H. L. [1986] "Independent Coordinates for Strange Attractors from Mutual Information," *Phys. Rev.* **A33**, 1134.

Friedman, J. H., Bentley, J. L. & Finkel, R. A. [1977] "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math.* Software **3**, 209.

Giovannini, F. & Politi, A. [1991] "Homoclinic tangencies, generating partitions and curvature of invariant manifolds," *J. Phys. A*, to be published.

Grassberger, P. [1983] "Generalized dimensions of strange attractors," *Phys. Lett. A* **97**, 227.

Grassberger, P. [1985] "Generalizations of the Hausdorff dimension of fractal measures," *Phys. Lett.* **A107**, 101.

Grassberger, P. [1986] "Do climatic attractors exist?," *Nature* **323**, 609 (1986); **326**, 524.

Grassberger, P. [1986a] "Information content and predictability of lumped and distributed dynamical systems," talk presented at Fritz Haber Conf. on *Chaos and Related Non-Linear Phenomena*, Kiryat Anavim, Dec. 1986; *Physica Scripta* **40**, 346 (1989).

Grassberger, P. [1988] "Finite sample corrections to entropy and dimension estimates," *Phys. Lett.* **128**, 369.

Grassberger, P. [1989] *Randomness, Information and Complexity, Proc.* 5th *Mexican Summer School on Statistical Mechanics*, ed. Rechtman, R. *et al.*.

Grassberger, P. [1990] "An optimized box-assisted algorithm for fractal dimensions," *Phys. Lett.* **A148**, 63.

Grassberger, P. [1990a] "Some comments on a proposed universal attractor on wrinkled tori," *Nonlinearity* **3**, 547.

Grassberger, P. & Kantz, H. [1985] "Generating partitions for the dissipative Henon map," *Phys. Lett.* **A113**, 235.

Grassberger, P., Kantz, H. & Mönig, U. [1989] "On the symbolic dynamics of the Henon map," *J. Phys.* **A22**, 5217.

Grassberger, P. & Procaccia, I. [1983] "On the characterization of strange attractors," *Phys. Rev. Lett.* **50**, 346.

Grebogi, C., Ott, E. & Yorke, J. A. [1988] "Unstable periodic orbits and the dimensions of multifractal chaotic attractors," *Phys. Rev.* **A37**, 1711.

Grossmann, A. & Morlet, J. [1985] in *Mathematics and Physics, Lectures on Recent Result*, ed. Streit, L. (World Scientific, Singapore).

Gunaratne, G. H., Jensen, M. H. & Procaccia, I. [1988] "Universal strange attractors on wrinkled tori," *Nonlinearity* **1**, 157.

Hammel, S. M. [1990] "Noise reduction for chaotic systems," *Phys. Lett.* **A148**, 421.

Hamming, R. W. [1977] *Digital Filters* (Prentice-Hall, Englewood Cliffs, N. J.).

Havstad, J. W. & Ehlers, C. L. [1989] "Attractor dimension of nonstationary dynamical systems from small data sets," *Phys. Rev.* **A39**, 845.

Hennequin, D. & Glorieux, P. [1991] "Symbolic dynamics in a passive Q-switching laser," *Europhys. Lett.* **14**, 237.

Henon, M. [1976] "A two-dimensional map with a strange attractor," *Commun. Math. Phys.* **50**, 69.

Holzfuss, J. [1987] "Zur Messung von fraktalen Dimensionen und Lyapunov-Spektren nichtlinearer Systeme am Beispiel akustisch erzeugter Kavitationsblasen," *Ph.D. thesis*, Univ. Göttingen.

Holzfuss, J. & Mayer-Kress, G. [1986] "An Approach to Error-Estimation in the Application of Dimension Algorithms," in *Dimensions and Entropies in Chaotic Systems,* ed. Mayer-Kress, G. (Springer) p. 114.

Honerkamp, J. [1983] "The heart as a system of coupled nonlinear oscillators," *J. Math. Biol.* **18**, 69.

Kaneko, K. [1989] "Pattern dynamics in spatiotemporal chaos," *Physica* **D34**, 1.

Kantz, H. & Grassberger, P. [1985] "Repellers, semi-attractors, and long-lived chaotic transients," *Physica* **D17**, 75.

Karlsson, M. & Yakowitz, S. [1987] "Nearest-neighbor methods for nonparametric rainfall-runoff forecasting," *Water Resources Res.* **23**, 1300.

King, G. P., Jones, R. & Broomhead, D. [1987] *Nucl. Phys. B* (Proc. Suppl.) **2**, 379.

Knuth, D. [1973] The art of computer programming, vol. 3: *Sorting and Searching* (Addison-Wesley, Reading, MA).

Kostelich, E. J. & Yorke, J. A. [1988] "Noise Reduction in dynamical systems," *Phys. Rev.* **A38**, 1649.

Kostelich, E. J. & Yorke, J. A. [1990] "Noise reduction: Finding the simplest dynamical system consistent with the data," *Physica* **D41**, 183.

Landa, P. S. & Rosenblum, M. G. [1989] *Zh. Techn. Fiz.* **59** (11), 1.

Landa, P. S. & Rosenblum, M. G. [1991] "Time series analysis for system identification and diagnosis," *Physica* **D48**, 232.

Lapedes, A. & Farber, R. [1987] "Nonlinear signal processing using neural networks: Prediction and system modeling," Los Alamos technical report LA-UR 2662.

Lathrop, D. P. & Kostelich, E. J. [1989] "Characterization of an experimental strange attractor by periodic orbits," *Phys. Rev.* **A40**, 4028.

Layne, S. P., Mayer-Kress, G. & Holzfuss, J. [1986] "Problems associated with dimensional analysis of EEG data," in *Dimensions and Entropies in Chaotic Systems*, ed. Mayer-Kress, G. (Springer) p. 246.

Ledrappier, F. & Young, L.-S. [1985] "The metric entropy of diffeomorphisms," *Ann. Math.* **122**, 509, 540.

Liebert, W., Pawelzik, K. & Schuster, H. G. [1991] "Optimal embeddings of chaotic attractors from topological considerations," *Europhys. Lett.* **14**, 521.

Liebert, W. & Schuster, H. G. [1989] "Proper choice of time delay for the analysis of chaotic time series," *Phys. Lett.* **A142**, 107.

Linsay, P. S. [1991] "An efficient method of forecasting chaotic time series using linear interpolation," *Phys. Lett.* **A153**, 353.

Lorenz, E. N. [1963] "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130.

Lorenz, E. N. [1969] "Atmospheric predictability as revealed by naturally occurring analogies," *J. Atmos. Sci.* **26**, 636.

Mackey, M. C. & Glass, L. [1977] "Oscillation and chaos in physiological control systems," *Science* **197**, 287.

Mandelbrot, B. B. [1985] *The Fractal Geometry of Nature* (Freeman, San Francisco).

Mayer-Kress, G., Yates, F. E., Benton, L., Keidel, M., Tirsch, W., Pöppel, S. J. & Geist, K. [1988] "Dimensional analysis of nonlinear oscillations in brain, heart and muscle," preprint.

Mees, A. I. [1989] "Modelling Complex Systems," technical report; Math. Dept., Univ. of Western Australia.

Meyer, T. P. & Packard, N. H. [1990] "Local forecasting of high dimensional chaotic dynamics," Urbana preprint.

Mitschke, F. [1990] "Acausal filters for chaotic signals," *Phys. Rev.* **A41**, 1169.

Moody, J. & Darken, C. [1989] "Fast learning in networks of locally-tuned processing units," *Neural Comput.* **1**, 281.

Nerenberg, M. A. H. & Essex, C. [1990] "Correlation dimension and systematic geometric effects," *Phys. Rev.* **A42**, 7065.

Nicolis, C. & Nicolis, G. [1984] "Is there a climatic attractor?" *Nature* **311**, 529.

Omohumdro, S. M. [1987] "Efficient algorithms with neural network behavior," *Complex Syst.* **1**, 273.

Osborne, A. R., Kirwan, A. D., Provenzale, A. & Bergamasco, L. [1986] "A search for chaotic behavior in large and mesoscale motions in the pacific ocean," *Physica* **D23**, 75.

Osborne, A. R. & Provenzale, A. [1989] "Finite correlation dimension for stochastic systems with power-law spectra," *Physica* **D35**, 357.

Oseledec, V. I. [1968] "A multiplicative ergodic theorem," *Moscow Math. Soc.* **19**, 197.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a Time Series," *Phys. Rev. Lett.* **45**, 712.

Paladin, G. & Vulpiani, A. [1987] "Anomalous scaling laws in multifractal objects," *Phys. Rep.* **156**, 147.

Pawelzik, K. [1991] *Nichtlineare Dynamik und Hirnaktivität* (Verlag Harri Deutsch. Frankfurt/Main).

Pesin, Ya. B. [1977] "Characteristic lyapunov exponents and smooth ergodic theory," *Russ. Math. Surveys* **32**, 55.

Pikovsky, A. [1986] *Sov. J. Commun. Technol. Electron.* **31**, 81.

Politi, A., Badii, R. & Grassberger, P. [1988] "On the geometric structure of non-hyperbolic attractors," *J. Phys.* **A21**, L763.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. [1988] *Numerical Recipes* (Cambridge Univ. Press).

Priestley, M. B. [1981] *Spectral Analysis and Time Series*, vols. I and II (Academic, London).

Provenzale, A., Osborne, A. R. & Soi, R. [1991] "Convergence of the K_2 entropy for random noises with power law spectra," *Physica* **D47**, 361.

Renyi, A. [1971] *Probability Theory* (North Holland, Amsterdam).

Rissanen, J. [1986] "Stochastic complexity and modeling," *Ann. Statistics* **14**, 1080.

Rössler, O. [1976] "An equation for continuous chaos," *Phys. Lett.* **A57**, 397.

Ruelle, D. [1986] "Resonances of chaotic dynamical systems," *Phys. Rev. Lett.* **56**, 405.

Ruelle, D. [1986a] "Locating resonances for Axiom A dynamical systems," *J. Stat. Phys.* **44**, 281.

Ruelle, D. [1990] "Deterministic chaos: The science and the fiction," *Proc. R. Soc. London* **A427**, 241.

Sano, M. & Sawada, Y. [1985] "Measurement of the Lyapunov spectrum from chaotic time series," *Phys. Rev. Lett.* **55**, 1082.

Sato, S., Sano, M. & Sawada, Y. [1987] "Practical methods of measuring the generalized dimension and the largest Lyapunov exponent in high dimensional chaotic systems," *Prog. Theor. Phys.* **77**, 1.

Schreiber, T. and Grassberger, P. [1991] "A simple noise reduction method for real data," preprint.

Shannon, C. E. & Weaver, W. [1949] *The Mathematical Theory of Information* (Univ. Illinois Press, Urbana, Ill.).

Sheinkman, J. & LeBaron, B. [1989] "Nonlinear dynamics and stock returns," *J. Business* **62**, 311.

Silverman, B. W. [1986] *Density Estimates for Statistics and Data Analyses*. Monographs on Statistics and Applied Probability vol. 26 (Chapman and Hall).

Smith, L. A. [1988] "Intrinsic limits on dimension calculations," *Phys. Lett.* **A133**, 283.

Sparrow, C. [1982] "The Lorenz equations: Bifurcations, chaos and strange attractors," *Appl. Math. Sciences* **41** (Springer).

Somorjai, R. L. [1986] "Methods for estimating the intrinsic dimensionality of high-dimensional point sets," in *Dimensions and Entropies in Chaotic Systems*, ed. Mayer-Kress, G. (Springer) p. 1370.

Stockbro, K. & Umberger, D. K. [1990] "Forecasting with weighted maps," Nordita preprint.

Stockbro, K., Umberger, D. K. & Hertz, J. A. [1990] "Eploiting neurons with localized receptive fields to learn chaos," Nordita preprint, to appear in *J. Complex Syst.*.

Stoop, R. et al., [1991] "Calculation of lyapunov exponents ...," to appear in *Z. Physik*.

Sugihara, G. & May, R. [1990] "Nonlinear forecasting as a way of distinguishing chaos from measurement errors in time series," *Nature* **344**, 734.

Takens, F. [1981] "Detecting strange attractors in turbulence," in Lecture notes in Mathematics vol. 898, (Springer) p. 366.

Takens, F. [1983] "Invariants related to dimension and entropy," in Atas do 13° Coloquio Brasileiro de Matematica.

Theiler, D. [1986] "Spurious dimension from correlation algorithms applied to limited time-series data," *Phys. Rev.* **A34**, 2427.

Theiler, J. [1987] "Efficient algorithm for estimating the correlation dimension from a set of discrete points," *Phys. Rev.* **A36**, 4456.

Theiler, J. [1987a] "Quantifying chaos," *Ph.D. thesis* (Caltech).

Theiler, J. [1990] "Statistical precision of dimension estimators," *Phys. Rev.* **A41**, 3038.

Theiler, J. [1990a] "Estimating fractal dimensions," *J. Opt. Soc. Am.* **A7**, 1055.

Tong, H. [1983] *Threshold Models in Non-Linear Time Series Analysis*, Lecture Notes in Statistics No. 21 (Springer).

Tong, H. [1990] *Non-Linear Time Series Analysis* (Oxford Univ. Press).

Torcini, A., Politi, A., Puccioni, G. P. & D'Alessandro, G. [1990] "Fractal dimension of spatially extended systems," preprint.

Tsonis, A. A. & Elsner, J. B. [1988] "The weather attractor on very short time scales," *Nature* **333**, 545.

Varosi, F., Grebogi, C. & Yorke, J. A. [1987] Simplicial Approximation of Poincaré maps of differential equations," *Phys. Lett.* **A124**, 59.

Vastano, J. A. & Kostelich, E. J. [1986] "Comparison of algorithms for determining Lyapunov exponents from experimental data," in *Dimensions and Entropies in Chaotic Systems*, ed. Mayer-Kress, G. (Springer).

Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. [1985] "Determining Lyapunov exponents from a time series," *Physica* **D16**, 285.

Yakowitz, S. [1987] "Nearest-neighbor methods for time series analysis," *J. Time Series Anal.* **8**, 235.

Appendix

A.1. *Box-assisted searching and sorting*

A common feature of many methods for nonlinear time sequence analysis is that they use or test local properties of attractors. Thus, most of the CPU time spent in many algorithms (dimension estimates, Lyapunov exponent estimates, noise reduction and forecasting based on local maps) is spent in searching for close neighbors in phase space within a time sequence. The most straightforward search in a series of N points needs a time $\propto N^2$. Much more efficient are methods based on tree-like data representations, (see Friedman *et al.* [1977], Omohumdro [1987], and Bingham & Kot [1989]). Though such methods are often considered to be the most efficient ones, we found box-assisted methods ("radix methods," in the terminology of Knuth [1973]) in general even faster and easier to use, though they need slightly more computer memory. Box-assisted methods are described in Theiler [1987], Kostelieh & Yorke [1990], and Grassberger [1990]. In the following we describe a variant of the algorithm in Grassberger [1990], written however as a general-purpose routine and not only applicable for correlation dimension estimates.

Assume we have a time series of length 'LENGTH' stored in an array 'series'. For any randomly chosen time 't', we want to find all points with distance smaller than ϵ, in two-dimensional delay coordinates with Euclidean norm. Algorithm 1 shows two C subroutines ("put-in-boxes" and "find-neighbors"). The first distributes the points into IMAX×IMAX boxes such that with the second subroutine one can find efficiently all neighbors closer than EPSILON. The time indices of the neighbors found (whose number is returned as 'number') are returned in the array 'foundlist.' The subroutines do not assume that the range of x_t is bounded in any way, though they are most efficient if $0 < x_t < 1$, and IMAX $\approx$ 1/EPSILON. Furthermore, IMAX must be a power of 2.

It should not be hard to rewrite this for higher-dimensional grids, though this is not always advisable even if the embedding is high-dimensional. Often it will be more efficient just to do the boxing in two coordinates (typically the first and the last coordinate in the embedding vector, since they are the least correlated), and to discard those pairs which are not close in the other coordinates.

Though this general-purpose program should already be quite efficient, efficiency can be enhanced further by combining both insertion and retrieval into the same subroutine. Such an algorithm for enumerating correlation sums is given in Grassberger [1990].

To illustrate the simplicity and power of box-assisted methods, algorithm 2 shows a FORTRAN subroutine for ranking NMAX real numbers stored in the array X, and storing their ranks in the array LIST. It does this by means of an auxiliary array BOX of size NBOX + 1. After calling the subroutine, LIST(I) contains the rank of X(I), which is defined as {1 + the number of X(J) less than X(I)}. Already for 5×10^4 points distributed uniformly in [0, 1] and with NBOX = NMAX/2, this routine is on a VAXstation 2000 about 4 times faster than quicksort (as implemented in Press *et al.* [1988]), and 6 times faster than heapsort (also from Press *et al.* [1988]). For nonuniformly distributed points, the CPU time increases $\propto$ NMAX^2/NBOX^{D_2}, where D_2 is the correlation dimension of the distribution.

It is easy to modify this algorithm so that it gives linked lists pointing upwards and/or downwards in rank. Using two or more keys, it is also easy to modify it so that it is still efficient for very nonuniformly distributed points.

A.2. *Information theoretic concepts*

In this paper we shall deal only with probabilistic information theory [Ash, 1965; Shannon & Weaver, 1949; Renyi, 1971], not with algebraic concepts (see, e.g., Chaitin [1987]).

We consider thus a discrete random variable **X** with possible outcomes $x_1, x_2, \ldots, x_N$. This might be obtained by coarse graining a continuous variable, i.e. the outcome x_i might indicate that some other random variable **W** (with possibly continuous distribution) is in the x_ith element of some partitioning. The probabilities prob($\mathbf{X} = x_i$) are denoted as P_i. They obviously satisfy the constraints $0 \leqslant P_i \leqslant 1$ and $\sum_{i=1}^{N} P_i = 1$. In the case of an underlying continuous variable w with measure μ, we have $P_i = \int_i d\mu(\omega)$.

The (Shannon) entropy of **X** is defined as

$$H(\mathbf{X}) = -\sum_{i=1}^{N} P_i \log P_i \quad . \tag{56}$$

It is measured in bits if the logarithm is taken with base 2, in "natural units" if it is the natural logarithm. It has the following properties:

(1) $H(\mathbf{X}) \geqslant 0$, and $H(\mathbf{X}) = 0$ only if all P_i are zero except for one i (if $P_i = 0$, we define $0 \log 0 = 0$).
(2) For fixed N, the maximum of $H(\mathbf{X})$ is attained if all P_i are equal, in which case $H(\mathbf{X}) = \log N$.
(3) If one uses an optimal strategy, the smallest average number of yes/no statements one needs to specify x_i is just H measured in bits (or, more precisely, the smallest integer larger than H). Thus, $H(\mathbf{X})$ is the

average information learned by an observer seeing the actual outcome x_i and the average uncertainty as long as he has not observed it. Notice that Eq. (56) is indeed formally an average value, whence $\log 1/P_i$ can be interpreted as the information (or "bit number") of event x_i.

Of the many other entropy-like quantities considered in the literature, the most relevant for us are the *Renyi* entropies [Ren;yi, 1971]

$$H^{(q)}(\mathbf{X}) = \frac{1}{1-q} \log \sum_{i=1}^{N} P_i^q \quad . \tag{57}$$

Here q is any real number different from 1, and $H(\mathbf{X}) = \lim_{q \to 1} H^{(q)}(\mathbf{X})$.

Among all these, the Shannon entropy is singled out by the fact that information about a composite event can be given in pieces, without any loss. More precisely, let us assume that $\mathbf{X}$ is actually a pair of variables, $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, and $x_{ik} = (y_i, z_k)$. We denote by p_k the probability for $\mathbf{Z} = z_k$, and by $p_{i|k}$ the conditional probability for $\mathbf{Y} = y_i$, conditioned on $\mathbf{Z} = z_k$: $P_{ik} = p_k p_{i|k}$. Similarly, $H(\mathbf{Z})$ is the entropy of $\mathbf{Z}$, and $H(\mathbf{Y}|z_k) = -\sum_i p_{i|k} \log p_{i|k}$ is the entropy of $\mathbf{Y}$ conditioned on z_k. From Eq. (56) we then get

$$\begin{aligned} H(\mathbf{X}) \equiv H(\mathbf{Y}, \mathbf{Z}) &= H(\mathbf{Z}) + \sum_k p_k \mathrm{H}(\mathbf{Y}|z_k) \\ &\equiv H(\mathbf{Z}) + H(\mathbf{Y}|\mathbf{Z}). \end{aligned} \tag{58}$$

The conditional entropy $H(\mathbf{Y}|\mathbf{Z})$ is always $\leqslant H(\mathbf{Y})$, with equality only if $\mathbf{Y}$ and $\mathbf{Z}$ are independent.

The mutual information of $\mathbf{Y}$ and $\mathbf{Z}$ is defined as

$$I(\mathbf{Y}|\mathbf{Z}) = H(\mathbf{Y}) + H(\mathbf{Z}) - H(\mathbf{Y}, \mathbf{Z}) \quad . \tag{59}$$

It can be viewed as the information on $\mathbf{Y}$ we obtain by being told z_k and vice versa. It would also be the redundancy if we describe the outcomes of $\mathbf{Y}$ and $\mathbf{Z}$ without taking into account that possible correlations between them could allow for shorter descriptions. Assume that the variables $\mathbf{Y}$ and $\mathbf{Z}$ are obtained by partitioning (coarse graining) continuous variables, and consider the limit where the partition is very fine. In this limit, $H(\mathbf{Y})$ and $H(\mathbf{Z})$ diverge, but $I(\mathbf{Y}|\mathbf{Z})$ remains finite.

Assume two different probability distributions P and Q over the same set of outcomes x_i. The quantity

$$K(Q|P) = \sum_i P_i \log \frac{P_i}{Q_i} \tag{60}$$

is called the Kullback-Leibler (or 'relative') entropy [Ash, 1965]. It is non-negative, and vanishes only if $Q_i = P_i$ for all i. It has several interpretations, the most important being the following: Assume that P is the correct distribution, but since it is unknown and only the approximation Q is known, the outcomes are encoded such that the code would be optimized for Q. Then $K(Q|P)$ is the excess of average code length.

Let us finally assume that the probabilities P_i have to be estimated from the outcomes of $M \gg 1$ random trials. Let us denote by M_i the number of times the outcome was x_i, and let us assume that some M_i are of order 1. Then the naive estimate $P_i \approx M_i/M$ systematically underestimates the entropy. Laplace's successor rule (see Bulmer [1979]) $P_i \approx (M_i + 1)/(M + N)$, on the other hand, in general overestimates it. In Grassberger [1988], corrections asymptotic in $1/M_i$ were given (also for Renyi entropies), the first and most useful one being

$$H(\mathbf{X}) \approx \frac{1}{M} \sum_i M_i [\log M_i - \psi(M_i)] \tag{61}$$

where $\psi(x) = d \log \Gamma(x)/dx$.

Instead of summing systematically over all possible outcomes i, the sum $\sum_{i=1}^{N} P_i^q$ can also be estimated from the outcomes $i_1, \ldots, i_M$ of a sample random with respect to P,

$$\sum_{i=1}^{N} P_i^q \approx \frac{1}{M} \sum_{j=1}^{M} P_{i_j}^{q-1} \quad . \tag{62}$$

This (and its generalization for the Shannon case) is used in Sec. 6. For $q \neq 2$, this estimate involves again systematic corrections given in Grassberger [1988].

In applying these concepts to time sequences, we assume that $\mathbf{W}$ is a stationary stochastic process, i.e., an infinite sequence of identically distributed (due to stationarity) but not necessarily independent variables $\mathbf{S}_i$. The coarse graining needed to obtain $\mathbf{X}$ consists in lumping together all sequences looking the same inside a finite time window, $\mathbf{X} = (\mathbf{S}_{t+1}, \ldots, \mathbf{S}_{t+n})$. Each $\mathbf{S}_i$ can take a value ('letter') from an 'alphabet' $s_1, s_2, \ldots$. Due to stationarity, the probability for the outcome of $\mathbf{X}$ to be a string $(s_{i_1}, \ldots, s_{i_n})$ is independent of the time t, and will be written $P(s_{i_1}, \ldots, s_{i_n})$. The process is uniquely specified by the alphabet and by the set of all P's. It is called a Markov process of order k if $P(s_{k+1}|s_1, s_2, \ldots, s_k) = P(s_{k+1}|s_2, \ldots, s_k)$.

If $\mathbf{X}$ is a sequence of length n, then we denote its entropy (sometimes called 'block entropy') as H_n,

$$H_n = -\sum_{s_1 \ldots s_n} P(s_1, \ldots, s_n) \log P(s_1, \ldots, s_n) \quad . \tag{63}$$

In a redundancy-free encoding, the difference

$$h_n = H_{n+1} - H_n \tag{64}$$

is the average information for specifying the outcome s_{n+1}, provided the n previous outcomes $s_1, \ldots, s_n$,

are known. For an infinite sequence, the average optimal code length per letter is thus

$$h = \lim_{n\to\infty} h_n \quad . \tag{65}$$

Following Shannon, h is called the entropy of the source emitting the sequence, or simply the entropy of the sequence. As is obvious from their interpretation, the entropies h_n decrease with n. For a Markov process of order k, $h_n = h$ for all $n \geqslant k$. The difference $h_n - h_{n-1}$ is just the Kullback-Leibler entropy difference between the nth and $(n+1)$th Markov approximations to the true probability distribution.

Algorithm 1

```
int imaxl, box [IMAX] [IMAX], llist [LENGTH];
void fill_boxes ( ), find_neighbors ( );
float series [LENGTH];

void fill_boxes (series)
float series [LENGTH];
{   int i, j, il, i2, k;
    float epsinv;

    epsinv=1./EPSILON;
    imaxl=IMAX-1;
    for (i=0; i<=imaxl; i++)  for (j=0; j<=imaxl; j++)  box [i] [j] = 0;

    il=(  (int)  (series [0] * epsinv) )  &  imaxl;
    for ( k=1;  k<LENGTH; k++)
    {   i2=(  (int)  (series [k] * epsinv) )  &  imaxl;
        llist [k] =box [il] [i2];
        box [il] [i2] = k;
        il=i2;
    }
}

void find_neighbors (t, series, foundlist, number)
int  t, foundlist [LENGTH], *number;
float series [LENGTH];
{   int  il, i2, jl, j2, ll, l2, kp;
    float xl, x2, dx, dx2, eps2, epsinv;

    epsinv=1./EPSILON;    eps2=EPSILON*EPSILON;
    *number = 0;
    xl = series [t-1];        il = (  (int) (xl * epsinv) )  &  imaxl;
    x2 = series [t];          i2 = (  (int) (x2 * epsinv) )  &  imaxl;
    for ( jl=il-1; jl<=il+1; jl++)
    {   ll=jl  &  imaxl;
        for( j2=i2-1; j2<=i2+1; j2++)
        {   l2=j2  &  imaxl;
            kp=box [ll] [l2];
            while( kp != 0)
            {   if (kp != t)
                {   dx=(xl - series[kp-1]);       dx2=dx*dx;
                    if (dx2<eps2)
                    {   dx=(x2 - series[kp]);     dx2+=dx*dx;
                        if (dx2<eps2)
                            {     foundlist [*number]=kp; (*number)++;  }
                    }
                }
                kp=llist [kp];
            }
        }
    }
}
```

Algorithm 2

```
      subroutine rank (nmax, nbox, x, list)
        real x (nmax)
        integer list (nmax), box (0:nbox)

        do 100 i=0, nbox
100         box (i) = 0
        do 200 n=1, nmax
            xn=x (n)
            i=int (xn*nbox)
            ip=box (i)
            if  ( (ip.eq.0) .or. (xn.le.x (ip) ) )  then
                box (i)=n
            else
2               ipp=ip
                ip=list (ip)
                if ((ip.gt.0) .and. (xn.gt.x(ip) ) ) goto 2
                list (ipp)=n
            endif
200         list (n) =ip
        n=0
        do 300 i=0, nbox
            ip=box (i)
3           if  (ip.gt.0) then
                n=n+1
            —   ipp=ip
                ip=list (ip)
                list (ipp)=n
                goto 3
            endif
300         continue
      end
```