Largest Lyapunov-exponent estimation and selective prediction by means of simplex forecast algorithms

Rudolf M. Dünki
Computer Assisted Physics Group, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
(Received 20 January 2000)

Limited predictability is one of the remarkable features of deterministic chaos and this feature may be quantized in terms of Lyapunov exponents. Accordingly, Lyapunov-exponent estimates may be expected to follow in a natural way from forecast algorithms. Exploring this idea, we propose a method estimating the largest Lyapunov exponent from a time series which uses the behavior of so-called simplex forecasts. The method considers the estimation of properties of the distribution of local simplex expansion coefficients. These are also used for the definition of error bars for the Lyapunov-exponent estimates and allows for selective forecasts with improved prediction accuracy. We demonstrate these concepts on standard test examples and three realistic applications to time series concerning largest Lyapunov-exponent estimation of an experimentally obtained hyperchaotic NMR signal, brain state differentiation, and stock-market prediction.

PACS number(s): 05.45.Tp, 07.05.Kf, 89.90.+n, 87.10.+e

I. INTRODUCTION

Chaotic dynamics has been investigated in a broad spectrum of systems, e.g., in astrophysics, meteorology, chemistry, biology, medicine, electronics, and finance. Despite this frequent occurrence, the analysis of a time series of apparently chaotic data is still not a trivial task. For example, an experimentalist searching for properties like topological invariants is often confronted with an irregularly oscillating time series. From such a series it is not always easy to deduce quantities indicative of chaos. In principle, there are two different types of procedures available for such an analysis. The first one tries to assess whether a fractal dimension can be ascribed to the underlying system. The most popular of these approaches is probably the so-called D_2 algorithm of Grassberger and Procaccia [1]. This approach, however, is somewhat ambiguous for two reasons. (i) Under certain conditions, colored noise can also mimic an apparent fractal structure [2]. (ii) In the overwhelming number of cases, fractality is a sign of a chaotic system, but theory forbids neither fractality combined with nonchaoticity [3,4] nor nonfractality with chaoticity (see, e.g., the logistic equation at $a=4$ or the Mackey glass system at $\tau=40$ [5]). The second type of procedures is closer to the essence of chaos itself: It tries to assess exponential separation of nearby trajectories, thus directly determining the system's predictability. Measures like K entropy [6] and Lyapunov exponent (LE) estimates [7–10] belong to these indicators. Having assessed at least the largest Lyapunov exponent to be positive one immediately knows that there is a positive K entropy and thus chaos.

There exist several algorithms for assessing Lyapunov exponents: Some for the estimation of the largest Lyapunov exponent (LLE) [7,11], or all positive LE [12], or all-positive and -negative LE [9], and even some for estimating local expansion spectra [10]. These algorithms, however, require one to select an evolution time for the nearby states which may introduce ambiguity. The role which the different spatial directions play is another common feature of these algorithms: The assessment of LE's usually requires frequent and specific changes of the reference system, e.g., through reorthogonalization or appropriate angular adaptions to a reference trajectory (though not strictly necessary [13]). By angular adaptions we mean the necessity of restricting the choice of a piece of a nearby trajectory to cases where the spatial orientation with respect to the reference trajectory is similar to the orientation of the preceeding piece [7].

In contrast to these restrictions, we propose a method based on the class of so-called simplex-forecast algorithms [14,15] which does not rely upon such specific changes in two respects: (1) We do not consider angular adaptions for the separation of trajectories. This is because our algorithm is based on the onset of dominance of the largest Lyapunov exponent and, at this onset, the separation of nearby trajectories is governed by this exponent. (2) Our treatment leads to a measure providing a suitable evolution time at which this onset of dominance sets in.

An algorithmic outline of the method is found in Sec. II. This is intended as a quick reference for the reader who wishes to skip the Methods section at a first read through. We formally deduce the framework in the Methods section (Sec. III) where we also present a method for calculating approximate error bounds. This method is demonstrated on an artificial didactic example in Sec. IV and then explored with several test systems in Sec. V. We then apply our concepts in Sec. VI onto a hyperchaotic NMR Raser system [16] with delayed feedback [17]. In Sec. VI we also show further and qualitative aspects to be potentially useful in the context of rather soft data (medical and/or financial). The usage of the distribution of local simplex expansion coefficients and the concept of selective prediction are applied. The latter allows the selection of potentially good candidates for forecasts and the rejection of potentially bad candidates, thereby leading to fewer forecasts with higher accuracy for the remaining ones.

In brief, the advantages of our method may be stated as follows. (1) It is a spin-off gaining more information from a class of existing forecast algorithms. (2) It allows for the definition of error bars. (3) It improves forecast performance compared to the original algorithm.

II. OUTLINE OF THE METHODS

Simplex forecasting as originally suggested by Sugihara and May [14] or improved variants thereof (e.g., the so-called local hyperplane approximation (LHA) approach [15]) are well known examples of those nonlinear forecasting algorithms that try to approximate an unknown part of a trajectory through the construction of a local linear map from states in the neighborhood. Consider such a trajectory in a m-dimensional phase space and a point $X_0^k(t_k)$ on it which is hit at time t_k $(k=1,\ldots,n)$. A simplex enclosing this reference point then consists of $m+1$ points in such a manner that the vectors between the reference point and the simplex points contain components in all m directions. Hence the technique requires constructing a suitable simplex around this reference point with the help of nearby points obtained from earlier observations. The forecast is then simply constructed from the evolution of the simplex, i.e., the image of the simplex one forecast-time-unit T later.

Let $X_0^k(t_k)$ denote the reference point and $X_i^k(t_k)$ its ith simplex neighbor. Then the ith simplex vector a_i^k is written as $X_i^k(t_k)-X_0^k(t_k)$, $1\leqslant i\leqslant m+1$. Resetting the time t to $t-t_k$ and omitting the index k for the moment, we may rewrite the reference point in the form $X_0=\Sigma_i X_i(0)$, i.e., as the (weighted) sum of its simplex neighbors. The forecast at time T will then be constructed from the evolved simplex neighbors: $X_T=\Sigma_i X_i(T)$.

It will be shown in Sec. III that this quantity may be used for the largest Lyapunov-exponent estimation together with an error bar. Anticipating some results from the next sections, an algorithmic scheme of the important steps in this procedure may be given as follows.

(1) From the time series under consideration, we select a population of reference points and construct simplices X_0 around each such reference point. Let the simplices evolve into X_T for a series of forecast times T.

(2) Build the basic quantity, namely, the correlations ρ_{TT} [cf. Eq. (2.10) below],

$$\rho_{TT}=\rho\left(\sum \log[|X_i(T)-X_T|/|X_i(0)-X_0|],|\tilde{X}_T-X_T|\right), \tag{2.1}$$

where $\tilde{X}_T$ stands for the true outcome.

(3) Determine the first maximum of ρ_{TT} with respect to T.

(4) Choose this time and build the corresponding largest Lyapunov exponent as [Eq. (3.20)]

$$\lambda_1^{\text{app}}=\frac{\mu\left(\sum_i \log[|X_i(T+\Delta T)-X_{T+\Delta T}|]-\sum_i \log[|X_i(T)-X_T|]\right)}{\Delta T}, \tag{2.2}$$

where μ denotes the mean.

(5) From the population of simplices build the first derivative with respect to T

$$d\rho_{TT}:=\frac{\Delta\rho_{TT}}{\Delta T}, \tag{2.3}$$

decompose it into its four basic terms [cf. Eq. (3.13)] and use them to calculate the error term (3.18b) which is shown to contain—in the statistical sense—the error $\Delta\lambda$ (see below).

(6) To quantify the statistical assessment, choose a confidence level α and determine the associated normalized deviation $\langle|\tilde{X}_T-X_T|_k\rangle$ from the zero level such that $P(\langle|\tilde{X}_T-X_T|_k\rangle)\leqslant 1-\alpha$.

(7) The error term together with this deviation allow calculating the error $\Delta\lambda$ corresponding to this level α [cf. Eqs. (3.24) and (3.25) below].

III. DERIVATION OF THE METHODS

As outlined above, a simplex forecast is obtained from the evolution of the simplex around the reference point: $\Sigma_i X_i(0)\rightarrow\Sigma_i X_i(T)$. The shape of the simplex thereby undergoes a change (cf. Fig. 1) and an easily obtained measure of this change is the average simplex' expansion, i.e., the averaged logarithmic expansion of the single nearby neighbors localized around X_0,

$$\sum_i \log[|X_i(T)-X_T|/|X_i(0)-X_0|]. \tag{3.1}$$

Similarly, we may estimate the difference δ between the forecast and the true outcome $\tilde{X}_T$,

$$\delta_T=\left|\tilde{X}_T-\sum_i X_i(T)\right|. \tag{3.2}$$

For dynamic systems, this may be thought of as the evolution (expansion) of an initial tiny, but unavoidable difference δ_0,

$$\delta_0=\left|\tilde{X}_0-\sum_i X_i(0)\right|\equiv|X_0-\tilde{X}_0|. \tag{3.3}$$

In terms of dynamic systems theory, we may linearize the evolution of this difference and write (up to phase factors [18])

$$\delta_T=\delta_0\cdot\left|\sum_j b_j(T)e_j\exp(\lambda_j T)\right|,\quad j=1,\ldots,m \tag{3.4}$$

where the e_i denote the unit vectors collinear with the eigenvectors of this linearized dynamics and $b_i(T)$ are the amplitudes,

$$b_j(T)=(e_j,\delta_0^v). \tag{3.5}$$

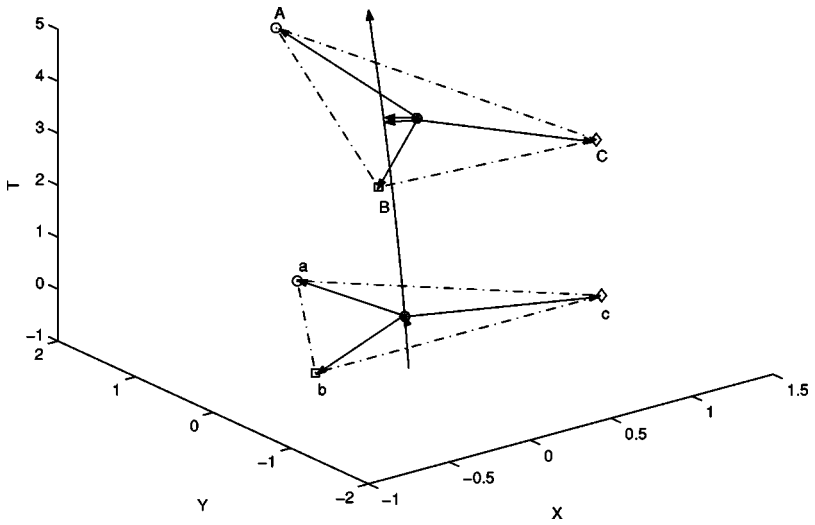


FIG. 1. Scheme of the evolution of small differences in two dimensions. An initial tiny difference δ_0 (filled triangle) at $T=0$ between a trajectory (solid line) and X_0 ($\otimes$) grows at $T=4$ to δ_T (double arrow) between the two. Similarly, the three simplex points (a,b,c) are mapped to their new locations (A, B, C) and the difference between the three points and X_0 or X_T, respectively, expands (arrows).

Here δ_0^v represents the difference vector $\tilde{X}_0-\Sigma_i X_i(0)$.

We now assume the difference between any of the $X_i(0)$ and X_0 to be small and analogously set,

$$a_{ij}(T)=(e_j,a_i)=(e_j,X_i(0)-X_0), \tag{3.6}$$

hence similarly

$$X_i(T)-X_T\simeq\sum_j a_{ij}(T)e_j\exp(\lambda_j T) \tag{3.7}$$

and therefore

$$\sum_i \log[|X_i(T)-X_T|/|X_i(0)-X_0|] = \sum_i \log\left(\left|\sum_j a_{ij}(T)e_j\exp(\lambda_j T)\right|\right) - \log\left(\left|\sum_j a_{ij}(0)e_j\right|\right). \tag{3.8}$$

In an intermediate time range, the evolution of the small scale dynamics is described by the largest Lyapunov exponent (LLE) λ_1 [18], namely, when dominance of the LLE has set in. In that case, the inner sum of Eq. (3.8) becomes

$$\left|\sum_j a_{ij}(t)e_j\exp(\lambda_j t)\right| = \exp(\lambda_1 t)\left(\left|a_{i1}(t)+\sum_{j=2} a_{ij}(t)e_j\exp[(\lambda_j-\lambda_1)]t\right|\right) \simeq \exp(\lambda_1 t)|a_{i1}(t)|. \tag{3.9}$$

Hence local expansion coefficients may be used to estimate the largest Lyapunov exponent, if this time range is known.

Let us drop the formal distinction between particular single values $(X_i^k(T),X_T^k)$ and the set of all such local values $\{X_i^k(T),X_T^k\}$, $k=1,\ldots,n$ for the moment. We regard instead the former to be a representation drawn from the distribution of the latter [Fig. 2(a) displays an example of such a distribution, cf. Sec. IV] and we describe it in terms of standard statistical measures. This view is first used to calculate the correlation

$$\rho_{TT}=\rho\left(\sum \log[|X_i(T)-X_T|/|X_i(0)-X_0|], \left|\tilde{X}_T-\sum X_i(T)\right|\right). \tag{3.10}$$

The usefulness of this quantity will be demonstrated in Secs. III A–III C when we (i) in Sec. III A show this correlation to provide the appropriate time range searched for (i.e., the onset of dominance of the largest exponentially growing term) when it passes the maximum with respect to T; (ii) in Sec. III B derive important properties at the onset of dominance to extract the LLE λ_1 at this point; (iii) in Sec. III C make use of these same properties to calculate an approximate error estimate $\sigma(\lambda_1)$. So far, we introduced the basic treatment of simplex expansion and outlined the ideas of our methods. We now turn to Sec. III A and focus on the properties of ρ_{TT}.

A. Derivation of the appropriate time range and maximum passage of ρ_{TT}

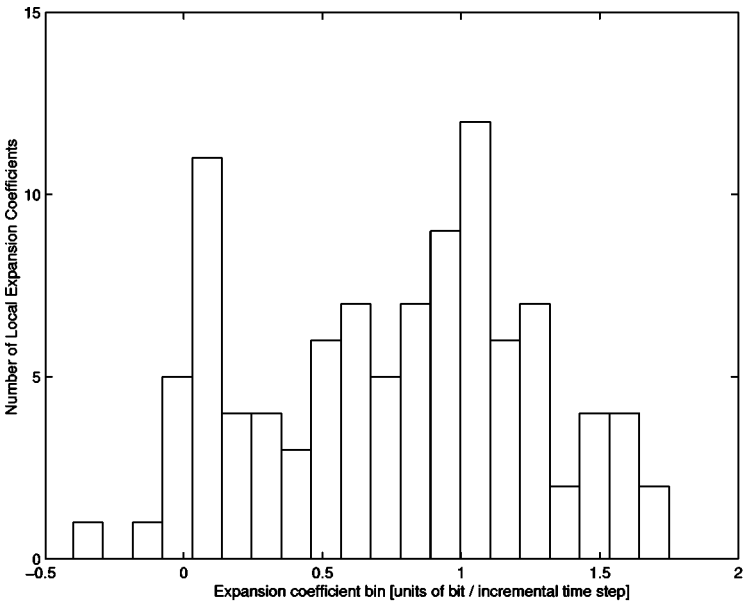
The essential point is our claim that an appropriate time range may be found when the correlation ρ_{TT} (3.10) passes the maximum with T increasing. The following conventions will be useful in this context:

(i) $\Delta X_T^i\equiv|X_i(T)-X_T|$,

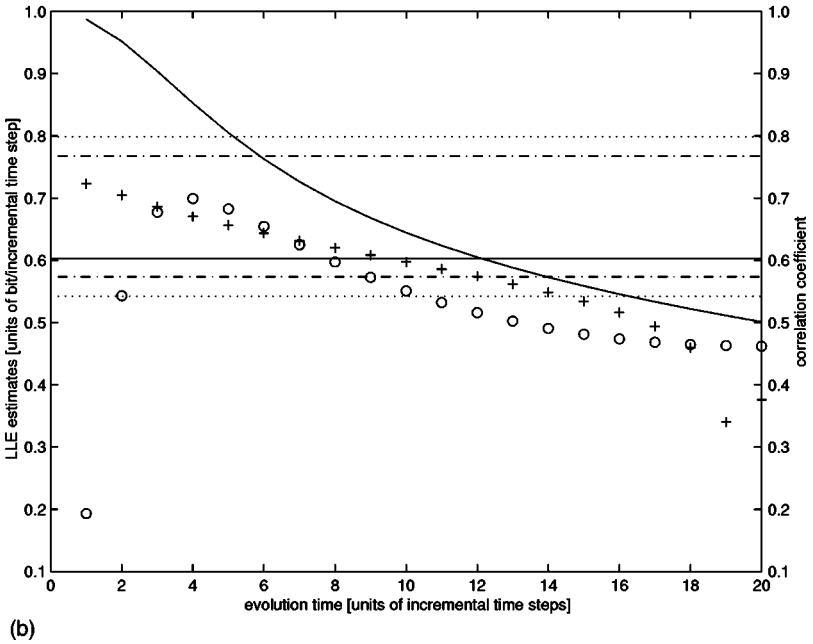
(ii) $\Delta X_{T0}^i\equiv[|X_i(T)-X_T|]/[|X_i(0)-X_0|]$,

(iii) $\log\Delta X_{T0}\equiv\sum_i\log(\Delta X_{T0}^i)$,

(iv) $\Delta X_{Ts}\equiv\left|\tilde{X}_T-\sum X_i(T)\right|$



(a)



(b)

FIG. 2. (a) Distribution (histogram) of the didactic examples' local expansions at $T=4$ (units of incremental time steps). (b) Various estimates of the didactic example vs evolution time. Straight line: λ; plus sign: λ_1^{app}; dotted line: 99% interval for λ_1^{app} at $T=4$; dash-dotted line: 95% interval (all in units of bit/incremental time step as shown on the left scale); circles: correlation according to Eq. (3.10); solid curve: correlation of the exponentially expanding terms used in Eq. (3.14) (right scale). Note the negative slope of the dormer after $T=4$ and that the two correlations become almost proportional after this point.

$$\text{(v)}\quad a_{\mathrm{res}} \equiv \frac{\sum_i \left| a_{i1}(T) + \sum_{j=2} a_{ij}(T) e_j \exp([\lambda_j - \lambda_1)T] \right|}{\left| \sum_j a_{ij}(0) e_j \right|},$$

$$\text{(vi)}\quad df \equiv \partial f/\partial t = \partial f/\partial T,$$

$$\text{(vii)}\quad d^2 f \equiv \partial^2 f/\partial t^2 = \partial^2 f/\partial T^2,$$

$$\text{(viii)}\quad \langle x \rangle \equiv [x-\mu(x)]/\sigma(x),$$

where μ denotes the mean and σ the standard deviation, i.e., $\langle x \rangle$ is a normalization to a (0,1) distribution. We note that the derivatives d with respect to T are, in practice, only assessible from the time series, i.e., in the numerical sense. The ranges of the subscripts i,j,k are $i=1,\ldots,m+1$; $j=1,\ldots,m$; $k=1,\ldots,n$ if not indicated otherwise, and we write a correlation coefficient ρ as

$$\rho(a,b) = \langle a \rangle \cdot \langle b \rangle, \tag{3.11}$$

where the $\cdot$ denotes componentwise multiplication first, then summing up over all n terms and, finally, dividing the result by n. Using these conventions we may rewrite Eq. (3.10) as

$$\rho_{TT} = \langle \log \Delta X_{T0} \rangle \cdot \langle \Delta X_{Ts} \rangle \tag{3.12}$$

and assess its first derivative

$$d(\langle \log \Delta X_{T0} \rangle \cdot \langle \Delta X_{Ts} \rangle = \langle \Delta X_{Ts} \rangle \cdot \left\{ \frac{(d \log \Delta X_{T0}) - d\mu(\log \Delta X_{T0})}{\sigma(\log \Delta X_{T0})}, \right. \tag{3.13a}$$

$$-\langle \log \Delta X_{T0}\rangle \frac{d\sigma(\log \Delta X_{T0})}{\sigma(\log \Delta X_{T0})}\Bigg\} \tag{3.13b}$$

$$+\langle \log \Delta_{T0})\rangle \cdot \Bigg\{ \frac{d(\Delta X_{Ts}) - d\mu(\Delta X_{Ts})}{d\sigma(\Delta X_{Ts})} \tag{3.13c}$$

$$-\langle \Delta X_{Ts}\rangle \Bigg\} \frac{d\sigma(\Delta X_{Ts})}{\sigma(\Delta X_{Ts})}, \tag{3.13d}$$

where d denotes the derivative with respect to T. Recalling our statement above, this has to be explored around the maximum of ρ_{TT}. Equation (3.13) reduces in the case of dominance of the exponentially expanding terms to

$$\begin{aligned} d\rho[\lambda_1 T, \exp(\lambda_1 T)] &= \{\langle \log[\exp(\lambda_1 T)]\rangle \cdot d\langle \exp(\lambda_1 T)\rangle\} \\ &= \langle \lambda_1\rangle \cdot d[\langle \exp(\lambda_1 T)\rangle] < 0 \end{aligned} \tag{3.14}$$

because the variance term [i.e., the equivalent to Eq. (3.13d)] dominates $d[\langle \exp(\lambda_1 T)\rangle]$. Accordingly, the slope becomes negative at the passage of this point [cf. Fig. 2(b) from the example below for an illustration]. Passing the maximum with T increasing indeed means the onset of dominance, thus indicating the intermediate time range searched for.

B. Important properties when extracting λ_1 at the onset of dominance

To get additional insight into Eq. (3.13), let the temporal increment ΔT be small and denote $Z := T + \Delta T$. Upon dominance of the exponentially expanding term, one may set

$$\rho_{TZ} := \langle \log \Delta X_{T0}\rangle \cdot \langle \log \Delta X_{z0}\rangle \simeq \rho(\lambda_1 T, \lambda_1 Z) = O(1). \tag{3.15}$$

Hence in the context of correlation estimates, the X_Z may be regarded to be approximately equivalent to X_T. Using Eq. (3.13c) this leads to

$$\begin{aligned} \langle \log \Delta X_{T0}\rangle \cdot \frac{d(\Delta X_{Ts}) - d\mu(\Delta X_{Ts})}{\sigma(\Delta X_{Ts})} &\simeq \frac{\sigma(\Delta X_{Z_s})\rho_{zz} - \sigma(\Delta X_{Ts})\rho_{TT}}{\Delta T} \\ &\simeq \frac{d\sigma(\Delta X_{Ts})}{\sigma(\Delta X_{Ts})}\rho_{TT} + \frac{\sigma(\Delta X_{Z_s})}{\sigma(\Delta X_{Ts})} d^2\rho_{TT}\Delta T/2. \end{aligned} \tag{3.16}$$

The latter holds because $\rho_{ZZ} \simeq \rho_{TT} + d\rho_{TT}\Delta T + d^2\rho_{TT}(\Delta T)^2/2$ and $d\rho_{TT}$ is vanishing at the maximum (up to numerical inaccuracies). The impact of Eqs. (3.13c) and (3.13d) thus depends on the second derivative (i.e., the curvature) of the correlation (3.13) amplified with the variance term

$$\frac{d\sigma(\Delta X_{Z_s})}{\sigma(\Delta X_{Ts})} \sim 1 + \exp(\lambda_1 \Delta T).$$

This is because $\sigma(ax) = a\sigma(x)$ and one expects $\sigma(\Delta X_{Ts})$ to grow to the order of $\exp(\lambda)$. Hence Eq. (3.13c) is easily extractable from experimental data.

To assess the remaining parts, Eqs. (3.13a) and (3.13b), we keep in mind Eqs. (3.4) and (3.7). For T not too small and again using $\sigma(ax) = a\sigma(x)$, we may approximate Eq. (3.13b) as

$$\begin{aligned} \langle \log \Delta X_{T0}\rangle \cdot \langle \Delta X_{Ts}\rangle \frac{d\sigma(\log \Delta X_{T0})}{\sigma(\log \Delta X_{T0})} &\simeq -\rho(\log \Delta X_{T0}, \Delta X_{Ts})/T. \end{aligned} \tag{3.17}$$

Equation (3.13a) may be decomposed into two terms,

$$\langle \Delta X_{Ts}\rangle \cdot \frac{d[\log \Delta X_{T0} - \mu(\log \Delta X_{T0})]}{\sigma(\log \Delta X_{T0})}$$

$$= \frac{\sigma(\lambda_1)}{\sigma(\log \Delta X_{T0})} \langle \lambda_1\rangle \cdot \langle \Delta X_{Ts}\rangle \tag{3.18a}$$

$$+ \frac{\sigma(da_{\rm res}/a_{\rm res})}{\sigma(\log \Delta X_{T0})} \left\langle \frac{da_{\rm res}}{a_{\rm res}} \right\rangle \cdot \langle \Delta X_{Ts}\rangle. \tag{3.18b}$$

It is hardly possible to deduce Eq. (3.18) from an experimentally obtained time series for each value of T. However, upon dominance of the LLE one may deduce it at least approximately: The term (3.18a) may, in this case be expanded as

$$\frac{\sigma(\lambda_1)\langle \lambda_1\rangle}{\sigma(\log \Delta X_{T0})} \cdot \langle \Delta X_{Ts}\rangle$$

$$= \frac{\sigma(\lambda_1 T)/T}{\sigma(\log \Delta X_{T0})} \langle \lambda_1\rangle \cdot \langle \Delta X_{Ts}\rangle \tag{3.19a}$$

$$= \frac{\sigma(\lambda_1 T)/T}{\sigma(\log \Delta X_{T0})} \langle \lambda_1 T\rangle \cdot \langle \Delta X_{Ts}\rangle \tag{3.19b}$$

$$\sim \frac{\sigma(\lambda_1 T)/T}{\sigma(\log \Delta X_{T0})} \langle \log \Delta X_{T0}\rangle \cdot \langle \Delta X_{Ts}\rangle \tag{3.19c}$$

$$= O(\rho(\log \Delta X_{T0}, \Delta X_{Ts})/T). \tag{3.19d}$$

Strictly speaking, the calculation of Eq. (3.18b) would require knowledge of $\langle \lambda_1\rangle$, i.e., about the quanitity we are after. But Eqs. (3.19) suggest that this quantity is at least approximately retrievable from the time series without further knowledge about λ_1. Now let ρ_{TT} (3.10) reach its maximum. The first derivative (3.13) must (up to numerical inaccuracies, cf. below) become zero at this particular value of T and this implies that the sum over Eqs. (3.13b)–(3.13d) and (3.19) is outweighted by the term (3.18b). At first glance, the motivation to actually calculate these terms is not obvious. In practice, however, these quantities provide the basis for an approximate error estimation when the largest Lyapunov exponent is extracted at the onset of dominance:

The apparent exponent $\lambda_1^{\rm app}$ is assessed from Eq. (3.9) as

$$\lambda_1^{\rm app} = \frac{\mu(\log \Delta X_Z - \log \Delta X_T)}{\Delta T} \tag{3.20}$$

and the relation to the true exponent λ_1 may be written as

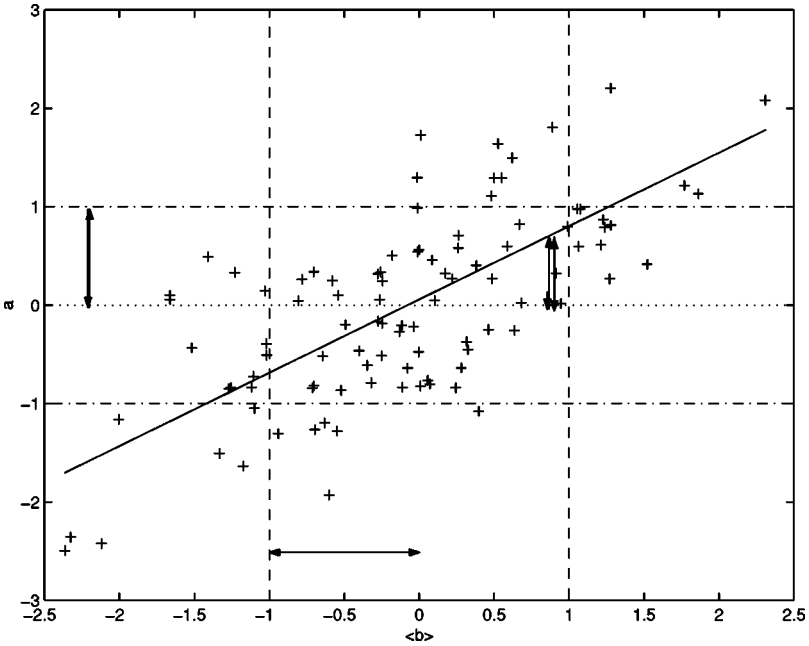


FIG. 3. Illustration of bivariate behavior. Standard deviations $\sigma(a)$: vertical arrow; $\sigma(\langle b\rangle_k)$: horizontal arrow; $\sigma_{\rm rel}$: double vertical arrow. Note the occurrence of $\langle b\rangle_k$ values crossing the zero line (dotted) when the linear regression curve (straight line) is only around $\sigma_{\rm rel}$ above this zero line.

$$\lambda_1^{\rm app}=\lambda_1+\mu_a\,, \tag{3.21}$$

where $\mu_a=\mu(da_{\rm res}/a_{\rm res})$. At the onset of dominance, the error $|\lambda_1-\lambda_1^{\rm app}|$ is thus implicitly contained in Eq. (3.18b). The error is not directly accessible, however, because of the normalization to a (0,1) distribution, but a related quantity is available:

Let $\sigma_a=\sigma(da_{\rm res})/(a_{\rm res})$ and $\rho_a=\rho(da_{\rm res}/a_{\rm res},X_{Ts})$. It then follows that Eq. (3.18b) provides the term $\rho_a\sigma_a/\sigma(\log\Delta X_{T0})$ rather than μ_a and all that remains is to gain statistical knowledge about μ_a when the quantity $\rho_a\sigma_a$ is known.

C. Calculation of approximate error bounds

To make use of the above-mentioned properties for the quantification of error bounds when $\rho_a\sigma_a$ is known, let us express the contribution of the kth individual value in the statistical sense by means of linear regression,

$$\left\langle\left\langle\frac{da_{\rm res}}{a_{\rm res}}\right\rangle\langle\Delta X_{Ts}\rangle\right\rangle_k\simeq\rho_a\sigma_a(\langle\Delta X_{Ts}\rangle_k)^2+\mu_a\cdot\langle\Delta X_{Ts}\rangle_k+N(0,\sigma_{\rm rel})\langle\Delta X_{Ts}\rangle_k\,. \tag{3.22}$$

This relates Eq. (3.18b) to the error $|\lambda_1-\lambda_1^{\rm app}|$ in terms of the regression line (3.22). Here the symbol $N(0,\sigma_{\rm rel})$ denotes the normal distribution with mean 0 and standard deviation $\sigma_{\rm rel}$. The latter may be regarded as describing the scatter around the regression line. This scatter can also be calculated by means of standard measures of linear regression:

$$\sigma_{\rm rel}=\sigma_a\sqrt{1-\rho_a^2} \tag{3.23}$$

(Fig. 3). We now may express the probability for $da_{\rm res}/a_{\rm res}$ to reach at least zero through $P(\langle\Delta X_{Ts}\rangle)$, namely, the probability of the occurrence of a particular value $\langle\Delta X_{Ts}\rangle_k$ such that

$$|\rho_a\sigma_a\cdot\langle\Delta X_{Ts}\rangle_k|\gtrsim|\mu_a+-\sigma_{\rm rel}| \tag{3.24}$$

with $|\mu_a+-\sigma_{\rm rel}|=\min(|\mu_a-\sigma_{\rm rel}|,|\mu_a+\sigma_{\rm rel}|)$. The assumptions entering in Eq. (3.24) are twofold: (a) for any given point $\langle\Delta X_{Ts}\rangle_k$, the probability of finding a nearby point $\langle\Delta X_{Ts}\rangle)_{k'}$ whose image scatters at least one standard deviation from the regression line may be set to 1; (b) this also applies when the absolute value $|\rho_a\sigma_a\langle\Delta X_{Ts}\rangle_k+\mu_a|$ is only one standard deviation away from zero (Fig. 3). Note that the term in absolute brackets is just the expectation value of the image of $\langle\Delta X_{Ts}\rangle_k$.

Conversely, one might regard the mean μ as the random variable instead of $\langle\Delta X_{Ts}\rangle$. $P(\langle\Delta X_{Ts}\rangle)$ then provides an estimate for the probability $P(|\mu|\geqslant|\mu_a|)$. This is the *a priori* probability for a certain mean μ to occur with value for μ_a when one fixes any hypothetical value $da_{\rm res}/a_{\rm res}$ to zero. From this, one does not get an exact error μ_a but, rather, a probability that it reaches a certain value.

The required quantity σ_a is also not known because Eq. (3.18b) contains the product $\sigma_a\cdot\rho_a$ only. To obtain a worst case guess, we fix σ_a, insert Eq. (3.23) in Eq. (3.24), and maximize with respect to ρ_a. In this way, one finds for ρ_a at a particular value $\langle\Delta X_{Ts}\rangle_k$,

$$\frac{-\rho_a}{\sqrt{1-\rho_a^2}}=\langle\Delta X_{Ts}\rangle_k\,. \tag{3.25}$$

Assuming normal distribution, we may set $P(\langle\Delta X_{Ts}\rangle)=0.95$ at $\langle\Delta X_{Ts}\rangle=1.96$, accordingly, $\rho_a(1.96)=0.89$. Using the Eqs. (3.23) and (3.24), the quantity $(1.96-1/0.89\sqrt{1-0.89^2})=1.45$ leads to

$$|\mu_a|_{95}\lessapprox 1.45\sigma_a\cdot\rho_a\,. \tag{3.26}$$

Similarly, we find for $P(\langle\Delta X_{Ts}\rangle)=0.99$ at $\langle\Delta X_{Ts}\rangle=2.34$, $\rho_a(2.34)=0.92$, and $(2.34-1/0.92\sqrt{1-0.92^2})=1.91$, leading to

$$\mu_a|_{99}\lessapprox 1.91\sigma_a\cdot\rho_a\,. \tag{3.27}$$

TABLE I. Individual contributions of the different equations to the error μ_a for the didactic example.

Quantity	(3.13b)	(3.13c)	(3.13d)	(3.19b)	(3.19d)	sum	$\sigma(\log \Delta X_{\mathrm{TO}})$	$\sigma_a \cdot \rho_a$
Value	−0.153	0.980	−1.058	0.182	0.175	0.048	1.392	0.067

Because $|\mu_a|_{95}$ and $|\mu_a|_{99}$ are our estimators of the 95% and 99% confidence intervals for μ_a, i.e., $P(|\mu_a| \leqslant |\mu_a|_\alpha) = \alpha$, we may (somewhat redundantly) define an approximate error $\sigma(\lambda_1^{\mathrm{app}})$ through

$$\sigma(\lambda_1^{\mathrm{app}}) \simeq (|\mu_a|_{95}/1.96 + |\mu_a|_{99}/2.34)/2 \tag{3.28}$$

because, for a standard error distribution, one gets $1.96\sigma(\lambda_1^{\mathrm{app}}) = |\mu_a|_{95}$ and $2.34\sigma(\lambda_1^{\mathrm{app}}) = |\mu_a|_{99}$.

IV. A DIDACTIC EXAMPLE

The above formalism may be best illustrated with a fictitious didactic example. We call it didactic because we simulate the properties of a dynamical system, but control certain parameters in a way to nicely illustrate our concept. In particular we may compare the exponentially expanding terms to the whole evolution of the expansion coefficients. Here we regard an evolution of nearby points of a fictitious two-dimensional system according to Eqs. (3.4) and (3.7),

$$\begin{aligned} \Delta X_{T0}^k &= \exp[-\lambda_1(t_k)T](a_1(T,t_k) \\ &\quad + a_2(T,t_k)\exp\{-[\lambda_2(t_k)-\lambda_1(t_k)]T\}) \\ \Delta X_{Ts}^k &= \exp[-\lambda_1(t_k)T](b_1(T,t_k) \\ &\quad + b_2(T,t_k)\exp\{-[\lambda_2(t_k)-\lambda_1(t_k)]T\}) \end{aligned} \tag{4.1}$$

where the incremental time step $t_{k+1} - t_k$ is arbitrariliy set to 1. This system allows one to mimic the temporal variation of the respective terms $a_i(t_k,T), b_i(t_k,T)$ with a finite dynamics with random coefficients,

$$\begin{aligned} a_i(t_k,T) &= a0_i(t_k) + a1_i(t_k)\sin(\omega T), \\ b_i(t_k,T) &= b0_i(t_k) + b1_i(t_k)\sin(\omega T), \end{aligned} \tag{4.2}$$

i.e., a set of n randomly choosen amplitudes $a0_i(t_k), b0_i(t_k)$ superimposed by another set of randomly choosen amplitudes $a1_i(t_k), b1_i(t_k)$ with a sinusoidal variation of a period of 40 incremental time steps. The motivation for this particular sinusoidal form was that (a) in the statistical sense, the first derivative passes a true zero at a quarter period and one has an exact estimate at this point to compare with and (b) a simultaneous zero of all $a_i(t_k), b_i(t_k)$ for a particular choice of t_k and T is avoided. The distribution of the local Lyapunov exponents $\lambda_1(t_k)$ and $\lambda_2(t_k)$ corresponding to the two directions were set to $(\mu,\sigma) = 10.60302,\ 033371)$ and $(-11.203, 6.3805)$, respectively. This is comparable to the values of the Henon system and the true estimate would be $\lambda = 0.603$ bit/incremental time step.

Applying our technique, we find the passage of the maximum after a time corresponding to four lags [Fig. 1(b)]. In this example, the onset of dominance of the exponentially expanding terms obviously sets in earlier than the true zero of the first derivative of $a_i(t), b_i(t)$, occuring at lag 10 (quarter period). This is displayed in Fig. 2(b) by the negative slope and the apparent proportionality between these exponential terms only and the correlation ρ_{TT} [Eq. (3.10)]. Our LLE estimate may thus (expectedly) differ somewhat from the true value and we find indeed $\lambda_1^{\mathrm{app}} = 0.670$ at $T = 4$, i.e., a difference of 0.0673.

This deviation, however, must be compared with the results of our error bound considerations: In this example we find the terms [Eqs. (3.13b), (3.13c), (3.13d), (3.19b), cf. Table I] to sum up to 0.048. This leads to $\sigma_a \rho_a = 0.067$ and to intervals $|\mu_a|_{95} \simeq 0.097$ and $|\mu_a|_{99} \simeq 0.128$, respectively. The latter is about twice the true error of 0.0673 which in turn is comparable with our $\sigma(\lambda_1^{\mathrm{app}}) = 0.0520$ [Eq. (3.28)]. These findings conform well to our analysis and the corresponding error bounds reveal the differences realistically.

V. CHAOTIC TEST EXAMPLES

Testing the formalism of Sec. II with data from well known model systems is another important step because such tests allow searching for systematic deviations between assessed and true estimates. Our testing was done with time series of the standard systems Henon, Lorenz, and Roessler (ordinary and hyperchaos). The corresponding time series were generated with standard parameter values from the literature [7,6] and a fourth-order Runge-Kutta integration scheme with an adaptive time step was used for integration. Our results were found through evaluating the behavior of 2000 simplices for each example. The largest LE estimate resulting from our method is then compared with the corresponding value from the literature.

We note that slight complications may occur when estimating the error from real-time series: In practice, the derivatives have to be evaluated numerically from the time series. This may lead to some mismatch because (1) Eq. (3.19b) can only be assessed approximately (cf. above). Equation (3.19d) was given as a rather intuitive example of such an approximation and is not intended to enter in the analysis. But similar considerations as those leading to Eq. (3.19d) suggest approximations such as Eq. (3.19b) is approximately

$$\rho_{TT}/\sigma(\log \Delta X_{T0}) d\sigma(\Delta X_{T0}), \tag{5.1a}$$

$$d[\rho_{TT}\sigma(\log \Delta X_{T0})]/\sigma(\log \Delta X_{T0}), \tag{5.1b}$$

$$d[\rho_{TT}\sigma(\log \Delta X_{T0})]/\sigma(\log \Delta X_{T0}) - d\rho_{TZ}, \tag{5.1c}$$

and these served as estimates for our test systems. (2) Equation (3.13) does not exactly cancel to zero but, rather, to a residual error term E_r (cf. above). We used a combination of equations (5.1a), (5.1b), and (5.1c), i.e., $\{[(5.1a)+(5.1b)]/2 + (5.1c)\}/2$ to deal with these imperfections and the differences considered in our error estimation were assumed to to sum up to E_r rather than to cancel out.

Table II displays the results obtained from these test systems: The differences between our LLE estimates and the

TABLE II. Largest Lyapunov exponent of test systems: λ (reference values from literature), λ_1^{app} [LLE derived from Eq. (3.20)], difference to reference values, estimated 95% [Eq. (3.26)], and 99% [Eq. (3.27)] interval, reference value, difference to reference value and $\sigma(\lambda_1^{app})$ [Eq. (3.28)] for several test systems. Units are bit/sec, except for the Henon system which are given in bit/iteration. Literature values are two times overestimated and three times underestimated, pointing to the absence of systematic derivations.

System	λ	λ_1^{app}	$\lvert\lambda_1^{app}-\lambda\rvert$	$\lvert\mu_a\rvert_{95}$	$\lvert\mu_a\rvert_{99}$	$\sigma(\lambda_1^{app})$
Henon [7]	0.607	0.583	0.024	0.104	0.141	0.056
Roessler (hyperchaos) [7]	0.16	0.172	0.012	0.035	0.048	0.019
Roessler [7]	0.13	0.103	0.027	0.029	0.040	0.016
Lorenz [6]	1.298	1.288	0.010	0.522	0.705	0.281
Lorenz [7]	2.16	2.286	0.126	0.497	0.671	0.267

reference values are well within our approximate error bounds. In addition, no systematic underestimation or overestimation seems to occur. We therefore regard our method to correctly estimate largest Lyapunov exponents. From Table II one may even get the impression that our approximate error bounds are somewhat too pessimistic (which is not excluded by our assumptions). To test for this possibility, we considered the ratios $[|\lambda_1^{app}-\lambda|/\sigma(\lambda_1^{app})]^2$. If our error estimates are correct, one expects these ratios to follow a χ^2 distribution. Accordingly, the occurrence of an overly small χ^2 value would be unlikely, thus indicating a systematic effect. In our case we find $\chi^2=3.60$, which seems to be somewhat small at a first glance. The random chance occurrence of an even smaller value is not unlikely $[P(\chi^2\langle 3.60)\rangle 39\%, df=5]$, however, and we may conclude that the result of our tests does not indicate systematic error overestimation.

We note that such an overestimation, if existent, would only mean that our error estimates should be regarded as upper rather than true error bounds. The estimate would nevertheless remain a valuable limit for error bounds in data analysis because, in that case, the true value would confidently lie between these bounds. As a last note, we mention that in most cases E_r was comparable to the estimated error and may thus provide a quick error estimate.

VI. REAL WORLD APPLICATIONS

A. NMR Raser

In a nonstandard example of chaotic dynamics we considered the time series of a NMR raser in the hyperchaotic region. In brief, this system consists of spins of Al atoms in a ruby crystal which are positioned in a radio-frequency field inside a cavity. Chaos is induced through temporally varying the quality factor of the resonant structure. The device provides the voltage induced by the rotating magnetization across the coil of the LC resonator as output signal. A detailed technical description of this NMR raser is given elsewhere in the literature [16,17,19].

Our study refers to the signal obtained under hyperchaotic rasing conditions when the dimensionality (D_2) is between 3 and 4 [20]. The signal was embedded in an appropriate m-dimensional space because only one component was available. This was achieved by means of delay-time coordinates [21] as usual. The embedding procedure introduces two additional free parameters, namely, the delay time τ and the embedding dimension m. For $m>D_2$, one expects to find a range where λ_1^{app} is independent of τ and m, however, because the LLE is an invariant of the system dynamics and may not depend on the parameter choice. The common value for λ_1^{app} within this range is regarded as the estimate of the

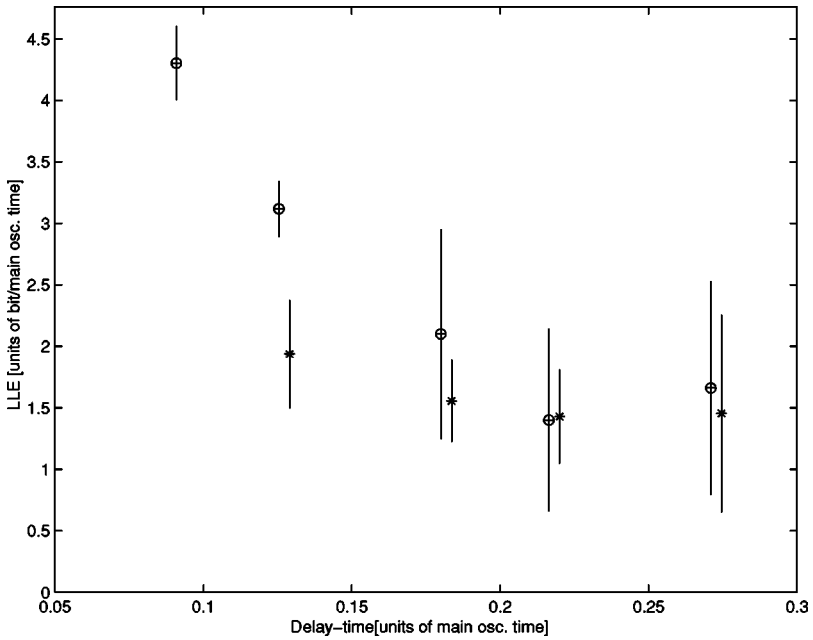


FIG. 4. Estimates of the LLE (units of bit per main oscillation time) of the NMR raser for several embedding conditions. Note the convergence towards a common value for embedding dimensions four (encircled crosses) and five (stars) and delay times between 0.15 and 0.3 (units of main oscillation time).

TABLE III. Largest Lyapunov exponent λ_1^{app} [Eq. (3.20)], estimated 95% interval $|\mu_\alpha|_{95}$ [Eq. (3.26)] relative to the main oscillation for several test systems and NMR raser. Units are bit/main osc.

Quantity	Roessler	Lorenz [6]	Lorenz [7]	Roessler hyperchaos	NMR raser
λ_1^{app}	0.648	1.020	1.087	1.178	1.523
$\|\mu_a\|_{95}$	0.182	0.413	0.236	0.240	0.581
Main osc. (Hz)	0.159	1.263	2.103	0.146	45.45

system's LLE. Our example displays indeed such a range (Fig. 4) and the corresponding LLE becomes 1.523 bit/main osc. We refer here to the somewhat unusual time units ''main oscillation'' because one might regard this time unit to represent the system's natural time base. The time span of one main oscillation may be assessed, e.g., through spectral analysis. A further advantage of this natural base is that it allows for comparison with the test system's outcomes (Table III). In this regard, our LLE estimate is found to be comparable with the values of the standard test systems

B. Distribution of local expansion coefficients and selective prediction

Section III made use of the distribution of simplex expansion coefficients to search for a suitable forecast time for the estimation of a time series' largest Lyapunov exponent. The important quantity considered was the correlation between local simplex-expansion coefficients and the goodness of the corresponding local forecast. Throughout the preceding sections, this has been shown to lead to suitable estimates of the largest Lyapunov exponent. In this section we present briefly two additional examples where (a) the distribution of local simplex expansions itself (distinguishing biomedical signals schizo control, Fig. 5) and (b) the correlation between local simplex expansion and the quality of the corresponding local forecast (second derivatives of monthly recordings of an economic time series) might be used as tools characterizing time series. The latter is shown to allow for some kind of selective forecasting (Fig. 6).

The former example (a) is drawn from a larger electroencephalogram (EEG) study comparing schizophrenic patients under different mental tasks with a control group matched for age, sex, education, qualification, and handedness [22,23]. The simultaneous EEG from four frontal electrodes was selected for our example. This particular choice was made because schizophrenia-specific effects may be expected in the frontal area. The resulting time series is then regarded as the time evolution of an unknown system in a four-dimensional phase space. Our simplex-expansion algorithm was applied to three such time series of 30 sec duration, each to obtain an averaged distribution of local expansion coefficients. This was done for a person belonging to the patient and also for one from the control group and the two distributions were compared. As is illustrated in Fig. 5, the distributions of the two persons seem to differ. Without putting too much emphasis on the psycho-physiological meaning of this result, it shows that the distribution of local expansion coefficients may contain characteristic features already assessible on a rather qualitative level.

As a last example (b), an analysis of second differences on monthly recordings of an economic time series, namely, the Dow Jones stock-market index between 1900 and 1995, was performed. Differencing is required because the raw series must be regarded as nonstationary [24]. After differencing twice, the remaining time series showed an autocorrelation of about -0.45 at lag one and no significant autocorrelations at higher lags. One may therefore deduce that standard linear autoregressive forecasts should allow for

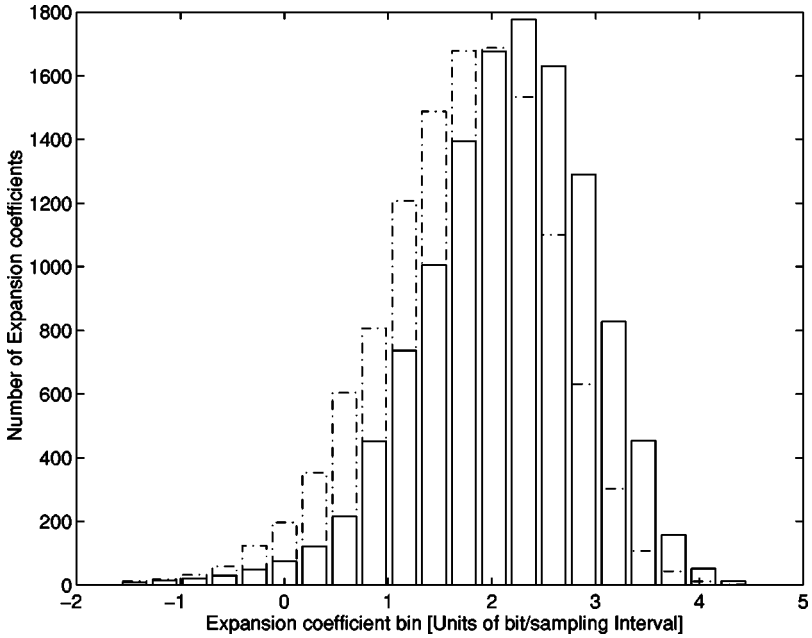


FIG. 5. Distribution (histogram) of local expansion coefficients for a schizophrenic patient (—·) and control (—). When compared to the control, the former seems to be more left sided.

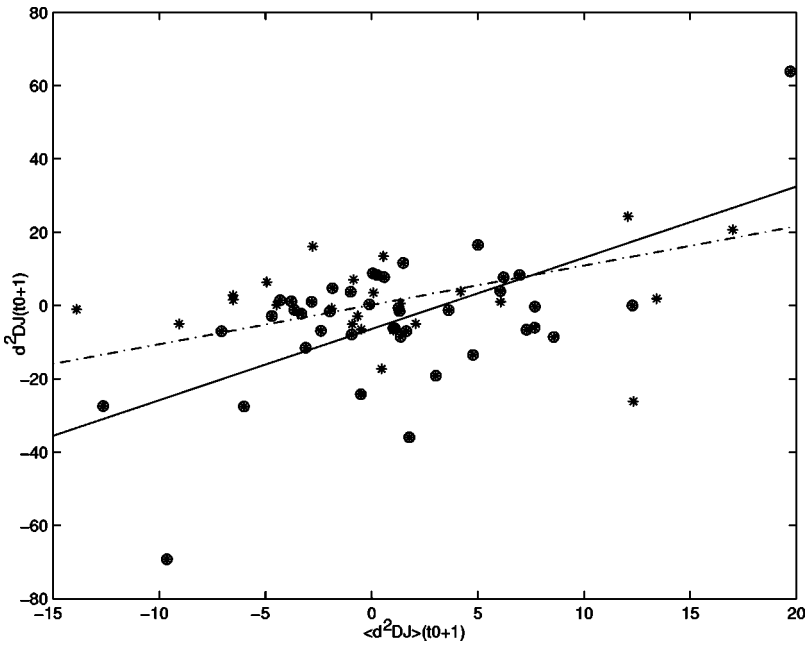


FIG. 6. One-step-ahead forecast of second differences of the Dow Jones index $\langle d^2D_J\rangle(t_0+1)$ vs $d^2\mathrm{DJ}(t_0+1)$, the true value of the second differences of the Dow Jones index after evolving one time step ahead. Stars: Forecast without considering local expansion coefficients; encircled: the forecasts remaining after deleting points having large local expansion coefficients. The correlation between forecast and outcome increases from 0.42 (all forecasts, broken line) to about 0.6 (selected forecasts only, solid line).

a correlation between forecast and differentiated time series of 0.45. The differentiated time series was subjected to the standard simplex-forecasting scheme resulting in a correlation ρ between the forecast and outcome of 0.42. Hence the standard nonlinear forecast seems to lead to a performance on the order of a linear autoregressive forecast.

In a second step, the local simplex expansion was also considered and the time series was reanalyzed. Remember that a high local simplex expansion is associated with a considerable amplification of small inacurracies. We therefore rejected forecasts in those cases where the local expansion coefficient exceeded a certain threshold value. This procedure inevitably leads to fewer forecasts, but the remaining ones perform better ($\rho\cong 0.6$, Fig. 6) and exceed the performance of the linear model. Our scheme thus provides a data-driven method for selecting those points in the time series whose forecasts are potentially better than the average forecasts are. By rejecting the rest, the scheme may be regarded as one way to practice selective forecasting.

VII. CONCLUSION

In conclusion, we propose an extension of a certain class of forecasting algorithms. This extension explores the distribution of local simplex expansion coefficients which allows for estimates of the largest Lyapunov exponents together with error bounds. These estimates are correct in the following sense: The differences between the estimated LLE and the true LLE were found to lie within the error estimates for the standard test systems. Considering the distribution of expansion coefficients rather than their means allows one to distinguish between processes where the distributions may have identical means (e.g., LLE's) but differ on the overall shape of the distribution. Our scheme further allows for selective forecasting which may considerably improve the performance of the forecast algorithm itself.

ACKNOWLEDGMENTS

I would like to thank J. Simonent for the NMR raser time series (University of Zürich Physics Institute), M. Dressel (Cantonal Psychiatric Clinic Rheinau) for providing the EEG signals, R. Bertschi (Credit Suisse Zürich) for the financial data, and R. Holzner and P.F. Meier (both from the University of Zürich Physics Institute) for many helpful discussions.

[1] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).
[2] A. R. Osborne and A. Provenzale, Physica D **35**, 357 (1989).
[3] C. Grebogi, E. Ott, S. Pelikan, and J. A. Yorke, Physica D **13**, 261 (1984).
[4] T. Zhou, F. Moss, and A. Bulsara, Phys. Rev. A **45**, 5394 (1992).
[5] J. D. Farmer, Physica D **4**, 366 (1982).
[6] J. G. Caputo and P. Atten, Phys. Rev. A **35**, 1311 (1987).
[7] A. Wolf, J. B. Swift, H. L. Swinney, and A. Vastano, Physica D **16**, 285 (1985).
[8] D. J. Wales, Nature (London) **350**, 485 (1991).
[9] J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986).
[10] H. R. Moser, P. F. Meier, and F. Waldner, Phys. Rev. B **47**, 217 (1993).
[11] M. Dämmig and F. Mitschke, Phys. Lett. A **178**, 385 (1993).
[12] M. Sano and Y. Sawada, Phys. Rev. Lett. **55**, 1082 (1985).
[13] G. Rangarajan, S. Habib, and D. Ryne, Phys. Rev. Lett. **80**, 3747 (1998).
[14] G. Sugihara and R. M. May, Nature (London) **344**, 734 (1990).
[15] H. D. Navone and H. A. Ceccatto, J. Phys. A **28**, 3381 (1995).
[16] R. Badii, E. Brun, M. Finardi, L. Flepp, R. Holzner, J. Parisi, C. Reyl, and J. Simonet, Rev. Mod. Phys. **66**, 1389 (1994).
[17] J. Simonet, E. Brun, and R. Badii, Phys. Rev. E **52**, 2294

(1995).
[18] H. G. Schuster, *Deterministic Chaos*, 2nd ed. (VCH, Weinheim, 1989).
[19] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).
[20] R. M. Dünki and G. B. Schmid, Phys. Rev. E **57**, 2115 (1998).
[21] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young, (Springer-Verlag, Berlin, 1981), pp. 366–381.
[22] M. Dressel, Ph.D. thesis, University of Konstanz, 1999 (unpublished).
[23] M. Dressel, B. Ambühl-Braun, R. Dünki, P. F. Meier, and T. Elbert, in *Workshop on Chaos in Brain*, edited by K. Lehnertz, J. Arnold, P. Grassberger, and C. E. Elger (World Scientific, Singapore, 2000), pp. 348–352.
[24] M. B. Priestley, *Spectral Analysis and Time Series* (Academic Press, New York, 1989).