

A Simple but Efficient Method for Nonlinear Parameter Estimation Based on Comparing Phase Space Structures

Kazuhiro Matsumoto^a and Hans H. Diebner^b

^a Hokkaido University, Department of Mathematics, Faculty of Science, Kita 10 Nishi 8, Kita-ku, Sapporo 060-0810, Japan

^b Institute for New Media, Schmickstr. 18, D-60314 Frankfurt am Main, Germany

Reprint requests to Dr. H. H. D.; E-mail: hans@diebner.de

Z. Naturforsch. **61a**, 239–248 (2006); received March 15, 2006

We introduce a simple method for nonlinear parameter estimation based on a structural comparison of target and model attractor. The parameters of the model are adapted by means of minimizing the structural difference of the attractors. For this quantitative comparison histograms derived from a coarse graining of the phase spaces are used. We present a time discrete as well as a continuous example to demonstrate the efficiency of this method. The target attractors are computed from the Hénon map and the Rössler system, respectively. The model systems are chosen to be fairly universal endowed with free parameters that are adapted so that the model attractor resembles the target. The estimations work accurate and acceptably fast up to four parameters.

Key words: Nonlinear Time Series Analysis; Parameter Estimation; System Identification; Attractor Reconstruction.

1. Introduction

Due to the sensitive dependence on initial conditions classical methods for parameter estimation like the least squares method applied to time series data fail for nonlinear dynamical systems. Therefore, a couple of new methods have been derived for chaotic time series [1]. Prominent methods rest upon synchronization of the model with the target (data) system [1 (chapter 14)–3]. The parameter values of the model system are adapted by means of minimizing the “force” for synchronization. The set of free parameters defines a class of model systems. The model that corresponds to the estimated set of parameter values then can be interpreted as the model with the least synchronization force. This is a temporal argument. In some cases, however, it can be observed that this procedure leads to a non-bounded model system in the phase space. In other words, sometimes a diverging system can be easily (using a very small force term) synchronized to a bounded chaotic attractor. Structurally, however, model and target system are quite different in these cases.

In this paper we argue for a structural comparison of target and model attractor. We suggest a very simple algorithm which turns out to be quite effective. In a nutshell, we compare the target with the model at-

tractor by defining distributions on a coarse grained grid in phase space. For the comparison of the resulting histograms we suggest two possibilities: a least squares procedure and a comparison based on a Poisson distribution, respectively. The least squares procedure can be justified in the limit of high “occupation numbers” in the coarse grained cells. Depending on the specifically chosen coarse graining a multinomial or a Poisson distribution may be more appropriate. We report on the usage of Gaussian, weighted Gaussian and Poisson distribution, respectively, applied to a discrete chaotic attractor. We also report on the application of the Gaussian and weighted Gaussian to a continuous chaotic attractor.

2. Data

The target time series are artificially computed from known systems in order to judge the quality of the estimations. We demonstrate the estimations of parameters for the discrete Hénon map as well as for the continuous Rössler attractor. The Hénon map is defined as

$$\begin{aligned}x_{n+1} &= \alpha_1 - \alpha_2 x_n^2 + \alpha_3 y_n, \\y_{n+1} &= \alpha_4 x_n,\end{aligned}\tag{1}$$

where the α_i are constant parameters. The transient phase is skipped and 600 points are kept to provide

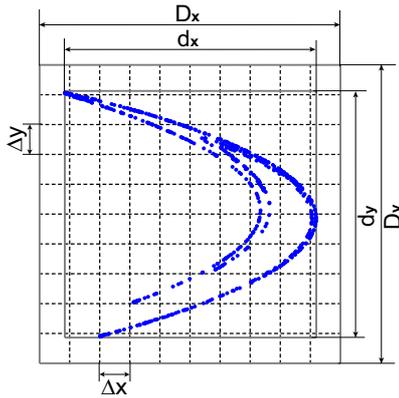


Fig. 1. Coarse graining for the Hénon attractor. The rectangular area (with sidelengths D_x and D_y , respectively) that encloses the attractor is chosen to be one tenth larger than the attractor's width (d_x and d_y , respectively) at each side with respect to the given coordinate. Then this area is subdivided into $\gamma \times \gamma$ cells, here $\gamma = 10$ (dotted lines). Since γ defines a relative spacing with respect to the maximum extension the actual coordinate labeling is suppressed. The Hénon attractor is represented by 600 subsequent points whereby the transient phase has been skipped.

a target system. The values of the target parameters have been chosen to be $\alpha_1 = 1.0$, $\alpha_2 = 1.4$, $\alpha_3 = 1.0$, $\alpha_4 = 0.3$.

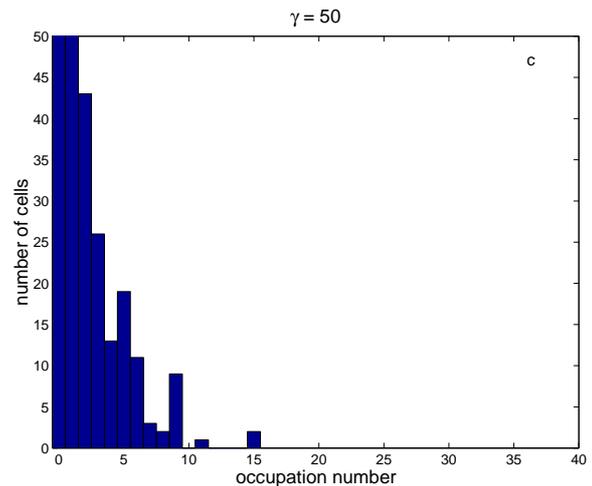
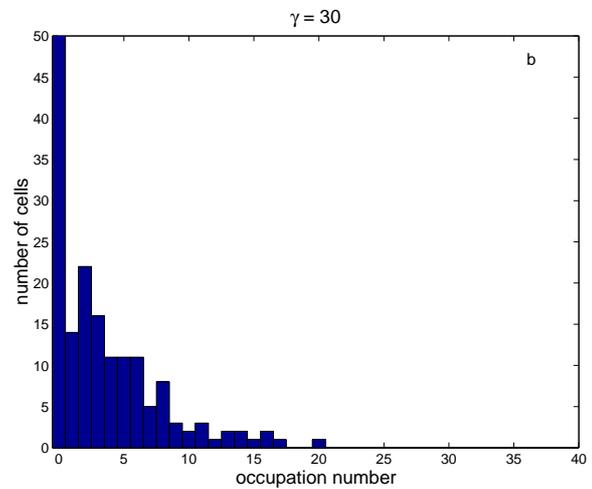
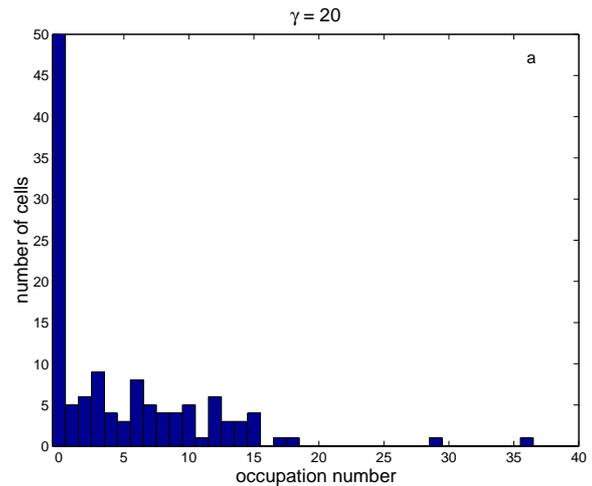
The Rössler system reads:

$$\begin{aligned} \dot{x}(t) &= -\beta_1 z(t) - \beta_2 y(t), \\ \dot{y}(t) &= \beta_3 x(t) + \beta_4 y(t), \\ \dot{z}(t) &= \beta_5 + \beta_6 x(t)z(t) - \beta_7 z(t), \end{aligned} \quad (2)$$

where \dot{x} denotes the time derivative, as usual. The target parameters are set to the values $\beta_1 = 1.0$, $\beta_2 = 1.0$, $\beta_3 = 1.0$, $\beta_4 = 0.2$, $\beta_5 = 0.2$, $\beta_6 = 1.0$, $\beta_7 = 5.7$ which lead to a chaotic attractor.

The phase spaces of the resulting attractors are coarse grained in order to derive a histogram. For this purpose we choose a rectangular (for the 2-dimensional Hénon case) or a cuboid (for the 3-dimensional Rössler case) area in phase space that runs parallel to the axes and encloses the attractor. The edge length of each side of the area symmetrically extends the width of the attractor by 1/10 with respect to the phase space axes. This situation is depicted in Fig. 1 for the Hénon attractor. The maximum widths of the attractor with respect to each axis are denoted

→ Fig. 2. Number of cells (y-axes) with a certain occupation number (x-axes) for $\gamma = 20, 30, 50$ (from top to bottom) for the Hénon attractor.



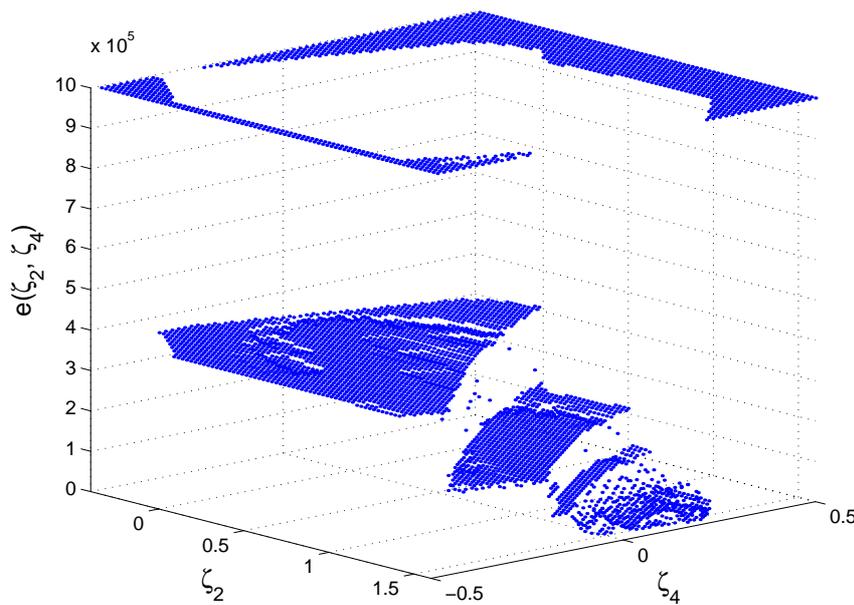


Fig. 3. Dependence of the error function on the variation of two parameters, ζ_2, ζ_4 , for the Hénon case (global view).

by d_x and d_y , respectively. This leads to edge lengths of the surrounding rectangular area of $D_x = 1.2d_x$ and $D_y = 1.2d_y$. The coarse graining $(\Delta x, \Delta y)$ is chosen to be $(\Delta x = D_x/\gamma, \Delta y = D_y/\gamma)$. The grid resulting from $\gamma = 10$ is shown in Fig. 1 as dashed lines. Since it is straightforward we refrain from showing the analogous picture for the Rössler attractor. Of course, in this case the surrounding volume is 3-dimensional.

The number of points in each cell of the grid is counted in order to construct a histogram. Figures 2a–c show the distribution of the occupation number of cells for $\gamma = 20$, $\gamma = 30$ and $\gamma = 50$, respectively, for the Hénon case. The x -axis is the occupation number and the y -axis is the number of cells. Since the attractor is localized in phase space a lot of cells are empty. Therefore, there is a large peak for the occupation number zero. This peak, however, is suppressed in Figs. 2a–c in order to have a visually meaningful scale for the comparison of the more interesting non-empty cells. In the case of $\gamma = 20$ the largest occupation number is 36 whereas for $\gamma = 30$ the largest number is 20 and for $\gamma = 50$ a maximum count of 15 is obtained. We refrain from showing the histograms for the other cases since the tendency can clearly be seen from the depicted cases. A graining finer than 50 does not make sense since most of the cells contain a single point of the attractor. This means that the adjustment pressure for the model attractor is too harsh. In addition, the simple least squares adaptation routine is

no longer applicable. If, to the contrary, the graining is too coarse the information on the spatial structure is lost. We recommend to use $\gamma = 30$ for the Hénon attractor in the given case of 600 data points. Also the Rössler attractor is represented by 600 points. In this case, however, a long time series of 6000 points is computed using a fourth order Runge-Kutta algorithm with step size $\Delta t = 0.05$ whereby each 10th point is kept in order to cover the whole attractor. We mention in passing that we observed no change in the behavior of the estimation procedure when using $\Delta t = 0.1$ for the Runge-Kutta algorithm and 600 subsequent points. In the latter case the trajectory shows roughly ten cycles of the attractor which seems to be enough information. In other words, if the transient phase is skipped the estimation does not depend on initial conditions.

The histogram is a positive distribution, of course. However, the peak at zero can virtually be stretched into the negative direction, so to speak, by using weight functions. Empty cells that are far away from the attractor are weighted stronger than empty neighboring cells. We come back to this point in the next section where we define the error function for the adaptation routine.

3. Error Function for the Hénon Attractor

In a first approach we choose a simple least squares routine for the adaptation of a model attractor to the

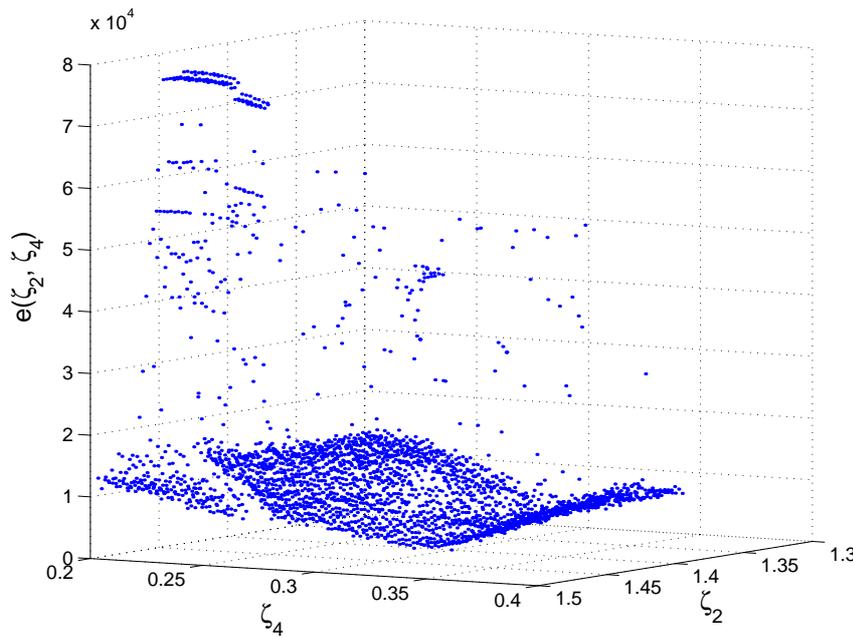


Fig. 4. Dependence of the error function $e(\zeta_2, \zeta_4)$ on the variation of two parameters, ζ_2, ζ_4 , for the Hénon case (local view around the minimum).

data. The number of points in cell (i, j) with $1 \leq i \leq \gamma$ and $1 \leq j \leq \gamma$ is denoted by h_{ij} for the target attractor and h'_{ij} for the model attractor, respectively. The error function reads

$$e(\zeta) = \begin{cases} \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} (h_{ij} - h'_{ij}(\zeta))^2, & \text{if all model points lie inside} \\ \Pi & \text{otherwise,} \end{cases} \quad (3)$$

which, of course, depends on the set of model parameters ζ . The constant Π is a penalty value for diverging model trajectories. This means, if any point of the model trajectory lies outside the searching volume then the corresponding set of parameters will be discarded by attributing a large penalty value to the error function.

The model which is expected to result in an attractor that resembles the Hénon attractor reads:

$$\begin{aligned} x'_{n+1} &= \zeta_1 - \zeta_2 x_n'^2 + \zeta_3 y'_n, \\ y'_{n+1} &= \zeta_4 x'_n. \end{aligned} \quad (4)$$

In order to give a vivid impression of the appearance of the error function in Fig. 3 we stepwise varied two parameters in equidistant steps, $-0.4 \leq \zeta_2 \leq 1.6$ in 80 steps and $-0.5 \leq \zeta_4 \leq 0.5$ in 60 steps, respectively,

whereas the other parameters have been fixed to the target values $\zeta_1 = 1$ and $\zeta_3 = 1$, respectively. Figure 3 shows a global view onto the error function whereas Fig. 4 shows details around the minimum, both for the case of $\gamma = 30$.

The plateaus emerging in the appearance of the error function can easily be explained. Some parameter values lead to fixed point attractors inside the searching region but outside the target attractor. These point attractors are robust against parameter changing within a quite large range of values leading to the same error function value. The plateau at $e(\zeta) = 10^6$ is due to a penalty of this magnitude when the model attractor diverges from the searching region.

The proposed method makes sense only if the dependence of the error function on the initial conditions is below an acceptable threshold. In other words, how does the error function behave when two Hénon attractors with exactly the same parameters but different initial conditions are compared? Figure 5 shows the result of an according calculation where the error function values for 100 model systems with randomly varied initial conditions but exactly the same values for the parameters as the target system are presented. Only one model system has been chosen to have exactly the same initial values as the target which, of course, leads to a vanishing error function value. The cloud of points has an approximative upper limit of 1500. This means,

that the precision in a parameter estimation routine is limited by this value which is, however, well below the bulk of the error function shown in Figs. 3 and 4, respectively. The stop criterium of the estimation procedure has been chosen accordingly.

4. Error Function for the Hénon Attractor Using Weights

The plateaus observed in the simple error function of (3) may lead to a bad convergence of the minimization routine and, as a result, to get stuck in a local minimum. We now define an improved error function using weight coefficients ω_{ij} for a given cell (i, j) . To this end we introduce a distance, s_{ij} , of the given cell to the attractor and weights depending on this distances in the following way:

$$s_{ij} = \min_{k,l} \left(\sqrt{(i-k)^2 + (j-l)^2} \right) \text{ with } h_{kl} \neq 0,$$

$$\omega_{ij} = \begin{cases} 1 & \text{if } h_{ij} \neq 0, \\ cs_{ij} & \text{otherwise} \end{cases},$$

$$e(\zeta) = \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} \omega_{ij} (h_{ij} - h'_{ij}(\zeta))^2. \quad (5)$$

This error function $e(\zeta)$ is capable to suppress the plateaus considerably and smooth the shape in comparison to the unweighted error function. The result can be seen in Fig. 6 for $\gamma = 30$ and $c = 1.0$. The two parameters ζ_2, ζ_4 have stepwise been varied in the same way as explained in the previous section for the unweighted case. One sees that the appearance of the error function is stretched in comparison with the unweighted case. The penalty value, Π , has been set to $5 \cdot 10^6$ which leads to the “ceiling” at the top. This dispersion of the plateaus avoids running into local minima. The convergence behavior of the minimization routine using the weighted error function is much improved. We abstain from showing the local view for this case since it does not supply new insight.

5. Error Function for the Hénon Attractor Using Poisson-Like Distribution

For the case of small occupation numbers the least squares approach in the above definitions of the error function may not be justified. We therefore introduce another error function based on a Poisson distribution for the occupation numbers. The new error function is

defined as follows:

$$e(\zeta) = \begin{cases} \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} \omega_{ij} \sum_{k=1}^{h'_{ij}(\zeta)} \log k - \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} h'_{ij}(\zeta) \log h_{ij}, & \text{if } h_{ij} > 0 \wedge h'_{ij} > 0, \\ - \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} h'_{ij}(\zeta) \log h_{ij}, & \text{if } h_{ij} > 0 \wedge h'_{ij} = 0, \\ \sum_{i=1}^{\gamma} \sum_{j=1}^{\gamma} \omega_{ij} \sum_{k=1}^{h'_{ij}(\zeta)} \log k, & \text{if } h_{ij} = 0 \wedge h'_{ij} > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

In this equation, the same weights, ω_{ij} , as defined in (5) are used. Equation (6) is the result of (negatively) logarithmizing the Poisson distribution (log-likelihood) $\frac{h'^h}{h!} e^{-h}$, whereby the factorial leads to the summation $\sum_{k=1}^{h'} \log k$. Figure 7 has been produced using $c = 1.0$ for the weights. The plateau resulting from the penalty for diverging model attractors is not shown in order to focus on the relevant part of the error function.

6. Results for the Hénon Attractor

For the minimization of the error function with respect to the parameters we used the Nelder-Mead simplex method along with some modifications as described in [4]. From this paper we derived our own simple code implemented in the MATLAB 6.0 environment. This means, that we abstained from highly elaborated tricks to improve the convergence behavior that may come along with a commercial software.

The procedure is as follows. For each suggested set of parameter values a model attractor as well as the corresponding histogram and therefrom the error function according to (3) or (5), respectively, is computed. In case the model dynamics produces a point outside the searching volume the computation of the time series is immediately interrupted and the error function set to a penalty value, Π , as defined in (3). This saves a lot of computational time. It is, therefore, also important to choose a proper searching volume which is not too large but also not too small since in the latter case the restriction is too harsh. Since the transient phase is skipped for each time series of the model sys-

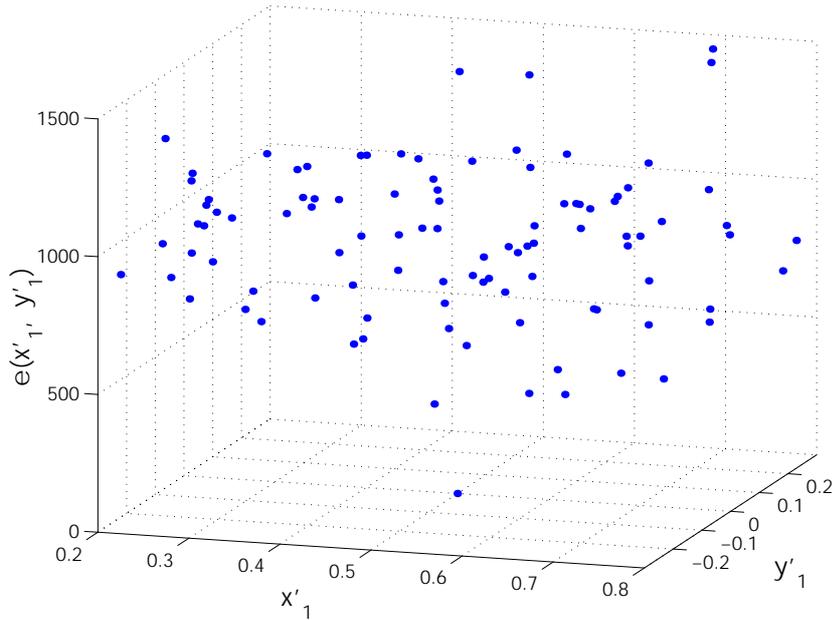


Fig. 5. Dependence of the error function on the initial conditions for the Hénon case. 100 model systems have been simulated with initial conditions (x_1, y_1) randomly varying around $x_1 = 0.5$ and $y_1 = 0$, respectively. Only one of the model systems has exactly the same initial conditions as the target system, $x_1 = 0.5, y_1 = 0$, which leads to a vanishing error function. The calculation is based on 600 subsequent iterations of the map after 250 time steps in the beginning (transient behavior) have been skipped. The coarse graining parameter is $\gamma = 30$.

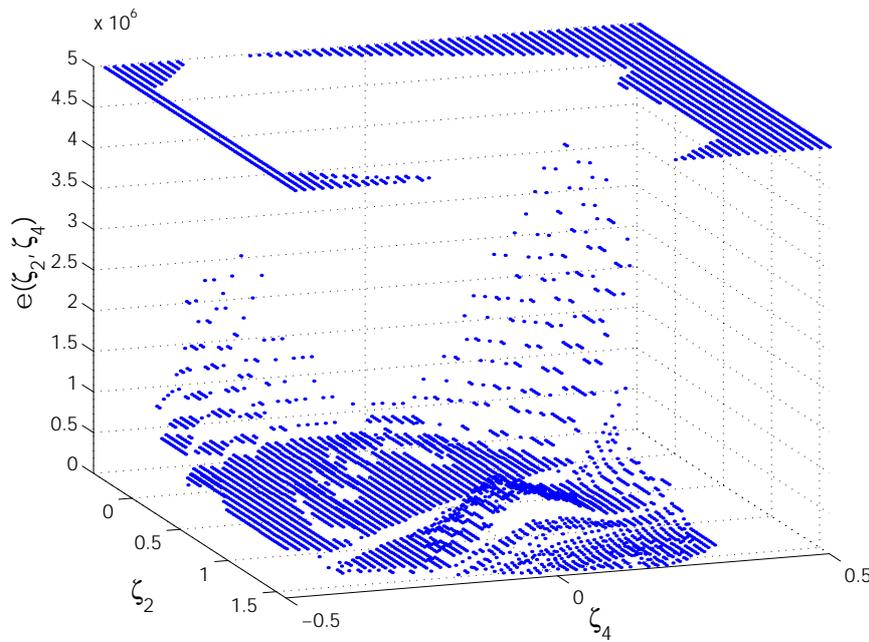


Fig. 6. Dependence of the error function on the variation of two parameters, ζ_2, ζ_4 , for the Hénon case using weights defined in (5).

tem the initial value of the dynamics is relatively unimportant unless not chosen extremely far from the target attractor. We indeed used one of the points of the target attractor to start the model dynamics. Given the correct parameter value for the model system the error function should vanish. The worst model for the un-

weighted error function is given by a point attractor inside the searching volume but not on the target attractor which leads to an error function value of slightly more than $N^2 = 600^2$ (where 600 is the length of the time series). We have chosen $s = 2 \cdot 10^{-3}N^2$ as a stop criterion for the minimization which turned out to be an

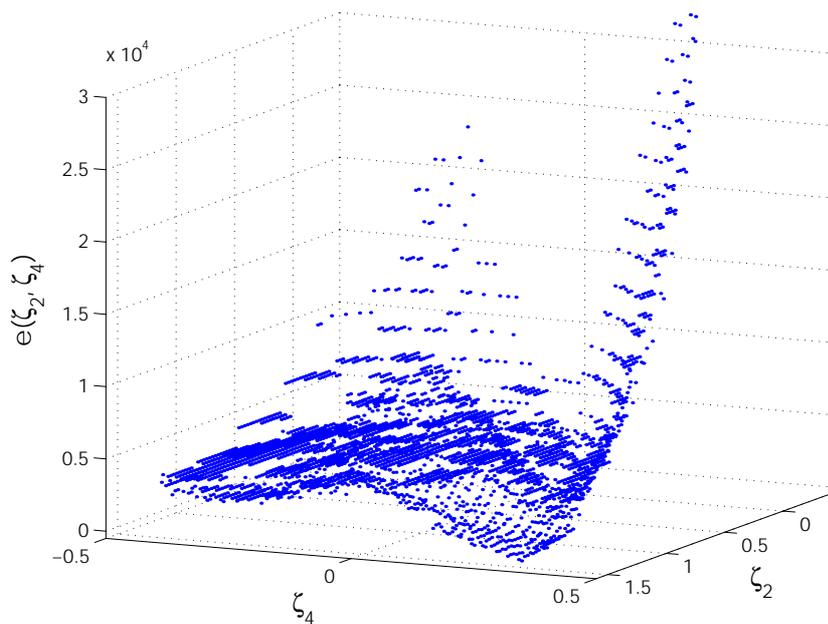


Fig. 7. Dependence of the error function on the variation of two parameters, ζ_2, ζ_4 , for the Hénon case using a Poisson-like distribution defined in (6).

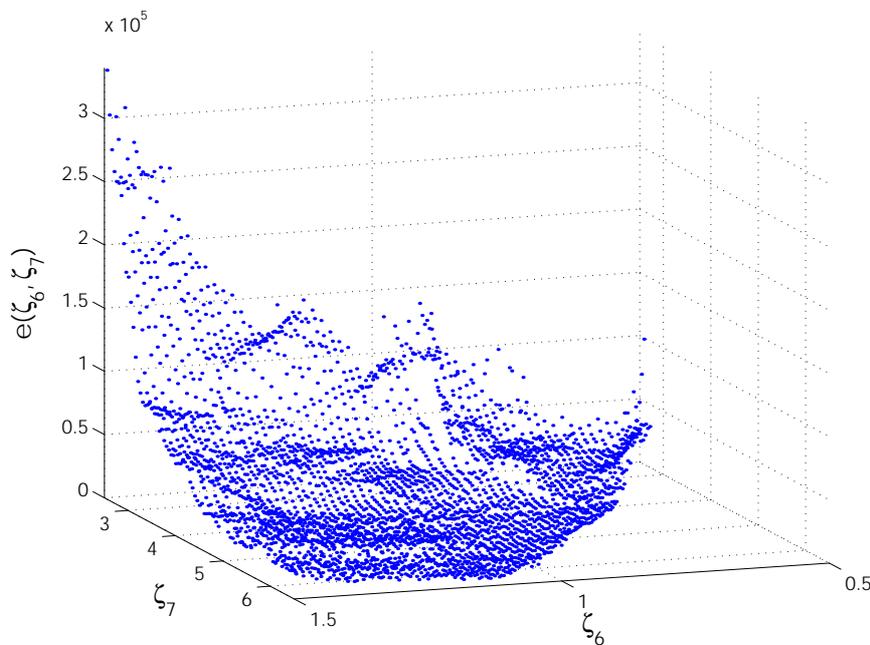


Fig. 8. Dependence of the weighted error function on the variation of two parameters, ζ_6, ζ_7 , for the Rössler case.

appropriate value (please compare with Fig. 5). This holds also for the weighted case according to our experience, although the weights have a slight impact on the maximum error function value.

The estimation is quick and robust for at least three parameters. The time consumption is about 15 minutes with a 1.6 GHz “Pentium M” laptop computer. We also

estimated six parameters by adding two terms to the Hénon system:

$$\begin{aligned} x'_{n+1} &= \zeta_1 - \zeta_2 x_n'^2 + \zeta_3 y_n', \\ y'_{n+1} &= \zeta_5 + \zeta_4 x_n' + \zeta_6 y_n'. \end{aligned} \tag{7}$$

In the latter case of six parameters it was necessary to

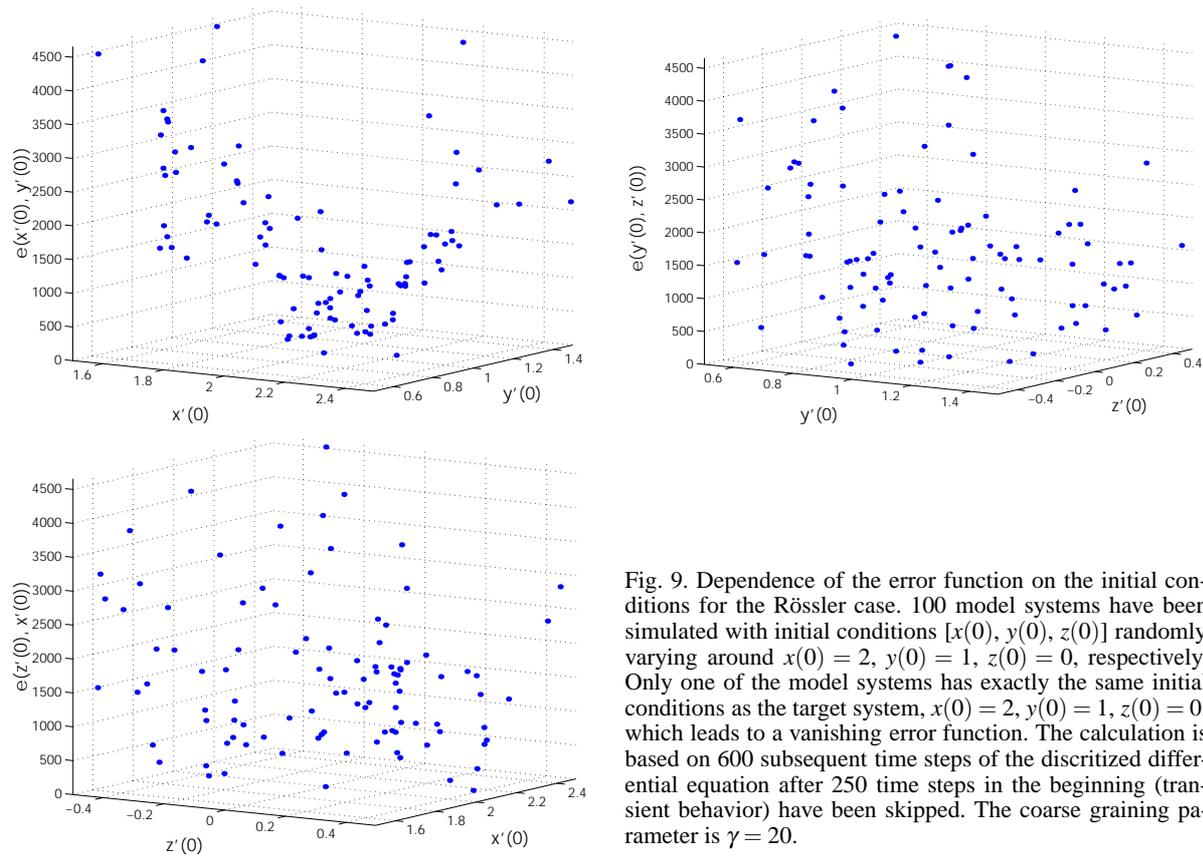


Fig. 9. Dependence of the error function on the initial conditions for the Rössler case. 100 model systems have been simulated with initial conditions $[x(0), y(0), z(0)]$ randomly varying around $x(0) = 2, y(0) = 1, z(0) = 0$, respectively. Only one of the model systems has exactly the same initial conditions as the target system, $x(0) = 2, y(0) = 1, z(0) = 0$, which leads to a vanishing error function. The calculation is based on 600 subsequent time steps of the discretized differential equation after 250 time steps in the beginning (transient behavior) have been skipped. The coarse graining parameter is $\gamma = 20$.

start the minimization routine several times with randomly chosen initial values for the parameters. One has to calculate with at least two hours for this estimation. For some unluckily chosen initial values the routine took roughly three hours to reach the minimum since obviously the plateaus retarded the convergence. In some cases it happens that the minimization gets stuck in a local minimum as a result of the plateaus which necessitates either restarts or a more sophisticated escape procedure within the minimization routine. The weighted error function, however, led to an enormous improvement in this respect. The best results turned out from a coarse graining using $\gamma = 30$. The estimated parameters are identical to the original ones up to three digits. We did not implement the computation of a goodness of fit measure so far which has of course to be done in a future version of the routine. Since in our case the original parameters are available we abstained from deriving such a measure in this paper.

7. Application to the Rössler Attractor

The model equation for the Rössler system reads:

$$\begin{aligned} \dot{x}'(t) &= -\zeta_1 z'(t) - \zeta_2 y'(t), \\ \dot{y}'(t) &= \zeta_3 x'(t) + \zeta_4 y'(t), \\ \dot{z}'(t) &= \zeta_5 + \zeta_6 x'(t) z'(t) - \zeta_7 z'(t). \end{aligned} \quad (8)$$

The error function for this case is a straightforward extension of (3) and (5), respectively, from two to three dimensions. Therefore, we refrain from explicitly showing the equations.

The appearance of the weighted error function can be seen in Fig. 8. For this figure, the two parameters ζ_6 and ζ_7 have been varied in 60 and 80 equidistant steps, respectively, in the range of $0.5 \leq \zeta_6 \leq 1.5$ and $2.5 \leq \zeta_7 \leq 6.5$. The coarse graining factor has been chosen to be $\gamma = 20$. The weighted error function (which is shown in Fig. 8) looks similar to the unweighted case (not shown). It is merely stretched by a factor that depends on the weight coefficient, c , enter-

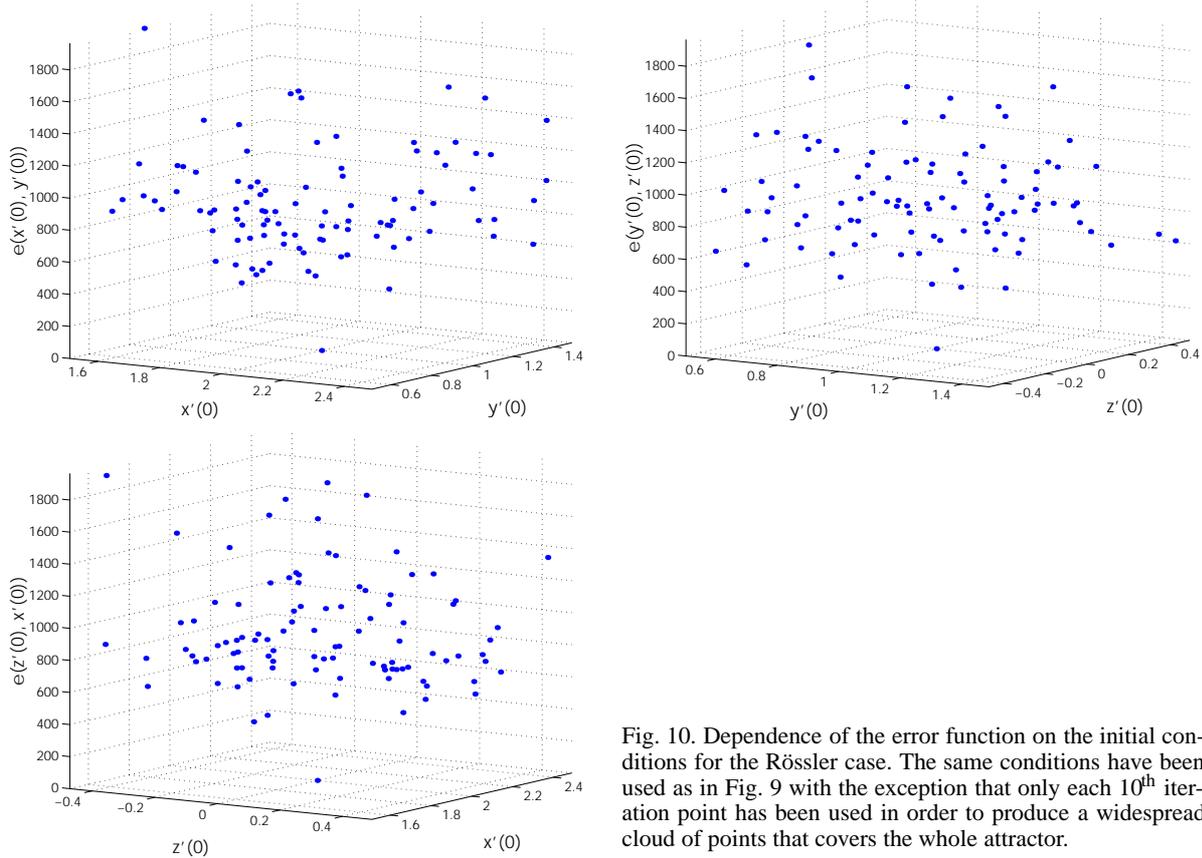


Fig. 10. Dependence of the error function on the initial conditions for the Rössler case. The same conditions have been used as in Fig. 9 with the exception that only each 10th iteration point has been used in order to produce a widespread cloud of points that covers the whole attractor.

ing the weights [cf. (5)]. For the computation underlying Fig. 8, $c = 20$ has been chosen. The minor influence of the weights on the global shape of the error function is due to the fact, that all model attractors lie on or in the vicinity to the target attractor. The parameter ζ_7 is a bifurcation parameter leading to a period-doubling behavior when increased from a small to a larger value. The resulting periodic solutions, however, approximately lie on the original chaotic attractor. The other parameter, ζ_6 , has merely a scaling behavior of the attractor. We point out, however, that the weights make sense nevertheless when estimating more than two parameters.

As already pointed out for the Hénon attractor, the proposed method makes sense only if the dependence of the error function on the initial conditions is below an acceptable threshold. Figures 9 and 10 show the result of a calculation where the error function values for 100 model systems with randomly varied initial conditions but exactly the same values for the parameters as the target system are presented. In Fig. 9 the time se-

ries consists of 600 subsequently calculated states with a time step of $h = 0.05$ used in the Runge-Kutta 4th-order procedure. Figure 10, in contrast, has been produced using only each 10th step of the numerical iteration in order to distribute the 600 points onto the whole attractor. This leads to a reduction of the upper border of the error function to less than half (~ 1800) of the limit for the first case (~ 4500). Again, like in the Hénon case, the cloud of points is well below the bulk of the error function shown in Figure 8. The stop criterion of the estimation procedure has been chosen accordingly.

For the estimation procedure we stick with $\gamma = 20$ and $c = 20$. The time consumption for the estimation of two parameters is less than half an hour. Again, the values of the parameters are equal to the original values up to at least three digits. The estimation of more than two parameters leads to an enormous increase of computational time. Three parameters are manageable in less than three hours. Four parameters need several hours, strongly dependent on the initial guess. How-

ever, the case of five parameters leads to an almost unacceptable time consumption in the order of magnitude of twenty hours, also depending on the initial guesses for the parameter values. As already mentioned for the Hénon attractor, in some cases of unluckily chosen initial values some restarts with randomly chosen parameter values are necessary.

8. Conclusions

We presented a procedure for parameter estimation for nonlinear dynamical systems based on a structural comparison in phase space. The shape of the model attractor that contains a set of free parameters is adapted to the target attractor. This spatial algorithm has some advantages compared to a temporal approach. Many techniques of temporal adaptations are based on synchronization. As we mentioned, it frequently happens that a diverging system can be easily synchronized to a target system leading to a non-acceptable model. Our algorithm in contrast focuses on the conservation of invariant structures in phase space. Independently of the

precision of the estimated parameters, all models resemble the target attractor. The method cannot result in a diverging attractor due to the structural approach.

Up to four parameters of two- or three-dimensional dynamical systems can be estimated acceptably fast and robustly. We note however, that some efforts put into the enhancement of the minimization routine and – most important – a combination with temporal approaches may improve the estimation efficiency considerably. For the case of time series data we expect that our method can be applied to attractor reconstructions and, therefore, considerably contributes to time series analyses.

Acknowledgements

We thank Ioan Grosu and Florian Grond for valuable discussions. This work results from a five months residence of K.M. at the Center for Art and Media (ZKM), Karlsruhe, Germany, from August till December 2005, supported by Ichiro Tsuda, Hokkaido University, Japan, and the ZKM.

- [1] H. Kantz and Th. Schreiber, *Nonlinear Timeseries Analysis*, Cambridge University Press, Cambridge 1997.
- [2] I. Grosu, *Int. J. Bifurcation and Chaos* **14**, 2133 (2004).
- [3] H.H. Diebner, A.A. Hoff, A. Mathias, H. Prehn, M. Rohrbach, and S. Sahle, *Z. Naturforsch.* **56a**, 663 (2001).
- [4] J.C. Lararias, J.A. Reeds, M.H. Wright, and P.E. Wright, *SIAM J. Optimization* **9**, 112 (1998).