Nonlinear noise reduction: A case study on experimental data

Holger Kantz, Thomas Schreiber,* and Ingo Hoffmann
Fachbereich Physik, Universität Wuppertal, Gauss-Strasse 20, 5600 Wuppertal 1, Germany

Thorsten Buzug and Gerd Pfister
Institut für Angewandte Physik, Universität Kiel, 2300 Kiel 1, Germany

Leci G. Flepp, Josef Simonet, Remo Badii,† and Ernst Brun
Physik-Institut der Universität, Schönberggasse 9, 8001 Zürich, Switzerland
(Received 8 February 1993)

We apply a recently proposed nonlinear noise-reduction method to time sequences from two different experiments. We demonstrate that it is not difficult to choose the parameters of this algorithm, even though we use no other information about the underlying dynamics than the data themselves. The noise reduction is very robust with respect to changes in the choice of parameters. The reliability of the result is tested by an analysis of the corrections. We discuss the effect of noise reduction on estimates of dimensions, entropies, and Liapunov exponents. For comparison we process one of the sets, densely sampled Taylor-Couette flow data, with a global filter based on singular value decomposition.

PACS number(s): 06.50.−x, 05.45.+b

I. INTRODUCTION

To a certain extent all experimental data are contaminated with noise. This is also true for time series from nonlinear sources, even if they are obtained under well-controlled laboratory conditions. Except in the case of gross oversampling, traditional (e.g, Fourier-based) filters fail since the signal itself can have a broad-band spectrum.

Nevertheless, it is desirable to reduce noise in such signals, in particular if one wants to extract information on small-length scales. This is the case for scaling quantities but also for predictions (see, e.g., [1] for a review on nonlinear time-series analysis in general).

A number of nonlinear noise-reduction methods has been proposed [2–8] but apparently they are not yet much in use. This might be partly due to the fact that they are nontrivial to implement (except maybe the lazy person's version [8]). A more serious concern is that only limited experience with the effect of nonlinear filters on nonlinear signals exists as compared to applying linear filters to linear signals.

Thus we want to demonstrate in this paper how careful nonlinear processing can lead to consistent and reliable noise reduction. This done, much more information can be extracted from the data.

As examples we apply the noise reduction scheme developed in [7] to two data sets: (1) a data set from the Taylor-Couette flow experiment [9] provided by Buzug and Pfister and (2) data from the NMR-laser experiment [10] taken by Flepp, Simonet, and Brun.

II. NOISE REDUCTION WITH LOCAL PROJECTIONS

In this section we briefly recall the main steps to the noise-reduction scheme proposed in [7], which can be seen as a synthesis of two other recent methods [4–6].

Noise reduction always begins with some idea of which features of the data we want to call signal and which noise. It is not *a priori* obvious what the noise part of a given data set is, and it is even less clear how the noise-reduced signal should look. Our approach to nonlinear noise reduction is to assume that the *true* signal is generated by a deterministic dynamics, whereas the noise is random.

So let us assume that our data are generated by a low-dimensional deterministic system with possibly chaotic dynamics, superposed by some additive noise. That is, one measures the noise signal $\mathbf{y}_t$

$$\mathbf{y}_t=\mathbf{x}_t+\boldsymbol{\sigma}_t, \quad \mathbf{x}_t=\mathbf{F}(\mathbf{x}_{t-1}) , \tag{1}$$

but there exists a "clean" trajectory $\mathbf{x}_t$ obeying the unknown deterministic dynamics $\mathbf{F}$ up to some noise $\boldsymbol{\sigma}_t$. We expect the noise to be independent of the signal, have zero average, δ (or fast-decaying) correlation, and some fixed probability distribution. $\mathbf{F}$ may be a map or the time-evolution operator of a first-order ordinary differential equation.

In many realistic situations this will not be the only perturbation of the data, which in addition may contain dynamical noise. This means that the system itself and not only the measurement is disturbed at every moment. Therefore *a priori* no nearby clean trajectory has to exist, which leads to the shadowing problem [11]. In such a case it might not even be excluded that what one would like to call noise is an essential although not deterministic part of the dynamics, as dynamical noise may, e.g., change the stability properties of competing attractors.

All algorithms we are aware of which use dynamics extracted from the noisy data can reduce noise by about one order of magnitude, the differences between them not

being dramatic. It seems that the quality of the fit of the dynamics is one limiting factor; lack of hyperbolicity of the dynamics may be another.

So let us now describe how to implement one efficient noise-reduction procedure. More details and its theoretical background are given in [7], where also its relation to other methods [4–6] is discussed.

In order to reconstruct a unique deterministic dynamics from a scalar signal one has to embed the data in some vector space of sufficiently high dimension, e.g., using delay coordinates [12]. In the latter the problem takes the form

$$y_t = x_t + \sigma_t, \quad x_t = f(x_{t-1}, \ldots, x_{t-m}) , \tag{2}$$

where the embedding dimension is m. The second equation defines the clean dynamics in delay coordinates. Rewriting it in an implicit form,

$$\tilde{f}(x_t, x_{t-1}, \ldots, x_{t-m}) = 0 , \tag{3}$$

shows that in an $(m+1)$-dimensional delay coordinate space the noise-free dynamics is constrained to an m-dimensional hypersurface. For the measured values y_t this is not true, but the extension of the cloud of data points perpendicular to this hypersurface is of the size of the noise level. Therefore one can hope to identify this direction and to correct the y_t by simply projecting them onto the subspace spanned by the clean data. Before this, one has to reconstruct this surface from the noisy data. These are the two main ideas underlying the noise-reduction algorithm below.

In an embedding space of dimension $m+1$ we compute the covariance matrix of all state vectors in a small neighborhood of a given point that we want to correct. The eigenvectors of this matrix are the semiaxes of the best approximating ellipsoid of this cloud of points. Now the important assumption is that the clean signal lives on a smooth manifold with dimension $d < m+1$ and the variance of the noise is smaller than that of the signal. Then for the noisy data the covariance matrix has large eigenvalues corresponding to the dominant directions of the attractor and small eigenvalues in all other directions. Therefore we project the vector under consideration onto the subspace of large eigenvectors to get rid of the noise components. Our fit of the assumed deterministic dynamics thus is a local and linear one, being implicitly contained in the construction of the linear subspace. In [7] this is formulated as a minimization problem. Here we only give the solution.

If we want to compute the correction for the nth embedding vector $\mathbf{y}_n = (y_n, y_{n+1}, \ldots, y_{n+m})$, we first form a small neighborhood $\mathcal{U}$ around this point. From these points $\mathbf{y}_k \in \mathcal{U}_n$ we construct the mean

$$\eta_i^{(n)} = \frac{1}{|\mathcal{U}_n|} \sum_{k(\mathbf{y}_k \in \mathcal{U}_n)} y_{k+i}, \quad i = 0, 1, \ldots, m \tag{4}$$

and the $(m+1)\times(m+1)$ covariance matrix

$$C_{ij}^{(n)} = \frac{1}{|\mathcal{U}_n|} \sum_{k(\mathbf{y}_k \in \mathcal{U}_n)} y_{k+i} y_{k+j} - \eta_i^{(n)} \eta_j^{(n)} . \tag{5}$$

We then introduce a diagonal weight matrix R and define a transformed version of the covariance matrix

$$\Gamma_{ij}^{(n)} = R_i C_{ij}^{(n)} R_j . \tag{6}$$

For $R = I$ we would obtain orthogonal projections as in [5,6]. We prefer to penalize corrections based on the first and last coordinates in the delay window by setting $R_{00} = R_{mm} = r$, where r is large. The Q orthonormal eigenvectors of the matrix $\Gamma^{(n)}$ with the smallest eigenvalues are called $\mathbf{e}_g^{(n)}$, $q = 1, \ldots, Q$. The projector onto the subspace spanned by these vectors is then

$$\mathcal{Q}_{ij}^{(n)} = \sum_{q=1}^{Q} e_{q,i}^{(n)} e_{q,j}^{(n)} . \tag{7}$$

Finally, the ith component of the correction θ_n is given by

$$\theta_{n,i} = \frac{1}{R_i} \sum_{j=0}^{m} \mathcal{Q}_{ij}^{(n)} R_j (\eta_j^{(n)} - y_{n+j}) . \tag{8}$$

This correction is done for each embedding vector, such that we end up with a set of corrected vectors in embedding space. Since each element of the scalar time series occurs in $m+1$ different embedding vectors, we finally have as many different suggested corrections, of which we simply take the average. Therefore in embedding space the corrected vectors do not precisely lie on the local subspaces but are only moved towards them. Furthermore, all points in the neighborhood change a bit such that the covariance matrix has new eigenvectors. Thus one has to repeat the correction procedure several times to find convergence.

A. Implementation

Before we describe more precisely how to implement this algorithm we want to point out that the embedding dimension, the dimension of the subspace, and the sizes of the neighborhoods are important parameters that have to be chosen appropriate to the data. But, as we shall show in the applications, this is in general not difficult, since due to the robustness of the algorithm good results are obtained for a whole range of values.

To apply the algorithm we proceed as follows.

(i) Embed the time series in an $(m+1)$-dimensional phase space using delay coordinates.

(ii) For each embedding vector $\mathbf{y}_n$ find a neighborhood containing at least K points. For efficiency it is advisable to use a fast neighbor-search algorithm, e.g., the one described in [13].

(iii) Compute the center of mass $\boldsymbol{\eta}_n$, the covariance matrix $C^{(n)}$ and the transformed matrix $\Gamma^{(n)}$. The diagonal weight matrix R in Eq. (6) is set to be

$$R_{ii} = \begin{cases} 10^3, & i = 0 \text{ or } i = m \\ 1, & \text{otherwise} . \end{cases} \tag{9}$$

(iv) Determine the eigenvectors of $\Gamma^{(n)}$ and compute the correction according to Eq. (8). Note that the ith element y_i of the time series appears as a component of the delay vectors $\mathbf{y}_{i-m}, \ldots, \mathbf{y}_i$. Therefore its correction is

the average of the corresponding components of the corrections $\theta_{i-m}, \ldots, \theta_i$.

(v) It is obvious that the linear approximation made by projecting onto a subspace induces some error, which can lead to a deformation of the attractor after several iterations. This error causes the local averages of all corrections to be nonzero. Thus we can partly remove this error by subtracting from each correction the local average. This improvement is due to Sauer [5].

(vi) When all corrections are computed, the time series is replaced by the corrected one and the procedure is repeated. Our experience with data from known dynamical systems shows that about eight iterations are a good compromise between accuracy and time consumption.

For a choice of the three variable parameters of the algorithm we follow the considerations in [7] as follows.

(1) *Embedding dimension.* For maps the dimensionality m should not be much smaller than twice the dimension d_{emb}, for which the noise-free dynamics in delay coordinates becomes deterministic (and which is $\leq 2D_f+1$, where D_f is the fractal dimension of the attractor [12]). For flows, especially if the sampling rate is high, m can be much larger because of the redundancy in consecutive data points.

(2) *Number of constraints.* The remaining subspace has to have a dimension as small as possible but big enough not to destroy the attractor, i.e., it should be about the supposed fractal dimension of the attractor.

(3) *Size of neighborhoods.* In [7] we derive a heuristic formula for the desirable size of the neighborhoods:

$$K \approx [\langle \sigma \rangle T^{2/D}]^{2D/(4+D)} . \tag{10}$$

Here T is the length of the time series, D the fractal dimension of the attractor, and $\langle \sigma \rangle$ the relative noise level. Independently of this, the diameter of a neighborhood must not be smaller than the noise level; otherwise the fit of the dynamics may be wrong. In parts of the phase space where the density of points is high this will lead to more than K neighbors. Both criteria give decreasing neighborhoods during the iterations since the noise level will decrease.

Our experience with time series with known clean trajectories shows that a suboptimal choice of these parameters is not fatal. Generally, the corrections are either simply too small or the curvature effects dominate the remaining error.

B. Practical data processing

Dealing with experimental data sets, we adopt the following general strategy. Before doing any processing we become acquainted with the data. First we look at iterates and phase portraits with different delays. This could already detect gross nonstationarity. To obtain an idea of the sampling rate and thus of the redundancy in the signal we compute the time-delayed mutual information [14] and locate its first minimum (if there is any). Following the discussion in, e.g., [1] we do not regard this or any other delay as *optimal* for embedding purposes but we use it as a guideline. If we wanted to extract as much information as possible without an explicit noise-reduction step it could be worthwhile to optimize the embedding parameters [15]. But as we use the noisy data only for qualitative studies we can afford a little sloppiness at this stage. After noise reduction we expect a much reduced sensitivity to embedding parameters.

Now we compute the correlation sum C_2 using the unprocessed data. The correlation sum [16] is defined as

$$C_2^{(m)}(\epsilon) = \frac{1}{T^2} \sum_{i,j} \Theta(\epsilon - \|\mathbf{y}_i - \mathbf{y}_j\|) , \tag{11}$$

where $\mathbf{y}_n$ are the embedding vectors of dimension m. [As usual, we exclude all terms with $|i-j| < t_c$ from the sum (11) to suppress spurious dynamical correlations, where the correlation time t_c has to be estimated from the data. For maps t_c equals unity; for flow data it depends on the sampling rate. See [1] for more details of the implementation]. For sufficiently high embedding m one expects to find a scaling range

$$C_2^{(m)}(\epsilon) \propto \epsilon^{D_2} , \tag{12}$$

where D_2 is the correlation dimension of the attractor. From this we already hope to get a rough estimate of a possible finite attractor dimension. We also expect the noise to affect scaling below a certain length scale which we then take as a first estimate of the noise level.

These estimates should suffice to choose the parameters for the first attempt on noise reduction. With the processed data we can check these estimates, taking the average size of the correction as a new estimator for the noise level. If there is a significant discrepancy we might start again with the new estimates.

The correlation sum Eq. (10) provides more information than only about the dimensionality. The correlation entropy h_2 (see, e.g., [1]) also can be extracted from $C_2^{(m)}(\epsilon)$ through the formula

$$h_2 = \lim_{\epsilon \to 0,\ m \to \infty} \frac{1}{m} \ln[C_2^{(m)}(\epsilon)] . \tag{13}$$

Intuitively, this quantity is related to the rate of production (or consumption) of information due to the chaoticity of the dynamics.

The last quantity we use as an indicator of noise on the data is the set of the Liapunov exponents λ_i. They describe the rate of exponential divergence or contraction of nearby trajectories with time. It is difficult to obtain accurate results for these quantities from a time series [17], but for our purposes it is sufficient to determine the local linear forecast maps in an embedding space of reasonable dimension and to compute the eigenvalues of the time-ordered product of their Jacobians [18].

The fractal dimension and the Liapunov exponents are related through the Kaplan-Yorke formula: After enumerating the Liapunov exponents in decreasing order the Liapunov dimension is

$$D_L = n + \frac{\sum_{i=1}^{n} \lambda_i}{\lambda_{n+1}} , \tag{14}$$

where n is the largest number such that $\sum_{i=1}^{n} \lambda_i > 0$. It is

an estimate for the fractal dimension, which is an upper bound for the correlation dimension D_2. Furthermore, the sum of all positive Liapunov exponents is equal to the entropy h_1, which is an upper bound of the correlation entropy h_2. Therefore we are able to check the consistency of these quantities.

III. TAYLOR-COUETTE FLOW EXPERIMENT

Rotational Taylor-Couette flow is a hydrodynamical system that shows many scenarios of low-dimensional chaos. For proper boundary conditions, period doubling [19], intermittency [20], homoclinic orbits [21], and breakup of two-tori [22] as well as stable three-tori [9] can be found.

A. Experimental setup

The motion of a viscous fluid between two coaxial cylinders is called Taylor-Couette flow. The rotating inner cylinder of the Taylor-Couette apparatus (Fig. 1) is made of stainless steel and has a radius of $r_i = 12.5$ mm. The bottom and top plates as well as the outer cylinder are at rest. The outer cylinder is made of optical polished glass and has a radius of $r_o = 25$ mm. Thus the ratio of the radii is $\eta = r_o / r_i = 0.5$ and $d = r_o = r_i = 12.5$. The accuracy of the radii is better than 0.01 over the entire length of 220 mm. The cylinder length h can be varied continuously by moving the metal collar which provides the top surface of the flow domain, so the aspect ratio $\Gamma = h/d$, used as a geometrical control parameter, takes values between 0 and 17.6.

To provide boundary conditions without cylindrical symmetry, the top plate of the apparatus can be inclined by a small amount. This inclination is a second control parameter of the system.

Silicon oil is used as a working fluid. The external control parameter is the Reynolds number, defined as $\mathrm{Re} = (\Omega d r_i)/\nu$, where Ω is the angular frequency of the inner cylinder and ν is the kinematic viscosity. The temperature of the fluid is held constant to within 0.01 K by circulating thermostatically controlled oil through a surrounding square box. A phase-locked-loop (PLL) circuit controls the speed of the inner cylinder to be better than one part in 10^{-4} in the short term and better than one part in 10^{-6} in the long-term average.

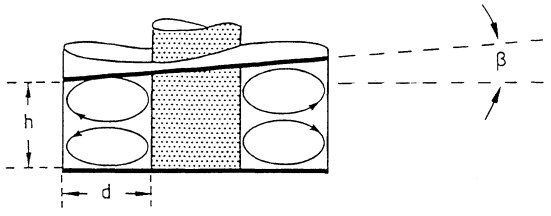


FIG. 1. Experimental setup of the Taylor-Couette flow experiment. The data were taken with $\beta = 0.23°$ and $h = 4.675$ mm.

B. Measurement technique

The local-velocity component of the flow field is measured by a real-fringe laser Doppler anemometer and recorded by a PLL analog tracker. The analog output voltage of the tracker is proportional to the velocity. Two Bragg cells are used to determine the velocity direction by applying a bias to the signal. The frequencies of both cells are 40 and 40.1 MHz, shifting the zero velocity to 100 kHz. The working fluid has to be seeded with light-scattering particles to obtain proper signal quality. The particles used are latex spheres of diameter 6.8 μm. According to the arrival statistics of scattering particles the signal shows Doppler-phase noise. For further signal processing the velocity signal is fed into an analog-to-digital converter (ADC) with a 12-bit resolution and then to a digital computer.

The data to be analyzed here were obtained with a Taylor cylinder of very small length. The aspect ratio Γ was chosen to be 0.374. In this range one finds period-doubling scenarios in the restabilized symmetric two-vortex state when the top plate is inclined. The corresponding bifurcation diagrams into chaos for an inclination of 0.23° can be seen in [19,22]. From this scenario a time series is measured in the chaotic regime at a Reynolds number $\mathrm{Re} = 705$. The sampling time is $T_s = 20$ ms and the total number of data points is 32 768.

C. Analysis of the Taylor-Couette data

A phase portrait of the original data is shown below [Fig. 2(a)] where it can be compared with the cleaned version. The time-delayed mutual information has a first minimum at a delay of 12 time steps. We compute the correlation sum (11) and plot the local slopes of a log-log plot of C_2 vs ϵ (Fig. 3). From this plot we can already expect an attractor dimension of about 3. Noise destroys the scaling behavior below $\frac{1}{16}$ the size of the attractor. Thus we make the following choice of parameters for the noise reduction.

(1) With a guess for the attractor dimension of 3 and the rather fine sampling we should choose an embedding dimension of at least 7 ($m = 6$ in the notation of [7]), rather larger. We try 9.

(2) To leave room for three local degrees of freedom we impose six constraints (which does *not* limit the attractor dimension to 3, as discussed in [23]).

(3) Since the effect of noise seems to destroy proper scaling below $\frac{1}{16}$ the attractor size, we require neighborhoods in the first iteration not to be smaller than this. In addition, we evaluate Eq. (10) and search for at least 32 neighbors for each neighborhood.

With these parameters we form a corrected version of the time series. If we suppose that the average size of the correction made is close to the size of the noise (which is a reasonable assumption even if it is not just exactly the noise we remove), we can give a new estimate of the noise level. It turns out that the data is corrupted with about 5% noise, which is slightly less than what we estimated above. Thus we rerun the noise reduction with the new requirement for the size of the neighborhoods.

A phase portrait of the second corrected version, which turns out to be hardly distinguishable from the first one, is shown in Fig. 2(b). The second run confirms a noise level of 5%. The drastically improved scaling region in the correlation sum (Fig. 4) suggests that the noise level has been reduced by about a factor of 4.

We started with the assumption that the measured data are additively composed of a low-dimensional deterministic part and random noise. If what we removed is really additive noise it should be uncorrelated with the signal (multiplicative noise or errors by nonlinear measurement devices *would* be correlated). If the noise is approximately white the corrections should have fast-decaying autocorrelations. In Fig. 5 we can see that indeed the cross correlation between signal and correction is very weak. The same holds for the autocorrelation of the correction (Fig. 6). Both figures show an interesting feature: The first set of parameters was chosen with too high an estimate of the noise level. Therefore we took rather large neighborhoods, which induces errors due to the linear approximation. These errors are proportional to the local curvature of the attractor and thus induce correlations between signal and corrections. These are visibly smaller for our second set of parameters, where smaller neighborhoods were used. We check that the corrections are distributed according to a normal distribution with width 5% of the attractor size. However, we think this is not a very significant observation. Added up through eight iterations even random corrections would be distributed almost as a Gaussian. This test is more useful if we know that the measurement errors are not normal. This is, e.g., the case when the data are coarsely digitized. Then we could check if the distribution of the corrections is consistent with what we know about the errors.

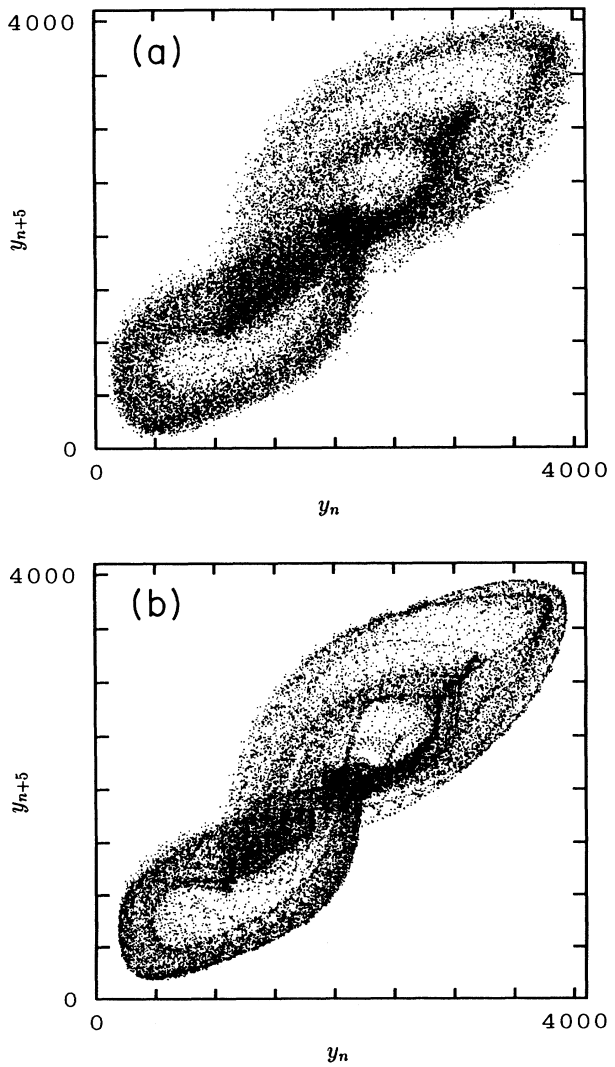


FIG. 2. Phase portrait of the Taylor-Couette data. In both panels the delay time is 5 steps. Panel (a) shows the unprocessed data, panel (b) the same data after noise reduction.

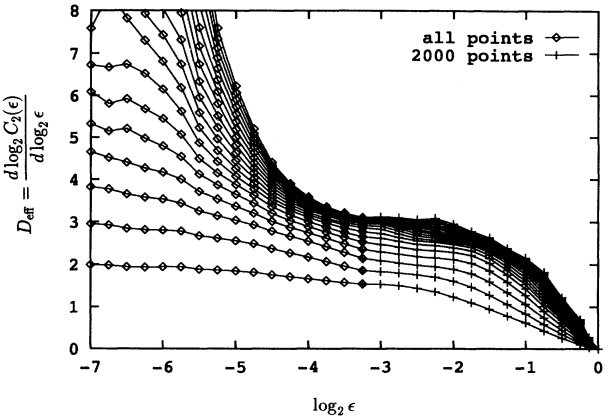


FIG. 3. Local slopes of the correlation sum for the unprocessed Taylor-Couette data. The time delay was 5 steps and embedding dimensions 2–20 are shown. For the coarse structure (above $\epsilon = 2^{-3}$) only 2000 points were used.

Encouraged by the clear scaling of C_2, which is consistent with a correlation dimension D_2 of 3, we also compute the corresponding generalized correlation sums C_q for $q \geq 1$ (see again [1] for details, e.g., concerning finite sample corrections). In Fig. 7 we shifted the curves

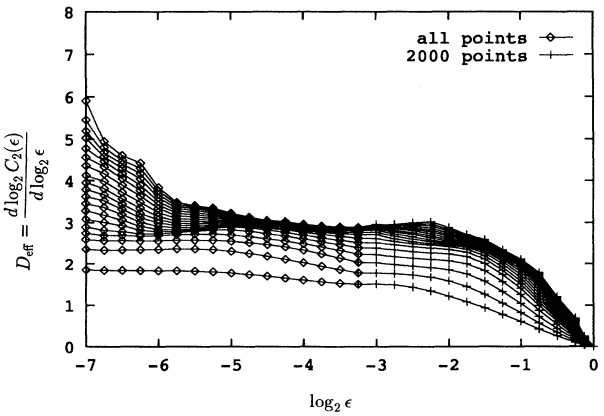


FIG. 4. Same as in Fig. 3 but using the set after noise reduction. Now a clear scaling region is visible.

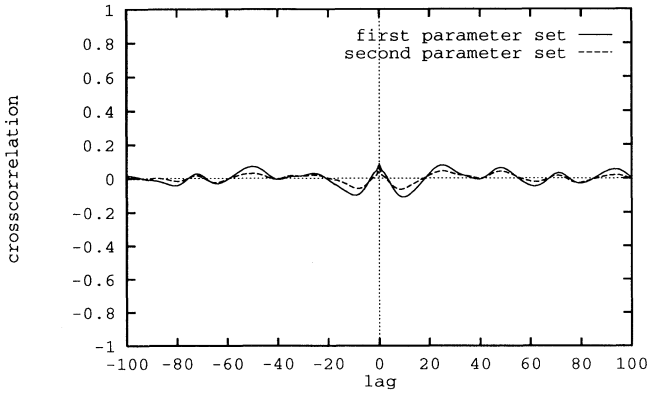


FIG. 5. Normalized cross correlation between signal and correction (Taylor-Couette data). A positive lag time l means that the product is formed between the signal at time t and the correction at time $t+l$. The two lines correspond to two attempts at noise reduction with different parameters.

by normalizing with $C_q(\epsilon=2^{-4})$. Thus we see that the data for $1\le q\le 11$ lie on straight lines with slightly different slopes. From these slopes we can extract numbers for D_q which decrease with increasing q, as they should. The different slopes are a sign of the multifractal structure of the attractor. However, with the material at hand we do not dare to estimate an $f(\alpha)$ spectrum or even part of it.

Following Eq. (13) we estimate the correlation entropy to be $h_2\approx 0.15$. The Liapunov exponents we find for a three-dimensional embedding are $\lambda_1=0.12\pm 0.02$, $\lambda_2=0\pm 0.01$, and $\lambda_3=-0.45\pm 0.03$, but an estimate of negative exponents is never very safe (see the discussion in [17]). From this we find a Liapunov dimension of about 2.25, which is smaller than our estimate of D_2 and thus not consistent with this, but which also depends very strongly on the size of λ_3. Note that $\lambda_2=0$, as it should be. On the original noisy trajectory we find one exponent close to zero and two negative exponents. This wrong result is a consequence of the fact that the diffusive deviation of trajectories due to noise is much stronger than the deterministic exponential divergence. If one wants to

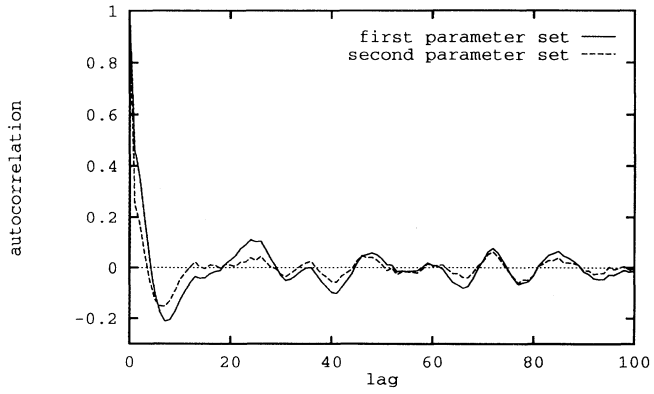


FIG. 6. Normalized autocorrelation of the correction for the Taylor-Couette data.

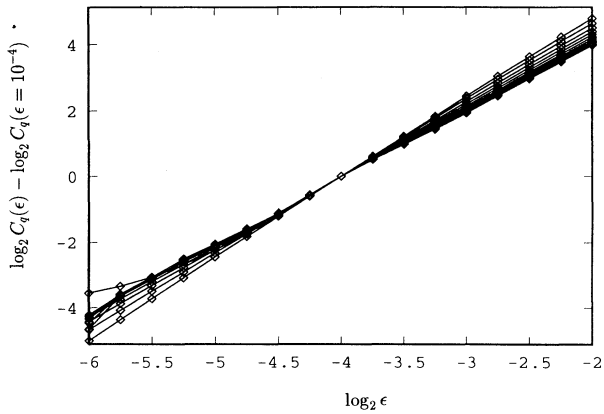


FIG. 7. Generalized correlations sums C_q for $1\le q\le 11$ (Taylor-Couette data). The curves are shifted vertically by suitable normalization. Thus the $\log_2 C_q$ axis has different offsets for the different curves. Delay 5 was used for a 10-dimensional embedding. Lower left-hand corner (top to bottom): $q=1$, $q=11,\ldots,2$; upper right-hand corner (top to bottom): $q=1,\ldots,11$.

compute Liapunov exponents on noisy trajectories, one has at least to use large time delays. By this the divergence rate is multiplied with the time step, whereas the noise level is the same as before.

Finally, we want to remark that this data set is a case where a global filter based on *singular value decomposition* [24,25] can be applied to reduce noise by a certain amount. This is due to its relatively fine sampling rate, which suggests even the use of a simple low-pass filter. Both, however, give only a very slight improvement of scaling properties, which shows that this kind of noise reduction is not more than a smoothing of the trajectories without restoring any structures.

IV. PARAMETRICALLY MODULATED NMR LASER

The second data set stems from a NMR-laser experiment.

A. Design of experiment and measurement

A ruby NMR laser is parametrically modulated with a sinusoidal driving signal near the relaxation oscillation frequency of the free-running system. In this way a natural time is introduced which facilitates greatly the analysis of the synchronously strobed laser output.

The NMR laser [10], like any laser, has two essential ingredients: the radiating "particles" (atoms, molecules, electrons, nuclei, etc.) and the radiation field produced by the particles. An external pump causes the population inversion where the higher-energy states of the particles are more strongly populated than the lower ones. A resonant structure (cavity), which endorses the particles, provides the feedback for the laser field thus causing the coherent excitation and radiation of the particles. In the ruby NMR laser the lasing "particles" are the nuclear spins of the ^{27}Al in $Al_2O_3{:}Cr^{3+}$, and the radiation field is a magnetic radio-frequency field sustained by the tuned

NMR coil that forms the cavity of the laser. Spin inversion beyond the first laser threshold is obtained by means of dynamic nuclear polarization (DNP) at the lattice temperature of 4.2 K. DNP is achieved by shining microwaves at Cr^{3+}, thus causing electronic-spin transitions which pump the nuclear spins to the lasing state.

The free-running NMR laser typically oscillates at the center frequency $\nu_a = 12.3$ MHz of the $(\frac{1}{2}, -\frac{1}{2})$ NMR line of ^{27}Al in a static magnetic field $B_0 = 1.1$ T. The NMR cavity is thereby tuned to ν_a. The NMR laser thus acts as a tuned single-mode, homogeneously broadened, unidirectional laser. The NMR laser field induces a voltage in the receiver coil L which can be tapped at the tuning capacitor C. The NMR laser signal is thus a rf voltage $V(t)$ across C. Since the free-running laser is essentially a two-dimensional dynamical system, an additional degree of freedom is required in order to bring it into a chaotic state. This is achieved, for example, by a small modulation of the quality factor $Q(t) = Q_0(1 + p \cos 2\pi\nu_m)$ of the NMR cavity. As laser chaos consists primarily of irregular variations of the amplitude of V, the rf laser signal is demodulated and its envelope $v(t)$ only is recorded. We call $v(t)$ the laser output. That output is proportional to the transverse nuclear magnetization $M_t = (M_x^2 + M_y^2)^{1/2}$ which can be theoretically calculated from the highly accurate extended Bloch model (EBL) [26].

The NMR laser turned out to be an extremely versatile device because of some outstanding properties: long-time stability, high signal-to-noise ratio, good control over system parameters, and low-dimensional dynamics. The crucial experimental parameters are set by a data-acquisition and control system based on an Olivetti M24 PC and a Mandax386, which are complemented by mathematical coprocessor, ADC board, IEEE control units, and various digital input/output ports. The experimental time series are collected by sampling the laser output with a 12-bit resolution analog-to-digital translation board at a frequency $\nu_s = 15\nu_m$. The modulation frequency is kept constant at $\nu_m = 91$ Hz. We choose a modulation amplitude p slightly below the crisis point p_c where a sudden change in the size of the strange attractor occurs [27,28]. The rather low modulation frequency allows a double-buffering acquisition technique to collect the time series consisting of $N = 600\,000$ integers.

B. Analysis of the NMR-laser data

The data set studied in this paper originally consisted of 600 000 12-bit integers, the last 1800 of which were discarded after inspection since the synchronization with the modulation seems to have been lost. Although from the point of view of good statistics this amount of data is desirable, for many purposes the size of the data set is rather inconvenient under the aspect of numerics. Thus we shall mainly concentrate on the analysis of Poincaré sections.

C. Poincaré sections

The only knowledge we use in addition to the data themselves is that they are sampled at a rate $\nu_s = s\nu_m$, which is an integer multiple of the modulation frequency ν_m. In this case $s = 15$. Thus defining a Poincaré section by requiring equal phase of modulation amounts to taking only every 15th data point.

Thus a Poincaré section consists of 39 880 points and there are $s = 15$ different possibilities to choose the phase which defines the surface of section. We choose the stroboscopic section which yields the largest variance to guarantee the highest signal-to-noise ratio.

But there is still a better way to form a Poincaré section than just taking one of the stroboscopic sections. As mentioned, all 15 different phases yield valid sections:

$$\hat{y}_t^{(p)} = y_{st+p}, \quad p = 0, \ldots, s-1, \quad T = 1, \ldots, 39\,880 . \tag{15}$$

In fact, almost all linear combinations of the 15 points within a cycle of the modulation

$$\hat{y}_t = \sum_{p=0}^{s-1} a_p y_{st+p} \tag{16}$$

formally give valid coordinates, but they are not equally good: they differ in their variance. We are interested in the linear combination with the highest variance, holding the norm $\sum_{p=0}^{s-1} a_p^2$ of the coefficient vector fixed. This is done by a principal component analysis on the $s \times s$ covariance matrix [25] of the whole series of $T = 598\,200$ points

$$C_{pq} = (T/s)^{-1} \sum_{t=0}^{T/s-1} (y_{st+p} - \eta_p)(y_{st+q} - \eta_q) , \tag{17}$$

where $\eta_p = (T/s)^{-1} \sum_{t=0}^{T/s-1} y_{st+p}$ is the average of the section with phase p. The eigenvector of C with the largest eigenvalue gives the linear combination with the highest variance. A phase portrait of the Poincaré section thus obtained is shown in Figs. 8 and 9. Indeed we

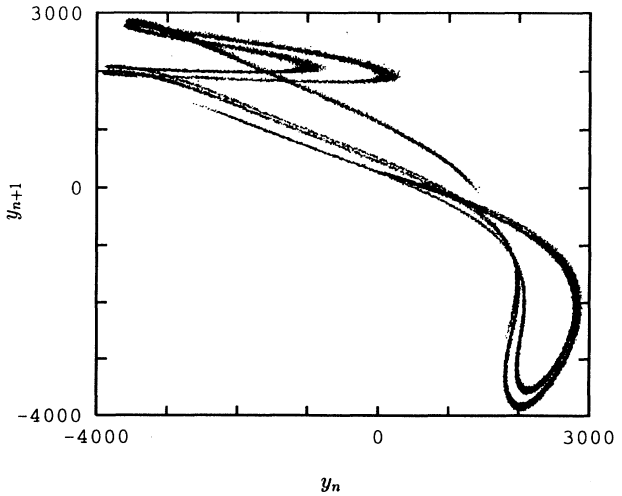
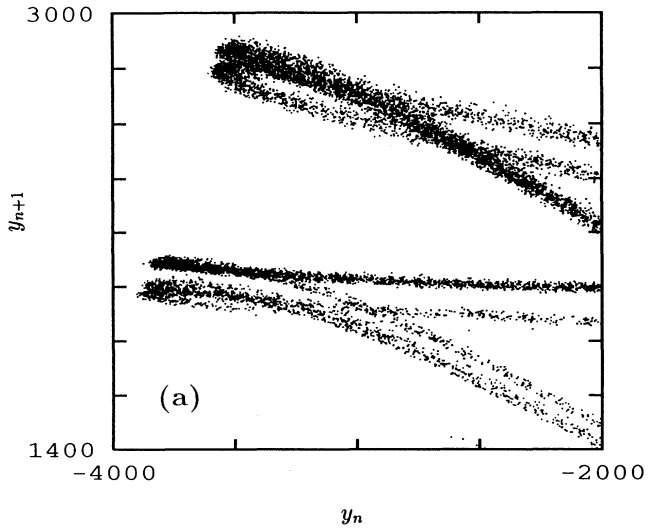


FIG. 8. Phase portrait of the Poincaré section of the NMR-laser data. The delay time is 1.

will see below that the noise level of this particularly chosen section is 1.1%, significantly lower than the best stroboscopic section containing 1.8% noise. The analysis in this section is done with this Poincaré cut. Similar results are obtained with a stroboscopic section.

Also for this set we estimate the correlation sum C_2 (Fig. 10). From this we might expect a dimension below 2 and a noise level of $\frac{1}{64}$ the attractor size. Thus we choose parameters for noise reduction as follows.

(1) With an attractor dimension below 2 and the map-like sampling we choose an embedding dimension of 7 ($m=6$ in the notation of [7]), i.e., slightly higher than necessary.



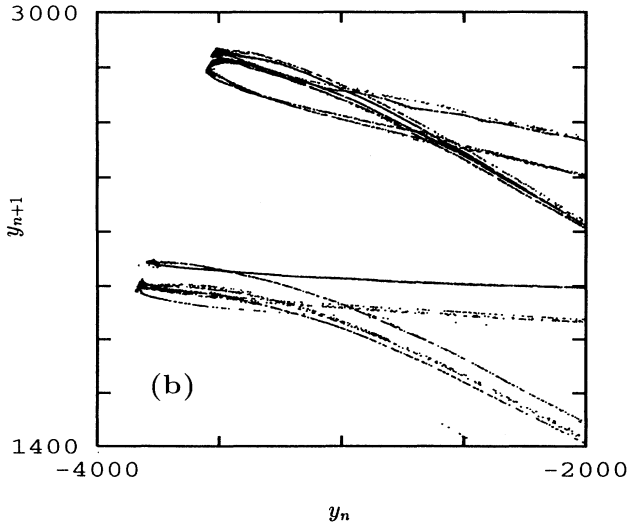


FIG. 9. Enlargements of phase portraits of the NMR-laser data. Panel (a) shows the unprocessed data (same as in Fig. 8), panel (b) the same region after noise reduction.

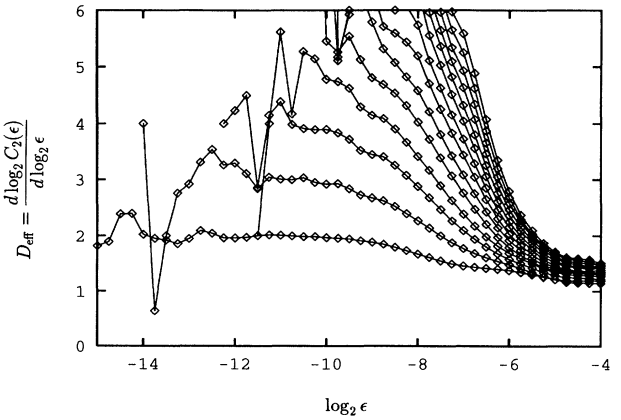


FIG. 10. Local slopes of the correlation sum for the unprocessed NMR-laser data (Poincaré section). Delay 1 was used in 2–15 dimensions.

(2) The five constraints we impose leave room for two local degrees of freedom.

(3) In the first iteration we form neighborhoods of at least $\frac{1}{64}$ the attractor size and a minimal number of 64 neighbors, as obtained from Eq. (10).

The corrections we compute during eight iterations have an amplitude amounting to 1.1% the amplitude of the signal. This is close to the $\frac{1}{64}$ we guessed, so we do not have to repeat the procedure with new parameters. We find that the autocorrelation of the corrections as well as the cross correlations between the signal and the corrections are very small (<0.04), as they should if the separation into signal plus additive noise makes sense.

Again we see an enormously enlarged scaling region of C_2 (Fig. 11). However, the values for D_2 we could read off for increasing embedding dimensions do not converge to a constant value as desired. Instead they increase with a rate of about 0.03 per new coordinate. This can have different explanations. Somehow, a higher-dimensional component is visible in the system on all length scales.

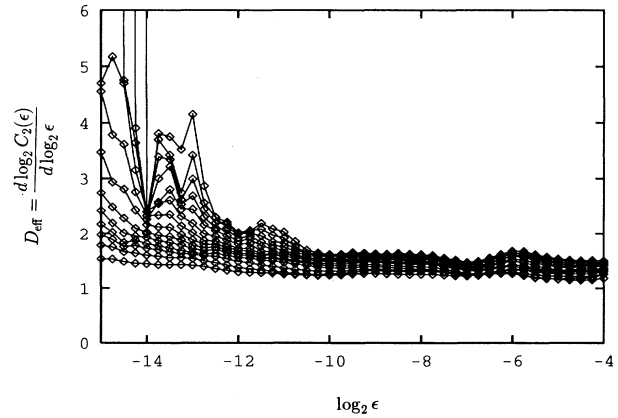


FIG. 11. Same as in Fig. 10, now after noise reduction. Scaling is visible over more than 2 decades.

This might come from some source of multiplicative or dynamical noise. It could be due to fast fluctuations in some parameter, or a mechanical or electronic disturbance of the apparatus, but also due to fluctuations in the synchronization of sampling and modulation. We want to stress that it cannot be an artifact of our noise-reduction algorithm, since it is visible also in the original data on scales greater than 2^{-5} (see Fig. 10). Nevertheless, the striking feature of the data is its scaling behavior over more than two decades.

From the correlation sums of different embedding dimensions we estimate the entropy $h_2=0.3\pm0.05$ with a slight increase towards smaller scales, which is related to the above mentioned increase of the correlation dimension. Using three-dimensional delay coordinates we find Liapunov exponents $\lambda_1=0.29\pm0.02$, $\lambda_2=-0.59\pm0.02$, and $\lambda_3=-1.22\pm0.02$. This is in good agreement with the entropy and yields a Liapunov dimension of $D_L=1.5$.

Despite the lack of convergence for high embedding dimensions we can try to estimate generalized dimensions D_q. We just have to agree on a fixed embedding dimension and range of length scales. In Fig. 12 we can see that all curves for $2\leq q\leq 11$ are close to a straight line with slope 1.45. A least-squares fit to the data shows, however, that the slopes decrease systematically with q from $D_2=1.5$ to $D_{11}=1.33$ with errors of about 0.005. Only the curve for C_1 corresponding to the information dimension deviates at smaller length scales. D_1 is much more sensitive to sparse or empty neighborhoods, where one systematically obtains too flat curves of C_1. Thus we have a weak indication of multifractality in this data set.

D. Flow data

Up to now we have concentrated on a Poincaré section of the data. Usually one expects that such a section shows an attractor which is exactly one dimension lower than that of the flow. The additional dimension can be thought of as being spanned by the coordinate obtained by parametrizing the orbit by the evolution time. However, the data at hand are taken in a very special way: between two intersections exactly an integer number of samples are taken, the intersections themselves coinciding with measurement times. Thus the additional coordinate takes only s discrete values and is not able to raise the attractor dimension by 1, but only s discrete and disjoint copies of the low-dimensional attractor are formed, which does not yield the expected higher dimension.

Therefore if one is interested in analyzing the flow data rather than Poincaré sections one has either to do a non-synchronized measurement or to artificially remove the synchronization of the data. This can be done by choosing a new sampling time which is not commensurate with the modulation time. Data sampled with this rate can be formed by interpolating between the measured values. We created a new time series by resampling the flow data with a sampling time which was a factor $e=2.718\,28\ldots$ longer than the original one, thus also reducing the length of the series by the same factor. We do not repeat all the above points for the flow data set. We indeed find a correlation dimension D_2 which is higher by 1.

Generally, the reason for not using a Poincaré section is that the intersection points are not equally spaced in time and that the resulting data set is too small. None of these problems occurs in this case. With the amount of data at hand, the loss of information in taking a Poincaré section (which we also minimized by taking the first principal component) is by far outweighed by the saving in CPU time: noise reduction on data sets of about 40 000 points takes less than 2 h on a DEC station 5000. In principle the time and memory needed grows linearly with the number of points to be processed (with a box-assisted neighbor-search method [13]). However, the flow data has to be embedded in a dimension at least higher by 2 to obtain satisfactory results. We did not attempt to clean the full 600 000 points but already the 220 000 points of the resampled set took about 30 h to be cleaned.

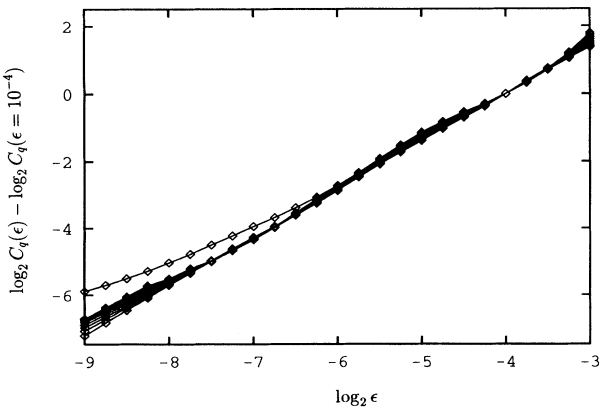


FIG. 12. Generalized correlation sums C_q for $1\leq q\leq 11$ (NMR-laser data, Poincaré section). Again the curves are shifted vertically by suitable normalization. The embedding dimension is 10; the delay is 1. Lower left-hand corner (top to bottom): $q=1$, $q=11,\ldots,2$; upper right-hand corner (top to bottom): $q=1,\ldots,11$.

V. CONCLUSIONS

The existing literature on noise reduction (major references include [2–8]) is to a large extent concerned with the development of algorithms. During this stage methods are mainly tested on artificial data sets where the dynamics is known and properties of the noise are controlled. Encouraged by the promising results of these tests we dedicated this paper to the next step, the application of nonlinear noise reduction to laboratory data. By straightforwardly running the algorithm given in [7] using the practical guidelines given there, we succeeded in removing most of the measurement error on two high-quality data sets. We found evidence that these sets can indeed be consistently separated into a low-dimensional deterministic part plus measurement noise.

These examples, together with earlier applications (e.g., in [2,4,8,23]), show that nonlinear noise reduction is becoming an indispensable tool for analyzing data from nonlinear dynamical systems. After more experience has been gained on data of comparable quality we can hope to extend the range of applicability to less clear cases, such as data with dynamical or parametrical noise, higher-dimensional data, etc.

ACKNOWLEDGMENTS

H.K., T.S., and I.H. are grateful to Peter Grassberger and Rainer Hegger for discussions about the data analysis. H. K. acknowledges support by the Deutsche Forschungsgemeinschaft, SFB 237. T.S. received a European Communities grant within the framework of the SCIENCE program, Contract No. B/SCI*-900557.

*Present address: Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark.

†Present address: Paul-Scherrer-Institut, CH-5232 Villingen, Switzerland.

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

[2] E. J. Kostelich and J. A. Yorke, Phys. Rev. A **38**, 1649 (1988); Physica D **41**, 183 (1990).

[3] J. D. Farmer and J. Sidorowich, in *Evolution, Learning and Cognition,* edited by Y. C. Lee (World Scientific, Singapore, 1988); Physica D **47**, 373 (1991); S. M. Hammel, Phys. Lett. A **148**, 421 (1990).

[4] T. Schreiber and P. Grassberger, Phys. Lett. A **160**, 411 (1991).

[5] T. Sauer, Physica D **58**, 193 (1992).

[6] R. Cawley and G.-H. Hsu, Phys. Rev. A **46**, 357 (1992); Phys. Lett. A **166**, 188 (1992).

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

[8] T. Schreiber, Phys. Rev. E **47**, 2401 (1993).

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

[10] P. Bösiger, E. Brun, and D. Meier, Phys. Rev. Lett. **38**, 602 (1977); D. Meier, R. Holzner, B. Derighetti, and E. Brun, in *Evolution of Order and Chaos,* edited by H. Haken (Springer, Berlin, 1982); E. Brun, B. Derighetti, R. Holzner, and D. Meier, Helv. Phys. Acta **56**, 825 (1983); E. Brun, B. Derighetti, D. Meier, R. Holzner, and M. Ravani, J. Opt. Soc. Am. B **2**, 156 (1985); G. Broggi, B. Derighetti, M. Ravani, and R. Badii, Phys. Rev. A **39**, 434 (1989); L. Flepp, Ph.D. thesis, University of Zürich, 1991 (unpublished).

[11] C. Grebogi, S. M. Hammel, J. A. Yorke, and T. Sauer, Phys. Rev. Lett. **65**, 1527 (1990).

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

[13] T. Schreiber, in *Predicting the Future and Understanding the Past,* edited by A. S. Weigend and N. A. Gershenfeld, Proc. SFI Studies in the Science of Complexity (Addison-Wesley, Reading, MA, 1993) Vol. XVII.

[14] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).

[15] Th. Buzug, T. Reimers, and G. Pfister, Europhys. Lett. **13**, 605 (1990); Th. Buzug and G. Pfister, Phys. Rev. A **45**, 7073 (1992); Physica D **58**, 127 (1992).

[16] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).

[17] J. P. Eckmann, S. O. Kamphost, D. Ruelle, and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986).

[18] M. Sano and Y. Sawada, Phys. Rev. Lett. **55**, 1082 (1985).

[19] Th. Buzug, J. v. Stamm, and G. Pfister, Phys. Rev. E **47**, 1054 (1993).

[20] G. Pfister, Th. Buzug, and N. Enge, Physica D **58**, 441 (1992).

[21] G. Pfister, A. Schulz, and B. Lensch, Eur. J. Mech. B **10**, 247 (1991).

[22] Th. Buzug, J. v. Stamm, and G. Pfister, Physica A **191**, 559 (1992).

[23] H. Kantz, in *Predicting the Future and Understanding the Past* (Ref. [13]).

[24] R. Vautard and M. Ghil, Physica D **35**, 395 (1989); R. Vautard, P. Yiou, and M. Ghil, *ibid.* **58**, 95 (1992).

[25] D. Broomhead and G. P. King, Physica D **20**, 217 (1986).

[26] L. Flepp, R. Holzner, E. Brun, M. Finardi, and R. Badii, Phys Rev. Lett. **67**, 2244 (1991).

[27] M. Finardi, L. Flepp, J. Parisi, R. Holzner, R. Badii, and E. Brun, Phys. Rev. Lett. **68**, 2989 (1992).

[28] N. B. Tufillaro, J. Holzner, L. Flepp, E. Brun, M. Finardi, and R. Badii, Phys. Rev. A **44**, R4786 (1991).