

Reconstruction of nonlinear time delay models from data by the use of optimal transformations

Henning Voss and Jürgen Kurths,
Institute of Physics,
University of Potsdam, D-14415 Potsdam, Germany

Physics Letters A **234** (1997) 336–344

Abstract

We propose a new technique for the numerical reconstruction of nonlinear delay differential equations from time series by applying the method of optimal transformations, a concept of multiple nonlinear regression analysis. By constructing a generalized correlation function, this method allows for testing time series for delay-induced dynamics. Also, the delay times and the governing differential equations are estimated with good numerical accuracy. We present several examples which show that this method is useful also for the analysis of short and noisy time series.

PACS numbers: 07.05.K

Keywords: Nonlinear time series analysis; delay differential equations; physiological feedback models; optimal transformations; maximal correlation

Time-delay induced instabilities, as described by univariate delay differential equations (DDEs), play an important role in modeling natural phenomena. Such models are used in many different scientific disciplines, like hydrodynamics, laser physics, physiology, engineering-, economic-, and cognitive sciences (cf. [1, 2] and references therein). Recently these delays have received special attention due to the fact that in conceptual models in physiology order parameter equations can often be expressed as DDEs, see also Ref. [2]. Most of the recent research in nonlinear dynamical systems has been done with respect to low-dimensional ordinary differential equations (ODEs). In close parallelism to that, methods of nonlinear time series analysis have been developed. The bulk of them consider the analysis of univariate time series with respect to ODEs, based on Takens' embedding theorem [3]. However, nonlinear instabilities can also be induced by a time-delayed feedback, modeled by nonlinear DDEs, and leading to a similar variety of dynamical behavior as in the case of ODEs [4]. Therefore, there is also a challenge for analysis methods with respect to DDEs; those concepts are essential for the comparison of time delay models with experimental data.

Dynamical systems given by a DDE have an infinite-dimensional state space, and attractors of the solutions can also be high-dimensional. By evaluating a

method for the identification of such systems, given only a single and finite solution, we demonstrate that these properties do not constitute severe limitations for time series analysis. The first method which satisfies these requirements has recently been proposed by Bünner et al. [1, 5]. It recovers functions with a single time delay by projection of the phase space to suitable chosen subspaces. The new technique presented here differs essentially from that method. It is based on multiple nonlinear regression analysis and allows to judge whether the data are in agreement with a description by a nonlinear feedback model with a single or even with a few time delays. Also, the governing nonlinear equations can be estimated. We found the nonlinear regression method to be very suitable for the analysis of data with non-optimal properties. Therefore, this method is well-suited to time series from experimental data — it works well also for rather short, noisy, and coarsely sampled time series.

The organization of this letter is as follows: First, we explain the aim of our analysis, then we describe the mathematics of the estimation procedure in general form before applying it to the reconstruction of DDEs, followed by some examples especially designed to reveal the potentials and limits of the method.

We describe a method to reveal the time delays and also the form of the governing differential equation from experimental data, for which one can assume from physical reasoning that they might be produced by systems with a single or multiple time delay. Even for data for which one cannot make this assumption, the method gives a quantitative criterion whether such an assumption might be justified. Although time series produced by such systems can also successfully be analyzed using embedding methods (see e.g. Ref. [6]), such an analysis would not reveal the natural properties of the system, like the importance of a time-delayed feedback; DDEs can also produce chaotic attractors with a low dimension [4], but the natural description would not be an ODE. Note that it is impossible to gain clear information about the delay times out of the time series by methods like spectral analysis or different kinds of nonlinear autocorrelation functions, as calculated in Ref. [7].

We consider the reconstruction of DDEs of the form

$$h(\dot{x}(t)) = f_0(x(t)) + f_1(x(t - \tau)) \quad (1)$$

for a single delay τ , and

$$h(\dot{x}(t)) = f_0(x(t)) + \sum_{i=1}^k f_i(x(t - \tau_i)) \quad (2)$$

for k coexisting delays. The functions h and f_0, \dots, f_k are assumed to be continuous. In most models, h is the identity. By “reconstruction” we precisely mean the numerical estimation of the functions rather than finding a model in a closed form. The dynamics of Equation (1) has to be described in an infinite-dimensional phase space, since one has to supply a whole interval $[x(t_0 - \tau), x(t_0)]$ on the real axis as initial condition for the unique evolution of the system from time t_0 on $[1, 2]$. In case of more than one delay, the initial condition has to

be specified by the interval $[x(t_0 - \tau_k), x(t_0)]$, where τ_k is the largest of the involved delays. However, in spite of the infinite dimensionality, the data analysis can be performed without using Takens' embedding theorem: By Equation (2) the time-evolution of the $(k+2)$ -tuple $(\dot{x}(t), x(t), x(t - \tau_1), \dots, x(t - \tau_k))$ is restricted to a $(k+1)$ -dimensional manifold. This is also reflected in each finite realization of the DDE, $\{y_t\}_{t=1}^N$, with an arbitrary initial condition: The time-evolution of the tuple $(\Delta y_t, y_t, y_{t-\tau_1}, \dots, y_{t-\tau_k})$, where $\dot{x}(t)$ has been approximated by $\Delta y_t := (y_{t+1} - y_{t-1})/2$, is also restricted approximately to the same $(k+1)$ -dimensional manifold via

$$\tilde{h}(\Delta y_t) = \tilde{f}_0(y_t) + \sum_{i=1}^k \tilde{f}_i(y_{t-\tau_i}) , \quad (3)$$

with the unit time step $\Delta t = 1$. The functions $\tilde{h}, \tilde{f}_0, \dots, \tilde{f}_k$ are then approximations of h, f_0, \dots, f_k in case of knowing only a finite realization $\{y_t\}_{t=1}^N$ of the process. Viewing Equation (3) as a defining equation for the unknown functions $\tilde{h}, \tilde{f}_0, \dots, \tilde{f}_k$, it expresses the inverse problem to Equation (2), i.e., the problem of estimating the DDE from data. It can be solved by estimating so-called optimal transformations.

These can be found by the multidimensional case of the *alternating conditional expectation algorithm* (ACE), as described in Ref. [8, 9]. It estimates optimal transformations for multiple nonlinear regression:

Given two random variables X and Y with zero expectation, $E[X] = E[Y] = 0$, the real-valued measurable mean-zero functions $\Theta(Y)$ and $\Phi(X)$ are called *optimal transformations*, if they satisfy

$$\frac{E[(\Theta(Y) - \Phi(X))^2]}{E[\Theta^2(Y)]} \stackrel{!}{=} \min . \quad (4)$$

This is equivalent to the maximization of the correlation coefficient between the transformed variables $\Theta(Y)$ and $\Phi(X)$. The simple proof is also given in [8]. In other words, we transform X and Y by suitable, generally *nonlinear*, transformations such as to yield a *linear* relationship between the new random variables $\Theta(Y)$ and $\Phi(X)$. The value

$$\Psi(X, Y) := \sup_{\Theta, \Phi} |R(\Theta(Y), \Phi(X))| , \quad (5)$$

with the correlation coefficient

$$R(X, Y) = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2]E[Y^2]}} , \quad (6)$$

is called *maximal correlation* between the “predictor” X and the “response” Y [10, 11]. Therefore, as an equivalent definition to Equation (4), optimal transformations are defined as those functions Θ^* and Φ^* for which the supremum of Equation (5) is achieved:

$$\Psi(X, Y) = |R(\Theta^*(Y), \Phi^*(X))| . \quad (7)$$

This concept can also be extended to higher-dimensional problems, specifically more than one predictor variable X , i.e.,

$$\frac{E[(\Theta(Y) - \sum_{i=0}^k \Phi_i(X_i))^2]}{E[\Theta^2(Y)]} \stackrel{!}{=} \min. \quad (8)$$

It can be shown that such optimal transformations do exist generally. The ACE algorithm to calculate sample-estimates for the optimal transformations, $\hat{\Theta}$ and $\hat{\Phi}$, converges for a variety of representations for Θ and Φ [8].

Here we give only a short description of the algorithm, in the form as implemented by us, using partly the same notation as in Ref. [8]. Without loss of generality, we assume that $E[\Theta(Y)] = E[\Phi_0(X_0)] = \dots = E[\Phi_k(X_k)] = 0$. The ACE algorithm for the one-dimensional case (4) works as follows: The function

$$\bar{\Phi}(X) = E[\Theta(Y)|X] \quad (9)$$

minimizes (4) with respect to $\Phi(X)$ for given $\Theta(Y)$. Similarly,

$$\bar{\Theta}(Y) = E[\Phi(X)|Y]/\|E[\Phi(X)|Y]\|, \quad (10)$$

where the norm is defined as $\|Z\| = \sqrt{E[Z^2]}$, minimizes (4) with respect to $\Theta(Y)$ for given $\Phi(X)$, keeping $E[\Theta^2(Y)] = 1$.

This motivates the ACE algorithm as an iterative procedure. Starting with the initial function

$$\Theta_1(Y) = E[X|Y], \quad (11)$$

from $i = 2$ on calculate

$$\Phi_i(X) = E[\Theta_{i-1}(Y)|X] \quad (12)$$

and

$$\Theta_i(Y) = E[\Phi_i(X)|Y]/\|E[\Phi_i(X)|Y]\|, \quad (13)$$

until $E[(\Theta_i(Y) - \Phi_i(X))^2]$ fails to decrease. Suppose this happens for $i = n$. Then $\Theta_n(Y)$ and $\Phi_n(X)$ are estimates for optimal transformations.

The ACE algorithm decreases $E[(\Theta_i(Y) - \Phi_i(X))^2]$ at each step. That it converges to optimal transformations $\Theta^*(Y)$ and $\Phi^*(X)$ is proved in [8]. For the minimization of Equation (8) one can use a double-loop algorithm [8]. In the newly introduced inner loop with the loop variable $j = 0 \dots k$ the quantities

$$\Phi_{j,i}(X_j) = E[\Theta_{i-1}(Y) - \sum_{p \neq j} \Phi_p(X_p) | X_j] \quad (14)$$

are calculated.

There are several possibilities to estimate conditional expectations from finite data sets and to represent the functions $\Theta, \Phi_0, \dots, \Phi_k$. To calculate conditional expectations from finite samples $x := \{x_i\}_{i=1}^N$, $y := \{y_i\}_{i=1}^N$, we use a special method of adaptive partitioning the data, which proved to be very successful in applications, as we have found: We first transform the samples x, y as to be

uniformly distributed on the interval $[0, 1)$, by calculating relative ranks [13]. The advantage of this is that after a partitioning of the range of x and y , each bin contains the same number of values (up to deviations of at most ± 1). This allows for a most accurate estimation of expectation values, independent of the form of the data distribution.

This transformation U of x, y to uniformly distributed samples, $U(x), U(y)$, is monotonic, because it preserves the ranking of values: If $x_i > x_j$, then $U(x_i) > U(x_j)$. It can easily be shown [11] that the maximal correlation is invariant under monotonic transformations, such that the results for $\Psi(\tau_1, \dots, \tau_k)$ in Equation (16) are not affected. Because U is invertible, it is also no problem to yield the true functions $\Theta(y), \Phi(x)$ back from the transformed results $\Theta(U(y)), \Phi(U(x))$. To calculate a function $\Theta(Y) = E[X|Y]$, as necessary for Equation (11), we therefore estimate $\Theta(U(y))$. For the sake of simplicity, from now on we treat y as uniformly distributed on the unit interval, such that we can drop the function U in the following. We now approximate the function $\Theta(Y)$ by a piecewise-linear function by the ansatz

$$\Theta(Y) = a_i Y_i + b_i \text{ for } \{Y_i \in [(i-1)\varepsilon, i\varepsilon)\}_{i=1}^{n_\varepsilon}, \quad (15)$$

n_ε the number of bins of the unit interval with constant bin-size $\varepsilon = 1/n_\varepsilon$. Now $\Theta(Y)$ can be fitted piecewise to the data $(x_i, y_i) \in (x, y)$ ($i = 1, \dots, N$). To calculate a function like $E[\Phi(X)|Y]$, as necessary for Equation (12), we use the same ansatz (15), but now taking for the data $(x_i, \Theta(y_i)) \in (x, \Theta(y))$. For the examples of this paper ε is always chosen such that each bin contains at least 20 values to allow for a proper statistics.

The reconstruction process consists of two parts: 1. Finding the number of involved delays, k , and the estimates for the delays, $\hat{\tau}_1, \dots, \hat{\tau}_k$. 2. Determining the function estimates, $\hat{h}, \hat{f}_0, \dots, \hat{f}_k$.

Until further notice, assume that one already knows k and τ_1, \dots, τ_k . The above method for finding optimal transformations can now be applied to solve the problem of finding the functions in Equation (3): If one solves Equation (8) for $Y \rightarrow \Delta y_t, X_0 \rightarrow y_t, X_i \rightarrow y_{t-\tau_i}$ ($i = 1, \dots, k$), the estimates $\hat{\Theta}, \hat{\Phi}_0, \dots, \hat{\Phi}_k$ can serve as estimates $\hat{h}, \hat{f}_0, \dots, \hat{f}_k$ for the functions $\tilde{h}, \tilde{f}_0, \dots, \tilde{f}_k$ of Equation (3). As a quantitative criterion how well the estimated functions $\hat{h}, \hat{f}_0, \dots, \hat{f}_k$ fit Equation (3), we use the absolute value of the correlation between $\hat{h}(\Delta y_t)$ and $\hat{f}_0(y_t) + \sum_{i=1}^k \hat{f}_i(y_{t-\tau_i})$,

$$\Psi(\tau_1, \dots, \tau_k) := |R(\hat{h}(\Delta y_t), \hat{f}_0(y_t) + \sum_{i=1}^k \hat{f}_i(y_{t-\tau_i}))|. \quad (16)$$

This definition is chosen in analogy to Equation (7) for the maximal correlation, yet the same symbol Ψ . Therefore, $\Psi(\tau_1, \dots, \tau_k)$ can be viewed as a generalized correlation function. It measures correlations which are explicitly caused by the time delays involved. The value of $\Psi(\tau_1, \dots, \tau_k)$ is restricted to the interval $[0, 1]$ by definition, as is the maximal correlation. For brevity we call $\Psi(\tau_1, \dots, \tau_k)$ also maximal correlation, although this term is actually restricted to the bivariate

case, Equation (5). The upper limit, $\Psi = 1$, can only be achieved for a perfect realization of Equation (2), i.e., a noise-free, infinitely long and arbitrarily high sampled time series. Thus, Ψ constitutes a measure for the closeness of the approximate manifold (3) and the exact one defined by the DDE (2). To avoid too many indices, in the following we restrict ourselves to a maximum number of two delays. The extension to a larger number will be straightforward from what follows.

The first part of the reconstruction process is now solved as follows: Estimate $\Psi(\tau_1, \tau_2)$ for every combination of $\tau_1 = 1, \dots, \tau_m$ and $\tau_2 \leq \tau_1$, where τ_m is the largest delay to test for. If the maximum of $\Psi(\tau_1, \tau_2)$ is close to one, its location points to an estimate for both delays. If the maximum is located at (τ_1, τ_2) with $\tau_1 = \tau_2$, then one concludes that only a single delay is of importance. If none of the $\Psi(\tau_1, \tau_2)$ appears to be close to one, one concludes that the time series can not be the solution of a DDE of the form (2) with $k \leq 2$. In practice, if one knows from the beginning that only a single delay is important, or if one wants to compare data with a one-delay model, to keep computational effort low one would use Equation (16) with $k = 1$, computing only $\Psi(\tau_1)$. What “close to one” means exactly depends, of course, on subjective judgment, taking all circumstances — like length of the time series, sampling rate, noise level, quality of measurement, and the used ACE implementation — into account. A deeper examination of those parameter dependence shall not be a subject in this letter, except an exemplary study of the noise behavior below.

In the following three examples we first demonstrate the method by using a standard example, the Mackey-Glass equation, for the reconstruction of DDEs with one and two delays. Then the method is applied to data from a physiological model with different parameters, to give an impression of both the robustness against different dynamical behavior and the sensitivity to slight delay variations. All examples are designed for rather few data points to demonstrate that the method can be applied also to experimental data which are often limited severely in length; throughout this paper, we use only time series with 500 data points.

1. In Figure 1 we apply the method to a time series produced by the Mackey-Glass DDE

$$\dot{x}(t) = -bx(t) + \frac{ax(t-\tau)}{1+x(t-\tau)^c}, \quad (17)$$

with $a = 0.2$, $b = 0.1$, $c = 10$, the standard parameters, and a delay of $\tau = 300$, which produces a dynamic on a high-dimensional chaotic attractor [4]. In this case, $f_0(x(t)) = -bx(t)$ and $f_1(x(t-\tau)) = \frac{ax(t-\tau)}{1+x(t-\tau)^c}$. The time series is sampled down to the natural time scale of the DDE, such that 300 points in the time series cover a period of time equal to the delay time $\tau = 300$. The time series has a length of only 500 points (Figure 1a), such that it is not only sampled very coarsely but also rather short. Note that for the largest tested delay, $\tau = 400$, there are effectively only 99 triples $(\Delta y_t, y_t, y_{t-400})$ left for the delay estimation procedure. Thus we get the oscillations of $\Psi(\tau)$ for large τ (Figure 1b). However, $\Psi(\tau)$ still exhibits a clear maximum for $\hat{\tau} = 300$. The regression of the variables

$\Delta y_t, y_t, y_{t-\tau}$ after transforming them by the estimated optimal transformations is also shown (Figure 1c,d). For the estimated delay the regression is nearly linear (Figure 1c), whereas for a slightly non-optimal delay of $\tau = 297$ there is a considerable deviation to a linear regression curve (Figure 1d). This is also reflected by the values of the maximal correlation: $\Psi(297) = 0.925$, $\Psi(300) = 0.991$.

Next, we reconstruct the functions $f_0(x(t)), f_1(x(t - \tau))$ using the estimates $\hat{h}(\Delta y_t), \hat{f}_0(y_t), \hat{f}_1(y_{t-\hat{\tau}})$ which we got already from the delay estimation. The function $\hat{h}(\Delta y_t)$ appears to be linear, and the inverse is applied to $\hat{f}_0(y_t), \hat{f}_1(y_{t-\hat{\tau}})$ to yield the correct estimates for Equation (17). Note that for the reconstruction of each of the functions there are only 199 triples $(\Delta y_t, y_t, y_{t-300})$ available. In Figures 1e,f the estimated functions are shown, together with the true functions. Since $(f_0 + c) + (f_1 - c) = f_0 + f_1$, both functions can only be estimated up to an arbitrary constant c , which is adjusted such as to minimize the differences between the estimates and the true functions. In practice this arbitrariness is no drawback since one is only interested in the complete right hand side of Equation (1). As a result, the linearity of f_0 and the nonlinear shape of f_1 are reconstructed well. The deviations appear due to the small sample size: The deviation of $f_0(y_t)$ for $y_t < 0.35$ (Figure 1e) and of $f_1(y_{t-\tau})$ for $y_t > 1.33$ (Figure 1f) are both caused by the fact that the time series covers for most values only the range $[0.35, 1.33]$ (Figure 1a), such that values outside are hard to estimate. This is a general problem of estimation and not of this method. Of course, one cannot estimate the functions for values which are not met by the time series at all.

To investigate the influence of noise, we perform the same analysis to the data heavily corrupted by additive gaussian white noise, namely with a standard deviation of 5, 10, and 20% of the standard deviation of the data. As a result, the maximal correlation is decreased (Figure 2a), but the shape of the graph $\Psi(\tau)$ remains similar up to a noise level of 10%, still allowing for a correct delay estimation. For a moderate noise level even the function reconstructions can be performed well, as shown for the reconstruction of the nonlinear part of the model equation in Figure 2b. For even higher noise levels the functions can still be reconstructed approximately (Figure 2c,d). This influence of noise is uncritical in the sense that it decreases with the length of the considered time series; the calculation of the optimal transformations uses only conditional expectation values, which can be estimated more properly the more data points are available.

2. The next example is a generalization of the Mackey-Glass equation by introducing a further delay:

$$\dot{x}(t) = -bx(t) + \frac{1}{2} \sum_{i=1}^2 \frac{ax(t - \tau_i)}{1 + x(t - \tau_i)^c} \quad (18)$$

with $\tau_1 = 23$, $\tau_2 = 80$, and the same parameters as in the first example. The delays are chosen such that Equation (18) produces a chaotic behavior, as the single-delay equation of the first example, but now with a shorter characteristic

time-scale. For a sample length of 500 points, for $\Psi(\tau_1, \tau_2)$ we yield the maximum value $\Psi(23, 79) = 0.962$, pointing, with a small estimation error due to the short length and the coarse sampling of the data, to the right delay times. For the reconstruction of the functions $f_0(x(t))$, $f_1(x(t - \tau_1))$, $f_2(x(t - \tau_2))$ of Equation (18) one now has to calculate $\hat{h}(\Delta y_t)$, $\hat{f}_0(y_t)$, $\hat{f}_1(y_{t-23})$, $\hat{f}_2(y_{t-79})$ altogether, by Equation (8) for $k = 2$. Again, $\hat{h}(\Delta y_t)$ is linear, and all functions show the expected shape with about the same accuracy as in the first example without noise.

We performed several numerical calculations to check whether one could mis-specify a model by starting the procedure with one delay less than in the model. We found that the maximal correlation points always also to the missing delay by a second peak. Therefore, one is led to perform the method with an additional delay term and decide by the value of the maximal correlation which model is the better estimate. Additionally, the estimated functions have still the right shape if a delay term is missing in the ansatz.

We finish this example with a note to the requirements for the data length: Whereas in the first example four chaotic “orbits” (Figure 1a) have been sufficient to reconstruct the model properly, here the time series contains about nine orbits. By examination of further examples of the Mackey-Glass equation with one and two delays, we found, as a rule of thumb, that for a proper estimation of both the delays and the model functions, it is sufficient that the time series contains about four to five orbits in the case of one, and twice as much in the case of two delays. This, for at least the same sampling rate as in these examples and for sufficiently separated delays. If the aim is to reveal only the delay time but not the model functions, even a less number of orbits is necessary, as will be shown in the next example.

3. As a final example we apply the method to data produced by an order parameter model. A typical order parameter model which is given by a DDE with the delay as one of the model parameters is the model of Tass et al. [12]:

$$\dot{x}(t) = -\alpha \sin\left(x(t) - \frac{\omega\tau}{2}\right) - \beta \sin\left(x(t - \tau) + \frac{\omega\tau}{2}\right) \quad (19)$$

where $\alpha = 2$, $\beta = 3$, $\omega = \pi$. This model describes the interaction of the visual and proprioceptive feedback of a human subject performing a certain tracking task; x is a phase difference and τ a fixed artificial delay. Equation (19) shows a rich variety of dynamical behavior. In Figure 3 the dynamics for nine different values of τ is visualized by the phase portraits of the realizations $\{y_t\}_{t=1}^N$ with $N = 2000$. Here the distance of two data points is 0.02 time units, or a time unit corresponds to 50 data points. The values of τ are chosen from three qualitative different dynamical regimes: More or less complex limit cycles, “cycle slipping”, and chaos. We want to test how robust the delay estimation procedure is with respect to changes in the dynamics, and if one can also resolve the change of the delay within the regimes where the data look very similar to each other. We apply the single-delay method to the first 500 data points of each realization, which are highlighted in Figure 3. The maximal correlations $\Psi(\tau)$ are given in Figure 4, together with a line marking the true delay. As a result, for each τ

the value $\hat{\tau}$ where $\Psi(\tau)$ has its global maximum coincides with the true delay up to a small estimation error of one sample time step.

This example also reveals a natural limit of the method: Because the 500 data points of each of the nine time series contain only about a single orbit, or in other words, the manifold $(\dot{x}(t), x(t), x(t-\tau))$ is covered only sparsely by the approximate manifold described by the evolution of $(\Delta y_t, y_t, y_{t-\tau})$, here we do not yield reliable estimates also for the functions in Equation (19). This can be achieved for chaotic data containing at least about eight orbits. Finally, we note that in case of limit cycles, which contain very few information of the dynamics, the estimation of the functions works even better after adding a small amount of white noise to the data.

We have presented a method for the reconstruction of delay-differential equations from univariate time series. Our delay estimation procedure works well even for very short and noisy time series, as being shown in several examples of numerically generated data.

The method consists of two closely related ingredients: A generalized correlation function, based on the maximal correlation, and the method of optimal transformations. The generalized correlation function provides a suitable measure for finding the delays, and the method of optimal transformations, by virtue of the ACE algorithm, gives estimates for the governing functions of the DDE. It has been discussed that the estimates for the functions are only reliable, providing the data covers enough space of the manifold described by the DDE. This is not a restriction of the method itself but a general limitation due to the inverse nature of the problem.

Related to that, we state also two other problems of general nature, which are not bound to this method: First, if the dynamics is so simple that it could also be produced by a simpler equation than the ansatz (2), no method can, of course, test if the data has been really produced by Equation (2). Second, due to the simplicity of the solutions, like in the last example, where limit cycles occurred, one cannot be sure that all features of the dynamics are captured, which can also lead to unreliable estimates. With respect to this problems one should use, if available, also data containing a transient phase, in contrast to most other methods of time series analysis.

Our future work will concentrate on extending this method to the reconstruction of multivariate DDEs and partial differential equations, further on the application of this method to experimental data.

Finally, we would like to thank the referees for valuable comments.

References

- [1] M. J. Bünner, M. Popp, Th. Meyer, A. Kittel, and J. Parisi, Tool to recover scalar time-delay systems from experimental time series, Phys. Rev. E 54 (1996) R3082–3085.

- [2] W. Wischert, A. Wunderlin, and A. Pelster, Delay-induced instabilities in nonlinear feedback systems, *Phys. Rev. E* 49 (1994) 203–219.
- [3] E. Ott, T. Sauer, and J. A. Yorke, *Coping with Chaos* (Wiley Series in Nonlinear Science, John Wiley & Sons, New York, 1994).
- [4] J. D. Farmer, Chaotic attractors of an infinite dimensional dynamical system, *Physica D* 4 (1982) 366–393.
- [5] M. J. Bünner, M. Popp, Th. Meyer, A. Kittel, U. Rau, and J. Parisi, Recovery of scalar time-delay systems from time series, *Phys. Lett. A* 211 (1996) 345–349.
- [6] J. D. Farmer and J. J. Sidorovich, Predicting chaotic time series, *Phys. Rev. Lett.* 59 (1987) 845–848.
- [7] A. M. Albano, A. Passamante, and M. E. Farrell, Using higher-order correlations to define an embedding window, *Physica D* 54 (1991) 85–97.
- [8] L. Breiman and J. H. Friedman, Estimating optimal transformations for multiple regression and correlation, *J. Am. Stat. Assoc.* 80 (1985) 580–619.
- [9] A. Buja, Remarks on functional canonical variates, alternating least squares methods and ACE, *Ann. Stat.* 18 (1990) 1032–1069.
- [10] A. Rényi, On measures of dependence, *Acta. Math. Acad. Sci. Hungar.* 10 (1959) 441–451.
- [11] A. Rényi, *Probability Theory* (Akadémiai Kiadó, Budapest, 1970).
- [12] P. Tass, A. Wunderlin, and M. Schanz, A theoretical model of sinusoidal forearm tracking with delayed visual feedback, *J. Biol. Phys.* 21 (1995) 83–112.
- [13] B. Pompe, Measuring statistical dependences in a time series, *J. Stat. Phys.* 73 (1993) 587–610.

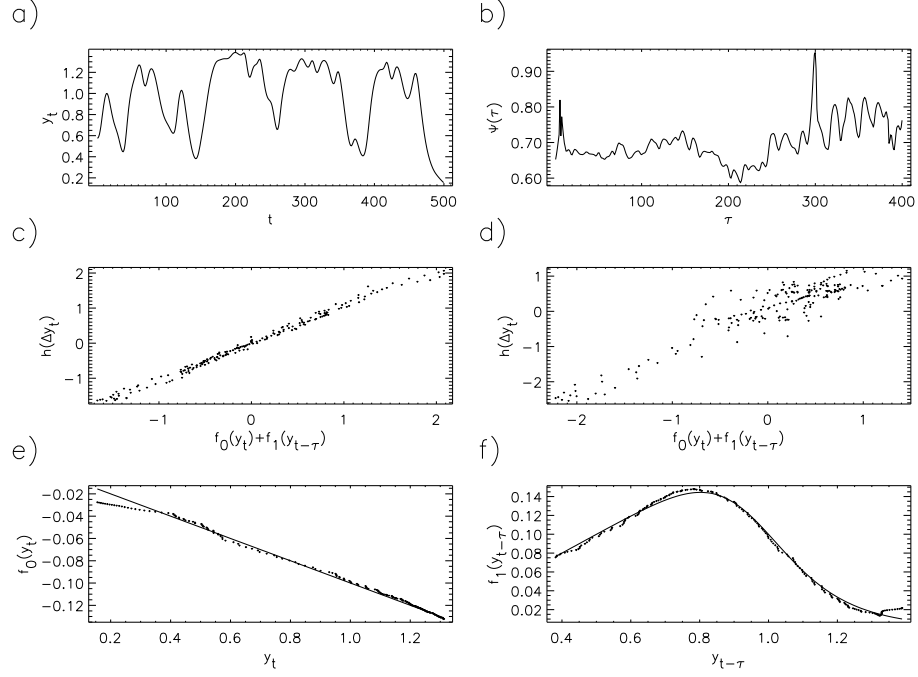


Figure 1: Delay estimation and function reconstruction for the Mackey-Glass equation with $\tau = 300$. a) 500 points of a realization of Equation (17). b) Maximal correlation $\Psi(\tau)$ for $\tau = 1, \dots, 400$. The peak indicates the estimated delay, $\hat{\tau} = 300$. c) Regression of the optimal transformations for the estimated delay, $\tau = 300$. d) The same as c) for a slightly shifted delay, $\tau = 297$. e,f) The true model functions (lines) and the estimated functions, given for the 199 corresponding data points y_t respectively y_{t-300} .

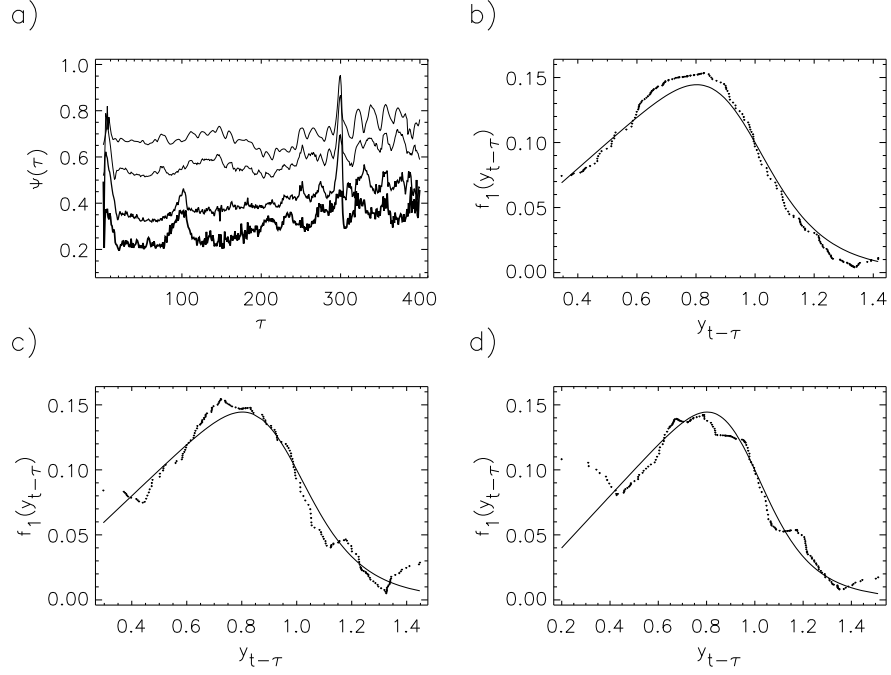


Figure 2: Delay estimation and function reconstruction for the Mackey-Glass equation corrupted by additive gaussian white noise of different magnitude. a) Maximal correlation $\Psi(\tau)$ for $\tau = 1, \dots, 400$, for noise levels of 0% (thinnest line), 5%, 10%, and 20% (thickest line) noise, as explained in the text. b)–d) The true nonlinear part of the model functions (lines) and the estimated functions for $\tau = 300$ and noise levels of 5% (b), 10% (c), and 20% (d).

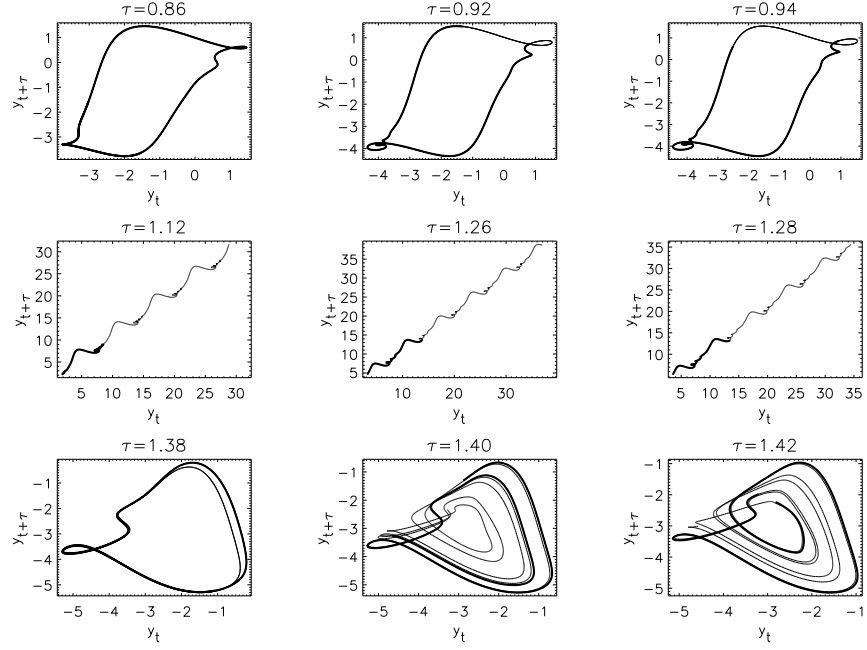


Figure 3: Phase portraits of realizations of Equation (19) for increasing delay τ given at the top of each figure. To give an impression of the dynamics, in each figure 2000 points are shown. The 500 points which are analyzed in Figure 4 are highlighted.

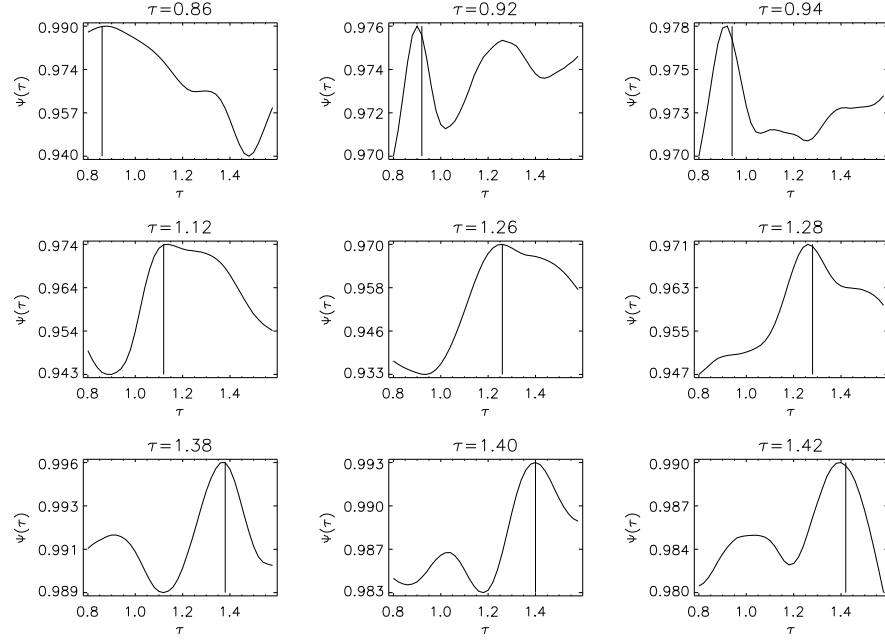


Figure 4: The maximal correlations $\Psi(\tau)$ for $\tau = 0.8, \dots, 1.42$ for the examples of Figure 3. The true delays, given at the top of each figure, are also marked by vertical lines.