Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Physics Letters A 310 (2003) 269–280
PHYSICS LETTERS A
www.elsevier.com/locate/pla

Asymptotic scaling laws for precision of parameter estimates in dynamical systems

W. Horbelt *, J. Timmer

Freiburger Zentrum für Datenanalyse und Modellbildung, Universität Freiburg, Eckerstr. 1, 79104 Freiburg, Germany

Received 21 November 2002; received in revised form 10 February 2003; accepted 13 February 2003

Communicated by C.R. Doering

Abstract

When parameters are estimated from noisy data, the uncertainty of the estimates in terms of their standard deviation typically scales like the inverse square root of the number of data points. In the case of deterministic dynamical systems with added observation noise, superior scaling laws can be achieved. This is demonstrated numerically for the logistic map, the van der Pol oscillator and the Lorenz system, where exponential scaling laws and power laws have been found, depending on the number of degrees of freedom. For some special cases, analytical expressions are derived.
© 2003 Elsevier Science B.V. All rights reserved.

PACS: 02.50.-r; 05.45.-a; 05.45.Tp

Keywords: Accuracy of parameter estimation; Precision of estimates; Scaling laws; System identification; Differential equation; Maximum likelihood

1. Introduction

When noisy time series are analysed, one often encounters the situation that the structure of the underlying dynamical process is known, but some coefficients are unknown. These must be estimated from measured data. In this Letter we investigate, how the accuracy of the estimated parameters depends on the number of data points, i.e., the length of the time series.

To be more specific, the parameter estimation problem is stated as follows. Consider an autonomous deterministic time-continuous dynamical system described by a differential equation that depends on a set of unknown parameters $\boldsymbol{p}$.

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}), \quad t \in [T_0, T_0 + T], \tag{1a}$$

$$\boldsymbol{x}(T_0) = \boldsymbol{x}_0. \tag{1b}$$

In the case of time-discrete systems, Eqs. (1) is replaced by

$$\boldsymbol{x}_n = \boldsymbol{x}(t_n) = \boldsymbol{f}(\boldsymbol{x}_{n-1}, \boldsymbol{p}), \quad n = 1, \ldots, N. \tag{2}$$

The dynamical parameters $\boldsymbol{p}$ and the initial values $\boldsymbol{x}_0$ are combined to the vector $\boldsymbol{\theta} = (\boldsymbol{x}_0, \boldsymbol{p})$. Assuming that the first component of $\boldsymbol{x}$ can be observed, the

* Corresponding author.
E-mail address: horbelt@physik.uni-freiburg.de (W. Horbelt).
URL address: http://webber.physik.uni-freiburg.de/~horbelt.

0375-9601/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0375-9601(03)00349-9

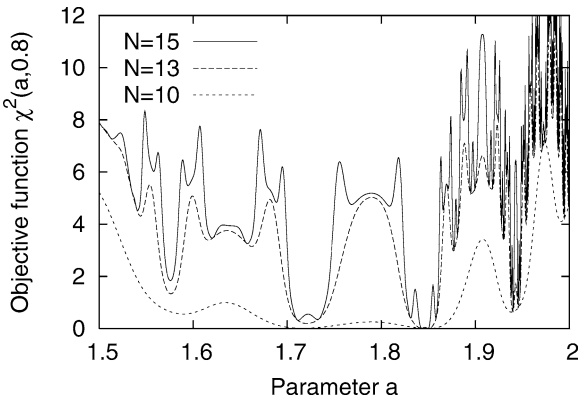

Fig. 1. Objective function of the logistic map Eq. (5) for noiseless data of different lengths N. For each length one time series was simulated with the true parameters $\boldsymbol{\theta} = (x_0, a) = (0.8, 1.85)$ and without noise. Then $\chi^2(\boldsymbol{\theta})$ was plotted as a function of a, x_0 being set to its true value. Curves for higher N are higher than those for smaller N since the summands of Eq. (4) are non-negative. The global minimum at $a = 1.85$ becomes successively sharper but the number of local minima increases.

measurement process is modelled by the *observation equation*

$$y_n = x_1(t_n, \boldsymbol{\theta}) + \eta_n, \quad n = 1, \ldots, N. \tag{3}$$

Here, η_n denotes independent normally distributed random numbers with zero mean and variance σ^2_{data}, accounting for measurement noise.

The maximum likelihood method can be used to estimate $\boldsymbol{\theta}$ from the measured data. It minimises the *objective function*, which is the sum of squared residuals between the data and the model trajectory, divided by the variance of the noise:

$$\chi^2(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(\frac{y_n - x_1(t_n, \boldsymbol{\theta})}{\sigma_{\text{data}}} \right)^2 \stackrel{!}{=} \min. \tag{4}$$

The ML estimator is asymptotically unbiased and efficient [1] because it takes into account all information known about the data. In the context of the n-step prediction error proposed in [2], Eq. (4) corresponds to the N-step prediction error.

$\chi^2(\boldsymbol{\theta})$ can have a quite irregular appearance [3]. Consider the logistic map

$$x_n = f(x_{n-1}, a) = 1 - a x_{n-1}^2, \quad n = 1, \ldots, N \tag{5}$$

with the parameter $a \in (0; 2]$ and the state variable $x_n \in [1 - a; 1]$.[1] In Fig. 1 its objective function is plotted against the dynamical parameter for different lengths of the time series. While the number of local minima increases with N, they become successively sharper, promising that the parameter estimate becomes successively more accurate, provided that the correct minimum is found. It is this increase of the accuracy that we intend to quantify in this Letter. For non-linear regression and stochastic auto-regression systems the standard deviation of the estimates scales like $N^{-0.5}$ [4,5]. For deterministic dynamical systems with added observation noise, stronger power laws and exponential laws are possible, as shown in the sequel.

The following section will provide mathematical definitions and numerical techniques. In Section 3 numerical results will be reported for some examples of dynamical systems. Analytical results will be given in Section 4.

2. Methods

In order to simplify this presentation, we confine ourselves to scalar time-discrete systems in this section. For time-continuous systems more comprehensive descriptions of the parameter estimation methods and applications to measured data can be found in [6–9].

In the scalar case, Eqs. (2) and (3) read

$$x_n = f(x_{n-1}, \boldsymbol{p}), \tag{6}$$

$$y_n = x_n(\boldsymbol{p}) + \eta_n, \quad n = 1, \ldots, N. \tag{7}$$

2.1. Fisher information

The least squares estimate $\hat{\boldsymbol{p}}$ is the parameter vector that minimises the objective function Eq. (4). In the limit of small noise levels we can assume that the estimation error $\Delta\boldsymbol{p}$ between the estimate $\hat{\boldsymbol{p}}$ and the true parameters $\boldsymbol{p}_0$ is also small and we can expand $x_n(\boldsymbol{p})$ with respect to $\Delta\boldsymbol{p}$:

$$x_n(\hat{\boldsymbol{p}}) = x_n(\boldsymbol{p}_0) + \boldsymbol{g}_n \cdot \Delta\boldsymbol{p}, \quad \boldsymbol{g}_n = \left. \frac{\partial x_n}{\partial \boldsymbol{p}} \right|_{\boldsymbol{p}=\boldsymbol{p}_0}. \tag{8}$$

[1] The alternative notation $z_n = r z_{n-1}(1 - z_{n-1})$ is related to Eq. (5) by the transformations $a = \frac{r}{4}(r-2)$ and $x = \frac{4z-2}{r-2}$.

Plugging this and Eq. (7) into the objective function Eq. (4) we get

$$\chi^2(\hat{\boldsymbol{p}}) = \sigma_{\text{data}}^{-2} \sum_{n=1}^{N} \bigl[y_n - x_n(\hat{\boldsymbol{p}})\bigr]^2 \tag{9}$$

$$= \sigma_{\text{data}}^{-2} \sum_{n=1}^{N} \bigl[x_n(\boldsymbol{p}_0) + \eta_n - x_n(\boldsymbol{p}_0) - \boldsymbol{g}_n \cdot \Delta\boldsymbol{p}\bigr]^2 \tag{10}$$

$$= \sigma_{\text{data}}^{-2} \sum_{n=1}^{N} \bigl[\eta_n - \boldsymbol{g}_n \cdot \Delta\boldsymbol{p}\bigr]^2 \stackrel{!}{=} \min. \tag{11}$$

At the estimated parameters the gradient of the objective function with respect to $\Delta\boldsymbol{p}$ vanishes. This leads to the normal equations

$$\boldsymbol{A}\Delta\boldsymbol{p} = \sum_{n=1}^{N} \eta_n \boldsymbol{g}_n \quad \text{with } \boldsymbol{A} = \sum_{n=1}^{N} \boldsymbol{g}_n \boldsymbol{g}_n^t. \tag{12}$$

Some algebra shows that the covariance matrix of the estimate is

$$\bigl\langle \Delta\boldsymbol{p} \Delta\boldsymbol{p}^t \bigr\rangle = \boldsymbol{I}^{-1} = 2\boldsymbol{H}^{-1}, \tag{13}$$

where $\boldsymbol{I} = \sigma_{\text{data}}^{-2}\boldsymbol{A}$ is the Fisher information matrix at the true parameters [1] and $\boldsymbol{H}$ is the Hessian matrix of χ^2. The sensitivities $\boldsymbol{g}_n$ follow from the dynamical equation (2):

$$\boldsymbol{g}_n = \boldsymbol{f}_p(x_{n-1}) + f_x(x_{n-1})\boldsymbol{g}_{n-1}$$
$$\text{with} \quad \boldsymbol{f}_p = \frac{\partial f}{\partial \boldsymbol{p}} \quad \text{and} \quad f_x = \frac{\partial f}{\partial x} \tag{14}$$

and by induction:

$$\boldsymbol{g}_n = \sum_{i=0}^{n-1} \boldsymbol{f}_p(x_i) \prod_{j=i+1}^{n-1} f_x(x_j). \tag{15}$$

In special cases the Fisher information matrix is singular, meaning that the desired parameters cannot be identified uniquely from the data [10]. These exceptions are not considered here.

2.2. Multiple shooting

Finding the global minimum of Eq. (4) can be a demanding task, as can be seen from Fig. 1. For this reason the multiple shooting technique developed for ordinary differential equations [6] was transferred to time-discrete systems [9]. In this approach all system states x_n, $n = 0, \ldots, N$, are treated as fit variables and the constrained non-linear optimisation problem

$$\chi^2(\theta) \stackrel{!}{=} \min, \tag{16a}$$

$$x_n = f(x_{n-1}, \theta), \quad n = 1, \ldots, N \tag{16b}$$

is solved iteratively with a generalised Gauss–Newton method [6] or a quasi-Newton method [11]. Since the constraints Eq. (16b) may be violated on intermediate iteration steps, the final solution can be reached through "forbidden ground". This freedom allows the method to stay close to the observed data and reduces the problem of local minima. As a final precaution the true trajectories were always provided as a starting guess in the optimisation procedure. For a detailed discussion of the numerical issues of the multiple shooting approach see [12].

3. Empirical results

In this section the scaling behaviour is investigated by simulation studies for some typical examples of dynamical systems. The following subsections contain three examples of different complexity.

3.1. Logistic map I: standard deviation of parameter estimates

Time-discrete systems allow us to focus on the essential properties of dynamical systems. The most prominent deterministic time-discrete system is the logistic map Eq. (5). For $a > 2$ the map is unstable, so the constraint $a \leqslant 2$ has to be implemented in the parameter estimation procedure. In order to circumvent boundary effects on the statistics, $a = 1.85$ was chosen as the true parameter.

Time series with length N ranging from 2 to 60 were generated according to Eq. (5). For larger N the sensitivity of the end of the trajectory with respect to its beginning would approach a limit imposed by the machine precision. The initial value was $x_0 = 0.8$. White Gaussian noise with a noise level of 10% was added, i.e., the standard deviation of the noise was 10% of that of the signal. Then the multiple shooting approach described above was applied to the data. At

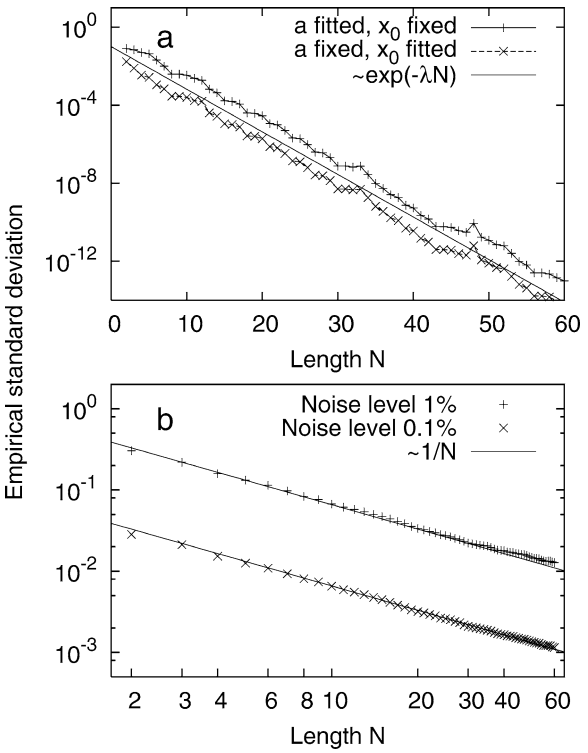

Fig. 2. Sample standard deviation of estimates for logistic map; (a) one degree of freedom; crosses: initial value x_0 fixed to 0.8, parameter a fitted; plus signs: parameter fixed to 1.85, initial value fitted; solid line: exponential curve with exponent $\lambda = 0.50263$; (b) two degrees of freedom; crosses: 0.1% noise; plus signs: 1% noise; solid lines: $1/N$ curves.

first only a was estimated while x_0 was assumed to be known.

For each N, 1000 time series were generated and the parameter estimates $\hat{a}_i$ were obtained. Although their distribution is not exactly Gaussian, especially for small N, their mean $\bar{a}$ and their sample standard deviation $\sigma = \left[\frac{1}{999}\sum_i (\hat{a}_i - \bar{a})^2\right]^{1/2}$ were computed as a measure of the accuracy. Fig. 2(a) displays σ as a function of N. It turns out to scale like $\exp(-\lambda N)$, where $\lambda = 0.503$ is the Lyapunov exponent of the logistic map, defined by

$$\lambda = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \log\left|f'(x_n)\right|. \tag{17}$$

In the next simulation, a was fixed to 1.85 and x_0 was fitted. Again 1000 time series were analysed for each N, leading to a similar result.

Finally, both variables were fitted simultaneously. In this case, $\sigma(N)$ came out as a rather discontinuous curve rather than a smooth function. The accuracy increases almost step-like at certain values of N which depend on the initial value x_0. The reason is that some data points in the time series contain more information than others. This problem was tackled by choosing a different initial value x_0 for each of the 1000 replications, so that the steps were smoothed out. The initial values were generated by iterating Eq. (5), thus their distribution reflects the invariant measure of the map. Furthermore a small number of failed fits was excluded, for which the final value of the objective function was ten times higher than expected. Although the multiple shooting approach helps in circumventing local minima, it cannot suppress them completely.

Again the standard deviation of the estimated parameter was plotted against N (see Fig. 2(b)). In this case it seems to scale like $1/N$. The deviation of the estimated initial value from the respective true value was also analysed. It did not show an equivalent scaling law. This can be understood as a consequence of the varying initial value. The sample standard deviation of the estimation error is not a meaningful number in this case.

Regarding these results it seems that the mere number of degrees of freedom determines the kind of scaling law while the difference between initial values and dynamic parameters is less important. The scaling curves shown in Fig. 2 are jagged due to the observation noise in the time series. The method used in the following subsection estimates the parameter uncertainties from noiseless data and yields, therefore, smoother curves.

3.2. *Logistic map II: standard deviation from Fisher information*

In the limit of small noise levels, the variance and covariance of the estimated parameters are given by Eq. (13). The sensitivities Eq. (15) can be computed directly from noiseless data. Nevertheless it is necessary to regard an ensemble of time series for each N since they depend on the initial values.

The sensitivities with respect to the initial value are calculated differently than those with respect to dynamic parameters. In order to simplify the computations in the case of two degrees of freedom, we intro-

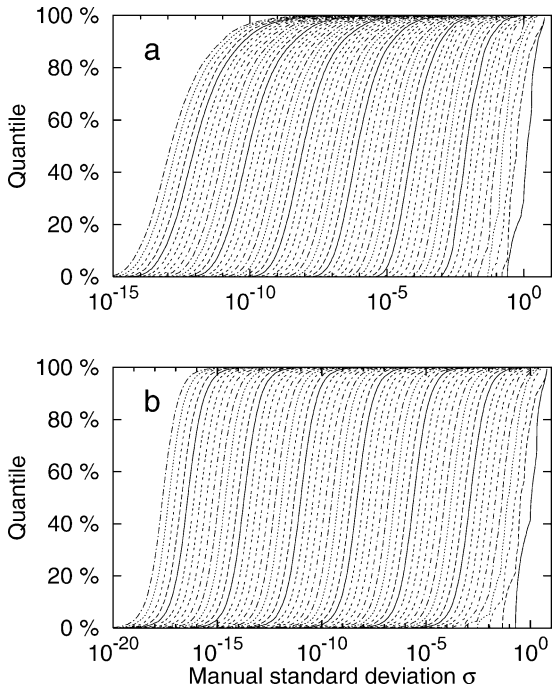

Fig. 3. Distributions of σ for the generalised logistic map with one degree of freedom, on a logarithmic scale, with $N = 2, \ldots, 60$ from right to left. The horizontal axis shows the standard deviation σ_i, computed manually as described in the text. The vertical axis shows the corresponding quantile; (a) $a = 1.85$; (b) $a = 2$.

duce the generalised logistic map

$$x_n = a_0 - a_2 x_{n-1}^2, \quad n = 1, \ldots, N \tag{18}$$

with two dynamic parameters, a_0 and $a_2 = a$. Though a_0 is still always equal to 1, it is now viewed as a fit parameter, either known or unknown.

For each N from 2 to 60, 10000 time series were computed with different initial values generated as before. At first a_0 was assumed to be known. For each realization i, the (scalar) matrix $\boldsymbol{A}$ in Eq. (13) was computed and inverted, giving a "manual standard deviation" σ_i. These are to be multiplied with σ_{data} to get the theoretical standard deviations in the limit of small σ_{data}.

In Fig. 3(a) the statistical distributions of the σ_i are shown on a logarithmic scale for all N. The rightmost curve corresponds to $N = 2$ and is jerky due to the finite number of data points. The leftmost curves are successively smoother, but also broader. An interesting result visible is that the logarithm of σ but not σ itself has a nearly Gaussian distribution. Fig. 3(b) shows the distributions for $a = 2$. In this case the shape of the curves is almost identical for all N, except for very small N. This reflects the fact that the logistic map has a higher symmetry for $a = 2$. The exponential scaling law is already visible: the distance between two adjacent curves is approximately constant in this semi-logarithmic plot. However, when looked upon on a linear scale, the curves would be dominated by those rare realizations with extraordinarily high σ_i. Therefore, the simple mean of σ_i does not provide a meaningful measure. A more suitable number is the log-averaged standard deviation $\tilde{\sigma}$, defined by

$$\log \tilde{\sigma} = \frac{1}{M} \sum_{i=1}^{M} \log \sigma_i. \tag{19}$$

It will be used in all following investigations.

In Fig. 4(a), $\tilde{\sigma}$ resulting from the data in Fig. 3(a) is compared with the corresponding exponential function $\exp(-\lambda N)$. This plot can be seen as an improved version of Fig. 2(a). In Fig. 4(c) the same can be seen for the data of Fig. 3(b), corresponding to $a_2 = a = 2$. When a_0 was the unknown fit parameter, the same results were obtained (see Fig. 4(a) and (c)). In these figures, no deviation from the exponential scaling laws is visible. In order to show how good the accordance really is, all $\tilde{\sigma}$ were multiplied with $\exp(\lambda N)$. This is shown in Fig. 4(b) and (d), respectively.

Fig. 5 shows the results for two degrees of freedom, i.e., both a_0 and a_2 were assumed to be unknown. In this case $\boldsymbol{A}^{-1}$ is a 2×2 matrix and the square roots of the two diagonal elements were plotted. Regardless of a_2, $\tilde{\sigma}$ scales asymptotically roughly like $1/N$, confirming the result of Fig. 2(b). For $a = 2$, the Lyapunov exponent is higher and the limit of the machine precision is reached earlier than for $a = 1.85$. Therefore the plot range in Fig. 5 was restricted to $N \leqslant 50$.

3.3. Lorenz system

In this subsection, as an example of a continuous dynamical system, we regard the Lorenz system

$$\dot{x} = s(y - x), \tag{20}$$

$$\dot{y} = -xz + rx - y, \tag{21}$$

$$\dot{z} = xy - bz \tag{22}$$

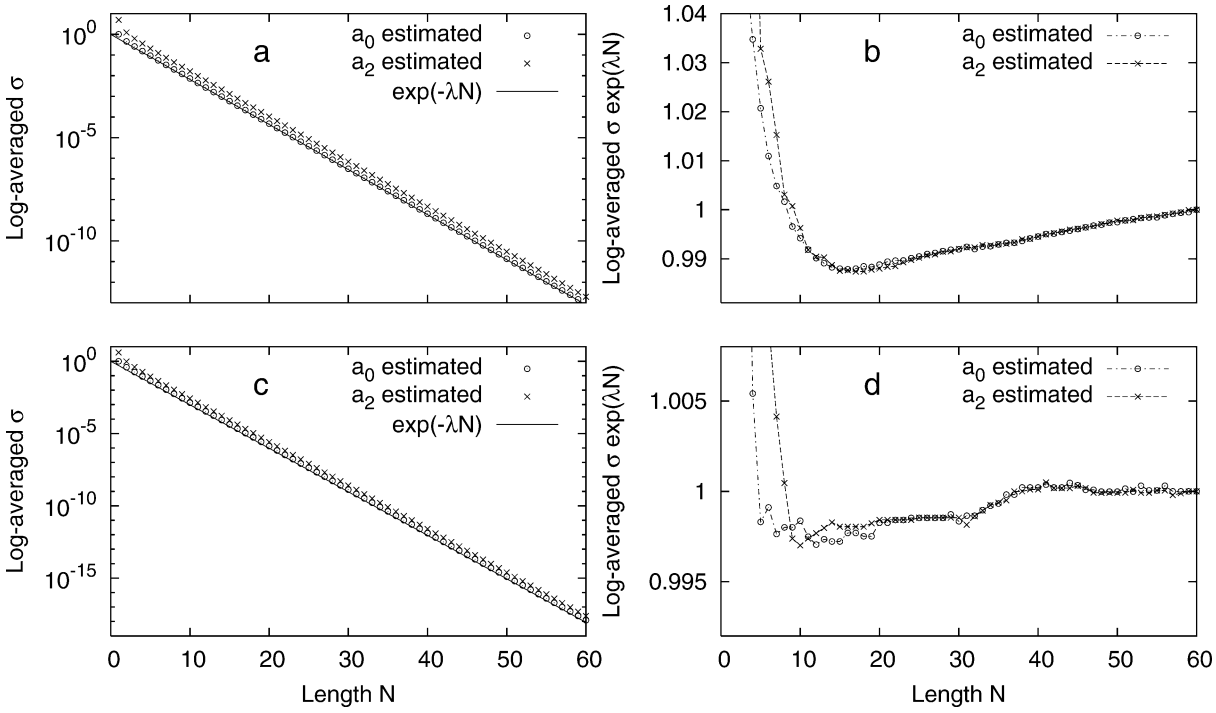

Fig. 4. Log-averaged σ as a function of N for the generalised logistic map with one degree of freedom. The true parameters were $a_0 = 1$, $a_2 = 1.85$ in the upper panels, corresponding to Fig. 2(a), and $a_0 = 1$, $a_2 = 2$ in the lower panels. In the left column (a), (c), $\tilde{\sigma}$ for the estimation of a_0 and a_2 is drawn with circles and crosses, respectively. The accuracy scales exponentially. The solid lines indicate exponential functions $\exp(-\lambda N)$, where $\lambda = 0.50263$ in panel (a) and $\lambda = \log(2)$ in panel (c). These are the Lyapunov exponents of the map for the respective parameters. The right panels show the deviation of $\tilde{\sigma}$ from the exponential law, i.e., $\tilde{\sigma}\exp(\lambda N)$, normalised so that the curves are equal to 1 at $N = 60$.

with $s = 10$, $r = 46$ and $b = 2.667$. For each T between 1 and 25, 1000 noiseless time series of length T were simulated with a sampling time $\Delta t = 0.04$. Then the parameter estimation procedure described in [8,9] was applied to the x component to minimise the objective function Eq. (4) with respect to one, two or three fitted parameters. The initial values were fixed to their true values in all cases. Finally the covariance matrix was evaluated at the solution point to obtain standard deviations σ_i for the fitted parameters. Fig. 6(a) shows the distribution functions for these σ_i in the case of one degree of freedom (s fitted, r and b fixed). The rightmost curve corresponds to $T = 1$. With increasing T, the shape of the curves becomes more symmetric. The log-averaged σ scales like $\exp(-\lambda T)$ to within linewidth (see crosses in Fig. 6(b)), where $\lambda = 1.24$ is the Lyapunov exponent of the system [13]. The same result is obtained when b is fitted (circles) or when r is fitted (not shown).

Fig. 6(c) and (d) show some of the results when two or all three parameters were fitted. Again the exponential law turns into a power law $\sigma \sim T^{\alpha}$. For two degrees of freedom, the exponent seems to be near -1.5. Numerical limitations bound the T range that can be exploited to below 15. In Fig. 6(d) a power law with $\alpha = -1$ is drawn for comparison although due to numerical errors the curves do not contain a large region of constant slope. It should be emphasized that in all cases the scaling behaviour is clearly better than the usual $1/\sqrt{N}$ law. This is because the complete dynamic information of the data is taken into account for the parameter estimation.

3.4. Van der Pol oscillator

For the logistic map and the Lorenz system with one degree of freedom, the accuracy of the parameter estimates appeared to be ruled by the positive Lya-

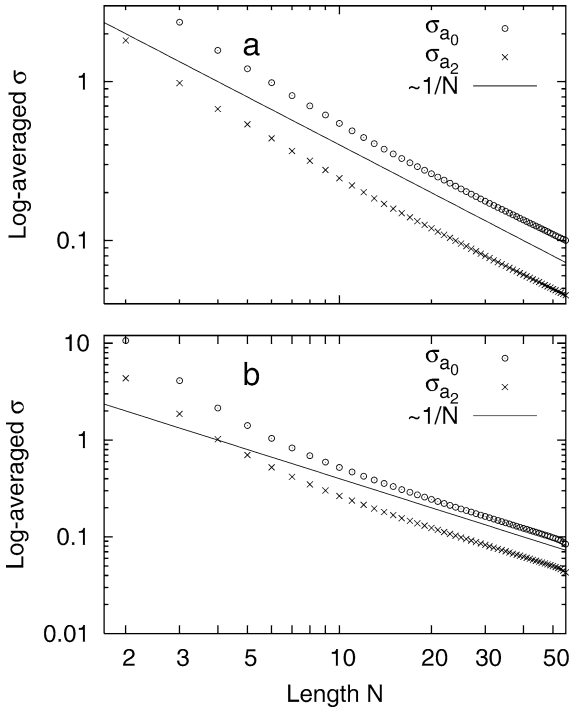

Fig. 5. Same as Fig. 4 for two degrees of freedom in double-logarithmic plot. The two point types correspond to the two square roots of the diagonal elements of the computed covariance matrix. The solid lines indicate the power law $\sigma \sim N^{-1}$. (a) $a = 1.85$; (b) $a = 2$.

punov exponent. This indicates that an exponential scaling law is only possible for chaotic systems. However, power laws with exponents below -0.5 are possible for other examples. In this section we consider the van der Pol oscillator, a system showing a limit cycle. The dynamical equation reads

$$\ddot{x} = k\dot{x}\left(1 - x^2\right) - x, \tag{23}$$

with $k = 1$. It generates oscillations with a period $\tau = 6.66$, which is used as a unit in the sequel. We simulated time series of length $N\tau$ with $N = 2, 4, \ldots, 200$, sampled with $\Delta t = 0.1\tau$.

To obtain initial values on the attractor, a master piece of length $\tau/2$ was simulated after removing the transient. Due to the invariance of the equation with respect to sign reversal, the second half period differs from the first one only in the sign and gives identical results with respect to the accuracy of estimates. Initial values were read off at 10 equidistant times (*phases*) on the master piece. For each N and each phase, the parameter estimation procedure was performed with noiseless data and standard deviations σ_i were determined from the Fisher information matrix, analogously to Section 3.3.

The σ_i were averaged for each N and plotted versus N in Fig. 7(a). They scale like $T^{-3/2}$. In Fig. 7(b) they are multiplied with $T^{3/2}$ to show the net effect. The net values are asymptotically constant. Finally in Fig. 7(c) we regard the dependence of the standard deviation on the initial value. For this simulation, 100 equidistant phases were chosen for each N. For N ranging from 5τ to 55τ in steps of 5τ, the computed standard deviation is plotted as a function of the phase of the initial value within the master piece. These functions are $\frac{\tau}{2}$-periodic as mentioned above. The effect of the scaling law $\sigma \sim T^{-3/2}$ was removed from these curves by multiplying them with $T^{3/2}$. The effect of the phase is like a boundary effect. It is most pronounced for short trajectories and flattens out for larger ones.

3.5. Summary of empirical results

In the preceding section scaling laws of various dynamical systems were determined empirically. These results are summarised in Table 1. Also included are results for the two-dimensional Hénon map that were not reported in detail. In this case the empirically determined scaling exponent $\lambda' = 0.3$ was not equal to the Lyapunov exponent $\lambda = 0.419$.

4. Analytical results

The preceding section has demonstrated that dynamical systems show a variety of different scaling laws. Now we will give results of a mathematical analysis of the behaviour in some selected cases.

4.1. Stationary stochastic Markov processes

Though this non-deterministic class of dynamical systems is not in the main scope of this Letter, it is well suited for comparison. Consider a stochastic Markov process without observation noise

$$\boldsymbol{x}_n = \boldsymbol{f}(\boldsymbol{x}_{n-1}, \boldsymbol{\eta}_n, \boldsymbol{p}), \tag{24}$$

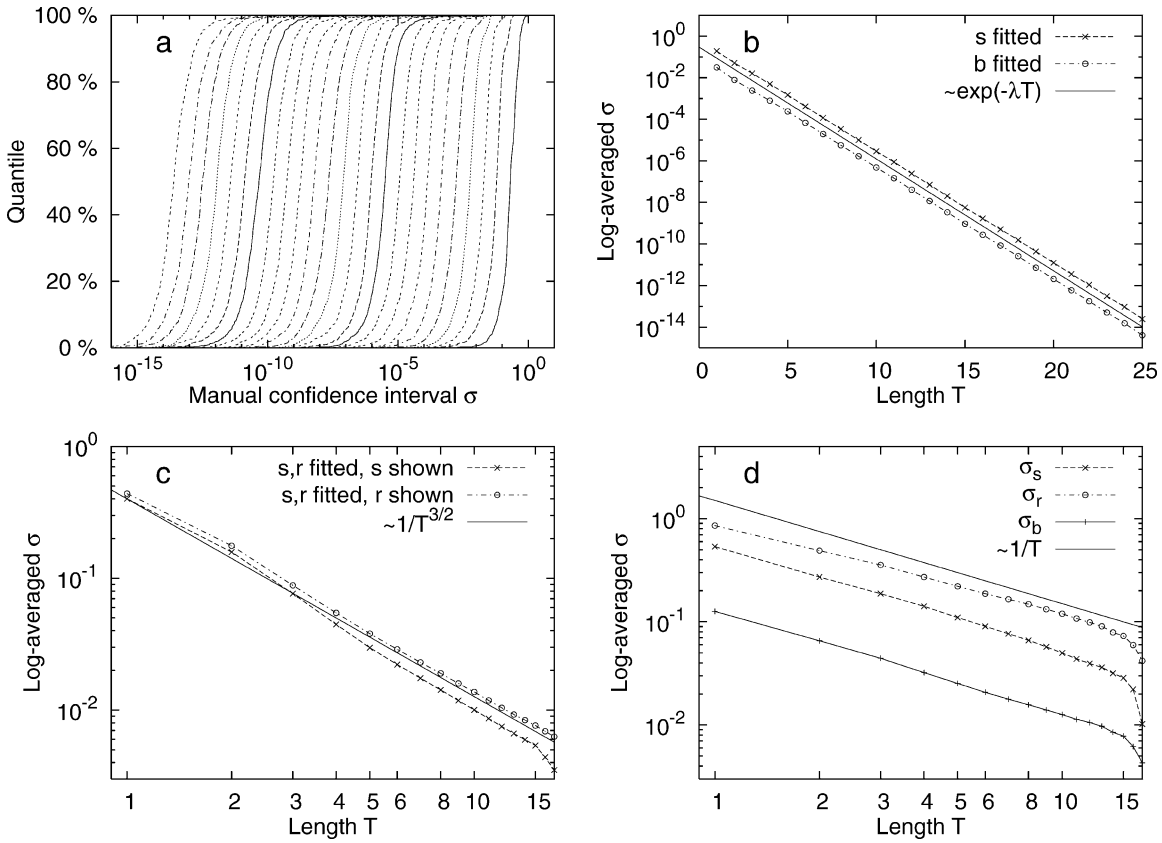

Fig. 6. Confidence intervals for the Lorenz system, computed from fits with noiseless data. (a) Empirical distribution functions of standard deviations for estimation of s. All other parameters and all initial values were fixed. T takes values from 1 to 25 from right to left. For each T 1000 fits were made. (b) Log-averaged standard deviations for a single degree of freedom (points). The solid line indicates the exponential scaling law $\sigma \sim \exp(-\lambda T)$, where $\lambda = 1{,}24$ is the Lyapunov exponent of the Lorenz equation for the chosen parameters. Each cross in the upper curve represents the mean of one curve in panel (a). (c) Same for two degrees of freedom (points). The solid line indicates a power law $\sigma \sim 1/T^{3/2}$. (d) Same for three degrees of freedom (points). The solid line corresponds to $\sigma \sim 1/T$. In this case no definite scaling law can be identified.

where $\boldsymbol{\eta}_n$ are independent identically distributed random vectors. Denote the true parameters with $\boldsymbol{p}_0$ and the maximum likelihood estimates with $\hat{\boldsymbol{p}}$. From the martingale central limit theorem follows under the assumption of ergodicity and stationarity, that $(\hat{\boldsymbol{p}} - \boldsymbol{p}_0)\sqrt{N}$ is asymptotically normally distributed with constant covariance matrix, i.e., the standard deviation of the estimation error scales like $1/\sqrt{N}$ (see [14], p. 131).

4.2. Scalar chaotic maps

This subsection derives analytically the exponential scaling law for scalar chaotic maps with one degree of freedom, seen, e.g., in Fig. 2(a). In order to simplify the discussion, we regard only the dependence on dynamic parameters while the initial value is assumed to be known.

Following up Eq. (15) we can write

$$\boldsymbol{g}_n = z_n \boldsymbol{u}_n, \tag{25}$$

$$\text{with} \quad z_n = \prod_{j=0}^{n-1} f_x(x_j), \tag{26}$$

$$\text{and} \quad \boldsymbol{u}_n = \sum_{i=0}^{n-1} \boldsymbol{f}_p(x_i) z_{i+1}^{-1}. \tag{27}$$

The numbers z_n embody the high sensitivity of the end of the trajectory with respect to $\boldsymbol{p}$. On average,

Table 1
Summary of scaling laws

Dynamical system	Technique used	Degrees of freedom	See Fig.	Scaling law
Logistic map	Empirical standard deviation, evaluated from ensemble of fits with noisy data	1	2(a)	$\exp(-\lambda N)$
		2[a]	2(b)	N^{-1}
Generalised logistic map	Manual standard deviation, computed by recursion, log-averaged	1	4	$\exp(-\lambda N)$
		2	5	N^{-1}
Hénon map	Empirical standard deviation, evaluated from ensemble of fits with noisy data	1	–	$\exp(-\lambda' N)$
		2	–	N^{-1}
Lorenz system	Noiseless fit	1	6(a)	$\exp(-\lambda T)$
		2	6(b)	$T^{-1.5}$
		3	6(c)	indefinite
Van der Pol oscillator	Noiseless fit	1	7	$T^{-1.5}$

[a] One parameter and one initial value.

$|z_n|$ increases exponentially. To be more specific, we regard

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log|z_n| \tag{28}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log\left|f_x(x_j)\right|, \tag{29}$$

which is exactly the definition of the Lyapunov exponent in Eq. (17).

On the other hand, $\boldsymbol{u}_n$ in Eq. (27) behaves like a noisy power series since $\boldsymbol{f}_p(x)$ is bounded and z_{i+1}^{-1} decays exponentially. We denote its limiting value by $\boldsymbol{u}_\infty$. Regarding only the asymptotics of $\boldsymbol{g}_n$, we can replace $\boldsymbol{u}_n$ with $\boldsymbol{u}_\infty$ in Eq. (25). Then we arrive at

$$\boldsymbol{A} \approx \boldsymbol{u}_\infty \boldsymbol{u}_\infty^t \sum_{n=1}^{N} z_n^2 \tag{30}$$

$$\approx \underbrace{\boldsymbol{u}_\infty \boldsymbol{u}_\infty^t \frac{1}{1-e^{-2\lambda}}}_{\text{const}} e^{2\lambda N}. \tag{31}$$

The latter relation is a rough estimate obtained by replacing $|z_n|$ with $\exp(\lambda n)$ according to Eq. (28). When a single parameter is estimated, $\boldsymbol{A}$ is a scalar and the standard deviation of the estimate is $\sigma = \sigma_{\text{data}}\sqrt{\boldsymbol{A}} \sim \exp(-\lambda N)$, in concordance with the empirical results of Figs. 2(a) and 4(a), (c).

In the case of two degrees of freedom, $\boldsymbol{A}$ is a square matrix. In the approximation of Eq. (30), it is singular, since $\boldsymbol{u}_\infty \boldsymbol{u}_\infty^t$ has rank one. That means that only one eigenvalue of $\boldsymbol{A}$ increases like $\exp(\lambda N)$ while the other one is much smaller. In most cases the estimated directions in the parameter space are linear combinations of both eigenvectors, so the corresponding variances do not scale exponentially. It would be desirable to understand the power law shown in Fig. 2(b) in a similar manner, however, this phenomenon is more difficult to access than the simple result for a single degree of freedom.

4.3. Periodic systems

The previous section considered a system with an exponential scaling law. In the sequel we will derive analytically the power law that is visible in Fig. 7. Consider a model that depends on a single parameter p and generates scalar periodic trajectories $x(t, p)$ with period $\tau(p)$ and frequency $\omega(p) = 2\pi/\tau(p)$, i.e.,

$$x(t, p) = f\big(\omega(p)t, p\big), \tag{32}$$

where $f(\phi, p)$ is 2π-periodic in ϕ. The true parameter shall be denoted by p_0. Furthermore assume that a measurement of the trajectory is made over the length $N\tau(p_0)$ with a sampling time $\Delta t = \tau(p_0)/m, m \in \mathbb{N}$.

The observation equation (7) holds with $n = im + j$, $i = 0, \ldots, N-1$, $j = 1, \ldots, m$. Eqs. (8)–(13) do

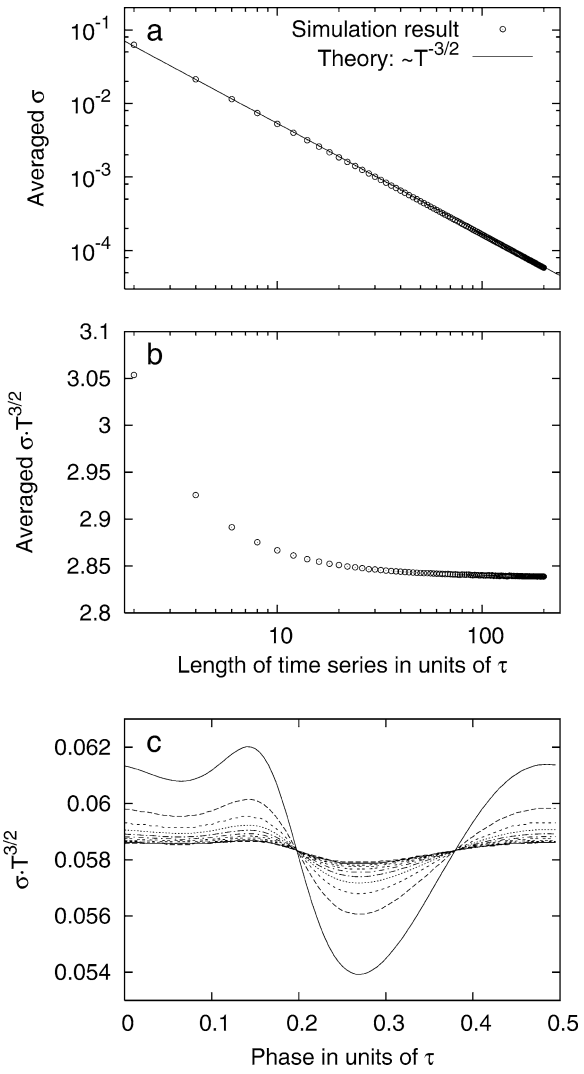

Fig. 7. Confidence intervals for the van der Pol oscillator, computed from fits with noiseless data; (a) averaged standard deviations as function of N in double-logarithmic plot (points) and a scaling law $\sigma \sim 1/T^{3/2}$ (solid line); (b) deviation of simulation from the scaling law. Note the small range of the vertical axis; (c) scaled standard deviation as a function of the starting point of the trajectory. Only the first half period is shown. The second half is identical due to a symmetry of the system. The different curves correspond to different lengths, ranging from 5τ (out-most curve) to 55τ (flattest curve). Each curve was multiplied with $T^{3/2}$ to remove the effect of the scaling law. For longer time series, the standard deviation is successively less dependent on the starting point.

also apply but the sensitivities are now

$$g_n = \frac{\mathrm{d}}{\mathrm{d}p} f\big(\omega(p)n\Delta t, p\big)\Big|_{p=p_0} \tag{33}$$

$$= \frac{\partial f}{\partial \phi}\big(\omega(p_0)n\Delta t, p_0\big)n\Delta t\frac{\mathrm{d}\omega}{\mathrm{d}p} + \frac{\partial f}{\partial p}\big(\omega(p_0)n\Delta t, p_0\big). \tag{34}$$

Case 1. The parameter has no influence on the period, i.e., $\frac{\mathrm{d}\omega}{\mathrm{d}p} = 0$.

Then the equation $\omega(p_0)m\Delta t = 2\pi$ and the 2π-periodicity of $\frac{\partial f}{\partial p}$ lead to

$$g_n = \frac{\partial f}{\partial p}\big(\omega(p_0)(im+j)\Delta t, p_0\big) \tag{35}$$

$$= \frac{\partial f}{\partial p}\big(\omega(p_0)j\Delta t, p_0\big) \tag{36}$$

and

$$\boldsymbol{A} = \sum_{n=1}^{Nm} g_n^2 = \sum_{i=0}^{N-1} A_1 = NA_1 \tag{37}$$

$$\text{with} \quad A_1 = \sum_{j=1}^{m}\left[\frac{\partial f}{\partial p}\big(\omega(p_0)j\Delta t, p_0\big)\right]^2. \tag{38}$$

A_1 is independent from i. According to Eq. (13) the standard deviation of the estimate scales like $N^{-0.5}$.

Case 2. ω depends on p.

If $\frac{\mathrm{d}\omega}{\mathrm{d}p}$ does not vanish, the first term of Eq. (34) is of the order $\mathcal{O}(n)$. The second term is bounded since $\frac{\partial f}{\partial p}$ is periodic. So the first term is dominant for $n, i \to \infty$. We expand g_n in terms of powers of i and retain only the highest order term, indicating the others with $\mathcal{O}(.)$:

$$g_n = \frac{\partial f}{\partial \phi}\big(\omega(p_0)j\Delta t, p_0\big)im\Delta t\frac{\mathrm{d}\omega}{\mathrm{d}p} + \mathcal{O}(1). \tag{39}$$

Again recalling Eq. (12) we get

$$\boldsymbol{A} = \sum_{i=0}^{N-1} i^2 A_2 + \mathcal{O}(i) = \frac{1}{3}N^3 A_2 + \mathcal{O}\big(N^2\big) \tag{40}$$

$$\text{with} \quad A_2 = \sum_{j=1}^{m}\left[\frac{\partial f}{\partial \phi}\big(\omega(p_0)j\Delta t, p_0\big)m\Delta t\frac{\mathrm{d}\omega}{\mathrm{d}p}\right]^2.$$

Again looking at Eq. (13), the standard deviation of $\hat{p}$ scales asymptotically like $N^{-1.5}$ in accordance with Fig. 7.

In summary this underlines that parameters influencing the period of a trajectory are more delicate than parameters that only affect the shape of a periodic orbit. A similar result was obtained in [15] for the chaotic Mackey–Glass equation. Though this delayed-feedback system does not produce periodic trajectories for the parameters regarded, the delay time has an important influence on intrinsic oscillations and accordingly the estimated delay parameter turned out to be extraordinarily accurate.

5. Summary

In this Letter we regarded the standard deviation σ of parameters estimated from noisy time series of deterministic dynamical systems and investigated its relation to the length of the time series. Depending on the number of degrees of freedom, different asymptotic scaling laws were found. Scalar time discrete systems as well as continuous chaotic systems with one unknown parameter show an exponential scaling law with the coefficient in the exponent being equal to the Lyapunov exponent of the system. For scalar time discrete systems this result could be explained analytically. When two parameters are estimated, the exponential law turns into a power law $\sigma \sim N^{\alpha}$. However, in all examples analysed, the exponent α was still well below -0.5. Finally, for periodic systems in which the period is influenced by a parameter, a power law $\sigma \sim N^{-1.5}$ was both derived analytically and demonstrated numerically.

The fact that the exponential law holds only for one degree of freedom has great practical relevance. It means that the extreme sensitivity of a chaotic system with respect to its initial values (quantified by its positive Lyapunov exponent) cannot be exploited to yield precise parameter estimates if nuisance parameters must be estimated simultaneously.

Other examples in which the accuracy of an estimate does not follow the usual $N^{-0.5}$ law, were reported in [16,17]. These effects cannot be reproduced when small amounts of observation noise is added to the data. In contrast, the scaling laws reported in this Letter were robust with respect to observation noise.

The reason for the superior scaling laws is that the objective function exploits the full dynamic information of the time series rather than all data points independently as in the case of regression or stochastic dynamical systems. This requires that the time series can successfully be modelled by a deterministic differential equation. Moreover we assumed that the structure of the model is known and that the global minimum of the cost function is found. For practical applications these conditions are not always fulfilled. Nevertheless one can say that for the purpose of parameter estimation and system identification it is important to take into account the complete dynamic nature of the data.

Note added in proof

After acceptance of the Letter, the authors were pointed to Ref. [18], in which the relation between the accuracy of initial value estimates and the Lyapunov exponent has been established.

References

[1] E.L. Lehmann, Theory of Point Estimation, Chapman & Hall, New York, 1991.
[2] L. Jaeger, H. Kantz, Chaos 6 (1996) 440.
[3] R.N. Miller, M. Ghil, F. Gauthiez, J. Atmos. Sci. 51 (1994) 1037.
[4] R.H. Shumway, D.S. Stoffer, Time Series Analysis and Its Application, Springer-Verlag, New York, 2000.
[5] P.J. Brockwell, R.A. Davis, Time Series: Theory and Models, Springer-Verlag, Heidelberg, 1991.
[6] H.G. Bock, Recent advances in parameter identification techniques for O.D.E., in: P. Deuflhard, E. Hairer (Eds.), Numerical Treatment of Inverse Problems in Differential and Integral Equations, Birkhäuser, Basel, 1983, p. 95.
[7] J. Timmer, H. Rust, W. Horbelt, H.U. Voss, Phys. Lett. A 274 (2000) 123.
[8] W. Horbelt, J. Timmer, M. Bünner, R. Meucci, M. Ciofini, Phys. Rev. E 64 (2001) 016222.
[9] W. Horbelt, Maximum likelihood estimation in dynamical systems, Ph.D. thesis, University of Freiburg, (2001), http://webber.physik.uni-freiburg.de/~horbelt/diss/.
[10] K. Judd, L. Smith, Physica D 151 (2001) 125.
[11] P.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, London, 1981.
[12] H.G. Bock, Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen, in: Bonner Math. Schriften, Vol. 183, 1987, Ph.D. thesis, Universität Bonn.
[13] E. Baake, M. Baake, H.G. Bock, K.M. Briggs, Phys. Rev. A 45 (8) (1992) 5524.
[14] H. Tong, Threshold Models in Non-Linear Time Series Analysis, Springer-Verlag, Heidelberg, 1983.

[15] W. Horbelt, J. Timmer, H.U. Voss, Phys. Lett. A 299 (2002) 513.
[16] K. Kaneko, Phys. Rev. Lett. 65 (1990) 1391.
[17] J. Theiler, L.A. Smith, Phys. Rev. E 51 (1995) 3738.
[18] M.D. Richard, Estimation and Detection with Chaotic Systems, Ph.D. thesis, MIT, 1994.