

Reconstruction of systems with delayed feedback: I. Theory

M.J. Bünner¹, M. Ciofini¹, A. Giaquinta², R. Hegger³, H. Kantz³, R. Meucci¹, and A. Politi^{1,4,a}

¹ Istituto Nazionale di Ottica Applicata, Largo E. Fermi 6, 50125 Firenze, Italy

² Instituto Pluridisciplinar, Paseo Juan XXIII 1, 28040 Madrid, Spain

³ Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Str. 38, 01187 Dresden, Germany

⁴ INFN, Unità di Firenze, 50125 Firenze, Italy

Received 14 June 1999 and Received in final form 4 November 1999

Abstract. High-dimensional chaos displayed by multi-component systems with a single time-delayed feedback is shown to be accessible to time series analysis of a scalar variable only. The mapping of the original dynamics onto scalar time-delay systems defined on sufficiently high dimensional spaces is thoroughly discussed. The dimension of the “embedding” space turns out to be independent of the delay time and thus of the dimensionality of the attractor dynamics. As a consequence, the procedure described in the present paper turns out to be definitely advantageous with respect to the standard embedding technique in the case of high-dimensional chaos, when the latter is practically unapplicable. The mapping is not exact when delayed maps are used to reproduce the dynamics of time-continuous systems, but the errors can be kept under control. In this context, the approximation of delay-differential equations is discussed with reference to different classes of maps. Appropriate tools to estimate the *a priori* unknown delay time and the number of hidden components are introduced. The generalized Mackey-Glass system is investigated in detail as a testing ground for the theoretical considerations.

PACS. 02.30.Ks Delay and functional equations – 05.45.Tp Time series analysis

1 Introduction

Complex time-dependence in laboratory systems, in our natural environment or in living beings can have a variety of origins. One of the most fascinating perspectives is represented by the description of aperiodic fluctuations in terms of deterministic dynamical models. In the last two decades, much work has been devoted to testing for this hypothesis and to reconstructing the underlying dynamics under the assumption that only a scalar time-series is available. Since the pioneering articles of Packard *et al.* [1], Takens [2], and Grassberger and Procaccia [3], a sound body of knowledge has been progressively acquired [4], leading to the establishment of a new discipline, the nonlinear time series analysis. The general approach consists in reconstructing the phase space from the observed scalar data, most often by making use of the time delay embedding. In a sequence of spaces of increasing dimension, one looks for the manifestation of deterministic structures such as finite attractor-dimension or enhanced predictability. Unfortunately, this approach suffers from severe limitations as soon as the dynamical complexity of the underlying dynamics becomes relatively large.

Systems with time delayed feedback can create arbitrarily complex dynamics already with very few variables and rather simple equations of motion. The

Mackey-Glass equation [5] is the best known such example. It is a first order scalar differential equation with a force field that depends on a past value of the variable itself. This model was suggested in a physiological context (regulation of the production of red blood cells), where the mechanism of time delayed feedback is rather common. Further examples range from such widespread scientific disciplines as biology, epidemiology, physiology, or control theory [6,7]. In physics, this class of systems has been largely ignored in the past. Only recently, time delayed feedback has been introduced as a means to either enhance or suppress chaotic properties of nonlinear systems. Examples of the former strategy are the CO₂ experiments discussed in references [8,9] and the theoretical studies described in [10,11], while some of the control techniques based on the addition of a delayed feedback can be found in references [12–15].

From the mathematical point of view, time delayed feedback leads to delay-differential equations (see [6] for some results about the existence and uniqueness of solutions of the initial value problem). The corresponding phase space is infinite dimensional, as the initial condition is a generic function defined on the interval $[-\tau_0, 0]$, with τ_0 being the delay time of the feedback loop. In practice, however, high frequency components are almost absent and thus a finite number of variables suffices to parameterize the asymptotic solutions. On the other hand,

^a e-mail: politi@ino.it

the fractal dimension D can be made arbitrarily large as it has been established that D is proportional to τ_0 for sufficiently large τ_0 [10,16].

As already mentioned, the direct reconstruction of attractors from scalar data through time delay embedding using Takens theorem is clearly limited to low dimensional objects. A recent estimate [17] which takes entropy-related folding effects of the embedding procedure into account, shows that the minimal number of points N required for a clear manifestation of determinism must be larger than $\sqrt{e^{hD}s^D}$, where s is the required scaling range (e.g. $s = 10$ represents one decade of scaling) and h is the Kolmogorov-Sinai entropy. In practice, attractors with dimensions larger than 5 can hardly be recognized by time series analysis using Takens theorem, since otherwise an unrealistic large amount of data and an unrealistically low noise-level would be required.

High dimensional attractors of systems with time delayed feedback are thus practically indistinguishable from colored noise. On the other hand, the underlying delay-differential equation couples only a few variables, so that it is natural to ask whether more effective techniques exist which are able to reproduce the observed dynamics. It turns out that a reconstruction, not of the attractor in a proper phase-space, but of the dynamical rule in what we call “state” space, is often easier and equally effective.

For scalar time delay systems, *i.e.* systems involving a single variable, the delay time of an unknown system was estimated from a time series with the help of appropriate indicators [18,19]. Later, it was shown that the dynamical rule itself can be reconstructed from the time series of scalar time delay systems [20–23] and thereby the Lyapunov spectrum [24]. Most importantly, the dimension of the state space does not depend on the delay time, opening up the possibility to model and characterize high-dimensional regimes as well.

Since the restriction to scalar time-delay systems is, in practice, too severe, some efforts have been made to extend the latter ideas to the case of multi-variate time-delay systems. On a phenomenological basis, it was demonstrated that the delay time can be estimated also in such systems by treating them analogously to scalar ones [25]. For a multi-variate delay system with a single time-delayed feedback, an embedding-like approach was applied to experimental laser-data [26]. Such a methodology follows from the extension of Takens theorem [27] to input-output systems conjectured by Casdagli [28] and later proven in [29]. For the general case of multi-variate delay systems (with multi-variate delays) until now a multi-variate measurement is required [30].

Our work on the reconstruction of the dynamics of systems with delayed feedback is divided into two parts. The first part (the present paper) is devoted to a thorough discussion of theoretical aspects ranging from the conditions for a faithful reconstruction, to the various intrinsic approximations. The second part [31] will be instead devoted to the application of this technique to experimental data of a CO₂ laser with delayed feedback.

Here, we start considering the problem of mapping the original dynamics (possibly characterized by several variables) onto scalar models under the only restriction of a single feedback process. In Section 2 we show that the reconstruction is possible also when the recorded variable does not coincide with the feedback variable (the only case previously discussed in the literature), although at the expense of requiring a larger state-space dimension m_d . In Section 3, we investigate the approximations involved in the modelization of continuous-time dynamics in terms of delayed maps.

In the following Sections 4 and 5, we use the generalized Mackey-Glass model as a testing ground for the implementation of various indicators to identify the minimal optimal model, as well as to estimate the delay time. Local indicators such as the one-step forecast error are introduced and studied in Section 4, while global indicators are considered in Section 5. Finally, the open problems are briefly reviewed in Section 6.

2 Embedding theory

In this subsection, we introduce multicomponent systems with delayed feedback and discuss the possibility to map them onto suitable scalar models. Besides addressing a general mathematical question (*i.e.* the equivalence between different classes of dynamical equations), our motivation resides in the possibility to reconstruct the dynamics of a delayed system from a single scalar variable.

As anticipated in the introduction, we shall refer to a general case with d variables. The only restriction that we impose concerns the number of feedback processes: we shall assume that only one variable (in our notation: the first one) is fed back. We believe that this is a sufficiently general standpoint to begin a meaningful study of delayed systems. Although the physically meaningful models are continuous-time systems, so called delay differential equations (DDEs) of the form

$$\dot{\mathbf{x}} = \mathbf{H}(\mathbf{x}(t), x_1(t - \tau)), \quad (1)$$

we will usually construct models which are discrete in time and are called delayed maps (DM). The way DDEs are implemented on digital computers is precisely by constructing a suitable DM and, more important, DMs are the natural models to be reconstructed from experimental signals recorded with a finite sampling rate. More precisely, we consider the class of d -component single-feedback DMs

$$\mathbf{y}(n + 1) = \mathbf{F}(\mathbf{y}(n), y_1(n - \tau_0)), \quad (2)$$

where $\mathbf{y}_n \in \mathcal{R}^d$ and the delay time τ_0 is a positive integer number. The initial condition of the DM consists of a $(d + \tau_0)$ -component vector, so that the phase space is $\mathcal{R}^{d+\tau_0}$. Again without loss of generality, the feedback variable is assumed to be the first component.

With reference to discrete-time systems, we now discuss the question of reconstructing the dynamics of a given component y_k in terms of the values of the same

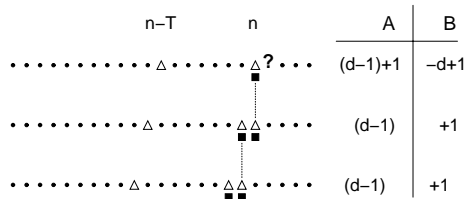


Fig. 1. Illustration of the coupling of the variables in the case where the embedding aims at the elimination of all the variables except the one with feedback. Full squares denote the hidden variables. Open triangles denote the variable with the time delayed feedback, which is accessible to measurement.

component at different previous times. The embedding theorems [2, 27] state that the knowledge of sufficiently many values of $y_k(n)$ for a given k allows to reconstruct the dynamics on the attractor. More precisely, the value of $y_k(n+1)$ is a unique function of its $2D$ previous values, where D is the attractor dimension. The essential point of this paper is the possibility to reconstruct the dynamics on high dimensional attractors with much less variables than required by these embedding theorems by exploitation of the particular structure of time delay systems.

We first derive the concept for a simple linear dynamical system

$$\mathbf{y}(n+1) = \mathbf{A}\mathbf{y}(n) + \boldsymbol{\alpha}y_1(n-\tau_0) \quad (3)$$

where \mathbf{A} is a $d \times d$ matrix and $\boldsymbol{\alpha}$ is a d -component vector. The structure itself of the above equation reveals a difference between the first variable (the only one being fed-back) and all the others. We thus have to specify which variable is actually recorded, since it plays an important role in the construction of an optimal model.

We first consider the case that the time series is a record of y_1 . Reconstructing the dynamics means finding an equation for $y_1(n+1)$. The obvious equation is just the first component of equation (3), which, however, requires the knowledge of all components of \mathbf{y} . Since we assume that they are not observed, we define all components of \mathbf{y} except y_1 as “hidden variables”. These unknowns must be explicitly eliminated to derive the desired dynamical law, involving exclusively past values of y_1 , but as few as possible.

In order to avoid lengthy technicalities, we will present the derivation of the desired equation of motion in a pictorial way, by referring to Figure 1. Each row in Figure 1 is a schematic description of the information involved in the application of equation (3) at a specific time. A full square positioned in the site n of the time-lattice indicates that all the $d-1$ hidden variables at time n are required in the iteration of the DM. Let us start from the uppermost row. As we have to determine only y_1 at time $n+1$ (see the question mark in the figure) we need to consider only one equation which in general will however depend on all variables at time n (see the corresponding full square at time n) and on y_1 at time $n-\tau_0$ (see the triangle). As a result, we have $(d-1)$ (the full square) plus 1 (the question mark) unknowns. This number is reported in column A on the right of the figure. The net difference between the

number of available equations and that of unknowns is instead reported in column B. We see that, in this case, since we have considered just one equation, such a difference is precisely $1-d$. Accordingly we reach the trivial conclusion that we cannot determine $y_1(n+1)$ from the knowledge of only $y_1(n)$ and $y_1(n-\tau_0)$.

Further information can be obtained from the past history. When we consider the application of the dynamical law at the previous time step (see the second row in Fig. 1), all the d equations enter into play. The number of unknown variables is $(d-1)$, *i.e.* the square at time $n-1$, while the difference between new equations (d) and new unknowns is 1. Therefore, the addition of this new step allows reducing the global gap between unknowns and equations. Iterating this procedure $d-1$ times will eventually allow to reach a break-even point, when the number of equations is equal to the number of unknowns. This means that we have to consider d rows in Figure 1, *i.e.* that $y_1(n+1)$ is unambiguously determined once y_1 itself is known in at least two windows of length d . Formally, we find that for $m \geq m_d^1$, with $m_d = d$,

$$y_1(n+1) = \boldsymbol{\beta} \cdot \mathbf{v}(n; m, \tau_0), \quad (4)$$

where

$$\mathbf{v}(n; m, \tau_0) = \left(y_1(n), y_1(n-1), \dots, y_1(n-m+1); \right. \\ \left. y_1(n-\tau_0), y_1(n-\tau_0-1), \dots, y_1(n-\tau_0-m+1) \right), \quad (5)$$

an expression stating that we have been able to transform the initial multicomponent DM (3) into a scalar equation (4) with, in some sense, multiple delays. The price we had to pay is that now the dependence on the past values of y_1 is not restricted to a single value as originally assumed in equation (3), but d consecutive values are needed.

If $\tau_0 < d$, the $2d$ variables appearing on the l.h.s. of equation (4) overdetermine $y_1(n+1)$, since in the above described process some unknowns are counted twice. As we have in mind applications to models with a few components compared to the delay, we shall not argue further about this point. Moreover, it is instructive to see that the dimension of the phase-space is $\tau_0 + d$ in the reconstructed model as well as in the original one: iteration of equation (4) indeed requires knowing $y_1(l)$ in the whole range $n \geq l > n - \tau_0 - d$. Accordingly, model (Eq. (4)) provides a faithful reconstruction of the whole dynamics including the convergence to the asymptotic attractor. This is to be contrasted with the possibility offered by the standard embedding technique to describe only the dynamics on the attractor itself.

The advantage over the standard application of the embedding theorem becomes more transparent if we also notice that the number of variables needed to reconstruct

¹ For the remainder of this paper we will term m_d the minimal “window size” of the model (4) that guarantees a proper embedding.

the dynamics is $2d$, independently of the delay τ_0 , *i.e.* independently of the phase-space dimension $\tau_0 + d$ that can be arbitrarily large. In particular, the technique can be equally effective also in the high dimensional regimes generally existing whenever $\tau_0 \gg 1$ (let us recall that the dimension of the attractor is generally proportional to the delay).

In the case of nonlinear DMs, the basic difference is that the function \mathbf{F} is, in general, non-invertible so that the knowledge of $2d$ variables is not sufficient for an unambiguous identification of $y_1(n+1)$. This problem is analogous to that of determining the minimum embedding dimension for a faithful reconstruction of the dynamics in the standard approach. In that context, a given attractor of dimension D_a can be generically reconstructed with a one-to-one correspondence only if the dimension of the embedding space is larger than $2D_a + 1$. In other words, one has to consider longer sequences in order to remove the ambiguities inherent to the lack of invertibility in a generic representation. More specifically, the model equation

$$y_1(n+1) = f(\mathbf{v}(n; m, \tau_0)), \quad (6)$$

with two windows of length $m \geq m_d$, and $d \leq m_d \leq 2d+1$, certainly suffices to faithfully reconstruct the dynamics in generic cases. The conjecture that $2(2d+1)$ represents an upper bound for the minimal dimension of the state-space is confirmed by interpreting the delayed feedback as an “external” driving (see Ref. [26]) and in turn invoking the analogy with input-output systems like those considered by [28]. In fact, this is the content of the embedding theorem generalized to that context in reference [29].

The feedback variable is certainly peculiar and different from all other variables involved in the dynamical process. It is therefore, interesting to ask oneself whether a compact reconstruction of the model is still possible if *not* the feedback variable is measured, but any other variable. The answer is yes, but the number of variables to acquire the necessary information is larger than before and the proof is also rather cumbersome so that a pictorial representation such as the one reported in Figure 2 will be very helpful. In this case, we must distinguish among three types of variables: $(d-2)$ hidden variables without feedback (represented by a full square); the hidden feedback variable (cross) and the variable experimentally observed (open triangle). In the first step of the procedure, there is one more unknown variable than before (since $y_1(n-\tau_0)$ is unknown, too) so that the gap between equations and unknowns is $-d$. Equally more negative is the second step, since the existence of an additional variable (the feed-back which is not recorded) prevents having a net gain. Therefore, recursively repeating the very same step does not allow removing all the unknowns. Nevertheless, we can still find a meaningful solution by modifying our strategy as described in the third step, where we consider the application of the mapping at time $n - \tau_0 - 1$. After comparing the newly involved variables with those already introduced in the two previous time steps, one sees that the additional gap is equal to $3 - d$. This result is strictly positive only if $d \leq 2$, thus suggesting that

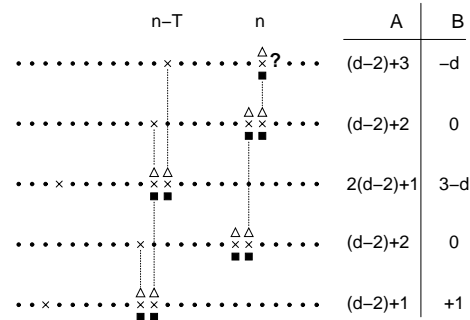


Fig. 2. Illustration of the coupling of the variables in the case where the embedding aims at the elimination of the variable with feedback. Hidden variables without feedback are denoted by full squares. The hidden variable with feedback is denoted by crosses. The experimentally observed variable is denoted by open triangles.

this new strategy leads in general to worse results. However, from now on, one can alternate steps of the previous and new type (see *e.g.*, the fourth and the fifth line in Fig. 2): this allows gaining 1 equation every second step. The break-even point is obtained after $2d - 1$ steps. This means that the recorded variable must be known in two windows of length $2d - 1$. Accordingly, the price to be paid for not dealing with the feed-back variable is that the number of “variables” is almost twice as large as before. Nevertheless, we can still consider this last result as positive, since the dimension of the space is still independent of the delay. An important difference with the previous case concerns the phase-space dimension. The iteration of the reconstructed model requires now to know a single variable over $\tau_0 + 2d - 1$ consecutive times, a number larger than the initial dimension $\tau_0 + d$ if $d > 1$. This means that our procedure has enlarged the phase-space dimension, introducing some spurious directions. We want to show now that the price for keeping the dimension of the phase-space equal to the original value is the construction of a much more complicated model. In fact, with reference again to Figure 2, we see that the steps of type 1 do not allow any gain only until we arrive at time $n - \tau_0$. However, from that point on the number of unknowns reduces by one unit per single step, since the variable y_1 was already taken into account, so that we eventually do not need to go beyond time $n - \tau_0 - d$. However, in doing so, all variables in the entire delay time are included, *i.e.* the standard embedding approach has been followed.

More in general, in the case of nonlinear systems, we expect that a model

$$y_2(n+1) = f(\mathbf{v}(n; m, \tau_0)), \quad (7)$$

exists for $m \geq m_d$, $d \leq m_d \leq 4d - 1$. However, it is honest to recognize that one will be hardly able to go beyond $d = 2$ in practical cases.

As in the standard embedding technique, the presence of nonlinearities with the possible noninvertibility of some functions might require doubling the number of variables necessary for a faithful reconstruction of the dynamics. Note that the number of spurious directions introduced

by the DM-model in the nonlinear case is at most $3d - 1$, and therefore much smaller compared to the number of spurious directions introduced by a Takens-type model, which can be up to $\tau_0 + d + 1$.

In the case of time-continuous models of the type (Eq. (1)), a similar reduction to a scalar equation can be developed with the difference that instead of giving rise to additional dependences on past values of the given variable, higher-order time derivatives have to be added. More precisely, the resulting model type is

$$x_1^{(m)} = h(\mathbf{w}(m, \tau_0)), \quad (8)$$

where

$$\mathbf{w}(m, \tau_0) = \left(x_1, x_1^{(1)}, \dots, x_1^{(m-1)}; \right. \\ \left. x_1(t - \tau_0), x_1^{(1)}(t - \tau_0), \dots, x_1^{(m-1)}(t - \tau_0) \right). \quad (9)$$

and $x_1^{(i)}$ stands for the i th derivative of the variable x_1 with respect to time.

3 From continuous to discrete time

In the previous section we have seen that vector-valued delay-models can be mapped onto scalar ones by embedding the attractors into suitable state spaces defined in terms of a single variable recorded in two windows of length m_d . These results provide the minimal framework for reconstructing the dynamics starting from the knowledge of just one observable. However, the exact inference of the model is formally possible only when a DDE (DM) dynamics is reconstructed in terms of a continuous (discrete) time model.

In reality, almost all physically meaningful processes stem from continuous time equations, while data are typically accessible as sequences of values sampled with a finite frequency. Accordingly, the typical situation consists in constructing a DM that mimics a DDE, *i.e.* we have to deal with the problem of passing from one to the other class of models. In this section, we discuss this problem, showing that the model mismatch implies that increasingly faithful reconstructions are only possible at the expense of increasing the state-space dimension. This can be done by lengthening either the first or the second window of each pair.

Before proceeding to the general discussion, it is important to stress that all the results reported in this section are derived under the assumption that (i) there exists a finite-dimensional attractor (this can be shown under quite general conditions [6]), (ii) the attractor dynamics is high-dimensional. In particular, we assume that the length of the window-pairs is smaller than the minimal embedding dimension required by the Takens theorem. This is because, as explained in the introduction, we want to consider cases where the usual embedding techniques fail to provide a faithful reconstruction.

For the sake of simplicity, we first consider a scalar DDE, $\dot{x} = H(x, x(t - \tau_0))$, and assume that the continuous variable $x(t)$ is recorded with a constant sampling time Δ on the discrete time-lattice $t_0 + n\Delta$ with $n \in \mathcal{Z}$. Let us call $x(n) = x(t_0 + n\Delta)$ from now on. In this framework we shall investigate the degree of accuracy that is possible to reach within the class of scalar DM-models. Let $\mathcal{A}(m_1, m_2)$ be the class of analytic functions h

$$\mathcal{A}(m_1, m_2) = \{h: R^{m_1+m_2} \rightarrow R\}. \quad (10)$$

Consider the DM-model

$$x(n+1) = h(\mathbf{v}(n; m_1, m_2, \tau)), \quad (11)$$

with $h \in \mathcal{A}$, and

$$\mathbf{v}(n; m_1, m_2, \tau) = \left(x_1(n), x_1(n-1), \dots, x_1(n-m_1+1); \right. \\ \left. x_1(n-\tau_0), x_1(n-\tau_0-1), \dots, x_1(n-\tau_0-m_2+1) \right), \quad (12)$$

with window pairs² (m_1, m_2) separated by a time τ_0 . We quantify the accuracy of the DM-model h in equation (11) with the help of the one-step forecast error (FCE):

$$\bar{\sigma}(h; m_1, m_2, \tau) = \sqrt{\frac{\langle (x(n+1) - h(\mathbf{v}(n; m_1, m_2, \tau)))^2 \rangle}{\langle x(n)^2 \rangle - \langle x(n) \rangle^2}} \quad (13)$$

where $\langle \cdot \rangle$ denotes a time average.

Any model h can be geometrically seen as an $(m_1 + m_2)$ -dimensional manifold in the state space augmented by the $x(n+1)$ direction (we shall call it, the \mathcal{S} -space). The FCE is trivially larger than zero whenever the original data lie on a manifold different from that one identified by the model. This is an error that can be removed by properly constructing the model. Conversely, if the data are distributed in a broader region, *i.e.* also transversally with respect to a hypothetical manifold, no exact model can be constructed and the FCE is bounded away from zero. This is precisely what we expect to happen because of the model mismatch: for any choice of window (m_1, m_2) , the variable $x(n+1)$ fluctuates in a small but finite interval, so that the FCE cannot be smaller than the average thickness of the distribution of points.

In order to clearly distinguish the latter fundamental limitation from trivial modelling errors, it is sufficient to define $\bar{\sigma}_{\mathcal{A}}(m_1, m_2, \tau)$ as

$$\bar{\sigma}_{\mathcal{A}}(m_1, m_2, \tau) = \min\{\bar{\sigma}(h; m_1, m_2, \tau) | h \in \mathcal{A}(m_1, m_2, \tau)\}. \quad (14)$$

² We have introduced the notation (m_1, m_2) to emphasize that the length of the two windows may be different. In that respect, the definition of \mathbf{v} contrasts with the one given in the previous section.

$\bar{\sigma}_{\mathcal{A}}(m_1, m_2, \tau)$ establishes the maximum level of accuracy that can be reached with a fixed window-system (m_1, m_2) and a delay time τ in the class of DM-models \mathcal{A} . From now on the function $h \in \mathcal{A}$, which minimizes the FCE, is called \hat{h} and therefore $\bar{\sigma}_{\mathcal{A}}(m_1, m_2, \tau) = \bar{\sigma}(\hat{h}; m_1, m_2, \tau)$. We shall see that there are at least two alternative procedures to increase the accuracy of the reconstruction. They amount to considering window pairs where either the first or the second window is increased beyond the dimension which would be required for a corresponding DDE model. For data from a scalar DDE, the optimal reconstruction is trivially a $(1, 1)$ -DDE model. We will show that the optimal DM for such data can be of either type of $(1, m_2)$ or $(m_1, 1)$, respectively.

Let us first discuss the $(1, m_2)$ case. Uniqueness and existence theorems [6] guarantee that the original DDE model can be written as a functional

$$x(t + \Delta) = G[x(t), \{x\}_d] \quad (15)$$

where $\{x\}_d = \{x(t') | t - \tau_0 \leq t' \leq t - \tau_0 + \Delta\}$. A simple example of the above functional dependence can be obtained in the case of the model class

$$\dot{x} = -\mu x + F(x(t - \tau_0)) \quad (16)$$

to which both the Ikeda [32] and Mackey-Glass [5] models belong. A formal integration of equation (16) yields

$$x(t + \Delta) = x(t)e^{-\mu\Delta} + \int_0^\Delta dt' F(x(t - \tau_0 + t'))e^{\mu(t' - \Delta)}. \quad (17)$$

If $x(t - \tau_0 + t')$ is nearly constant within the integration interval, one can approximate the functional dependence with a single value of the variable $x(t - \tau_0 + t')$ within the integration interval. This amounts to constructing a $(1, 1)$ -model and the uncertainty on $x(t + \Delta)$ is precisely the above introduced FCE $\bar{\sigma}_{\mathcal{A}}(1, 1, \tau_0)$, which is of Δ^2 -order³.

A better accuracy can be achieved if two or more consecutive points are assumed to be known in the vicinity of $x(t - \tau_0)$, since their knowledge allows constructing higher order approximations of F . Simple perturbative arguments suggest that the error made in the estimation of $x(t + \Delta)$ is of the order Δ^{m_2+1} , if m_2 consecutive points are used (*i.e.*, if a window-pair $(1, m_2)$ is considered) and Δ is small enough. In fact, the problem of estimating the error for fixed Δ and m_2 large enough is absolutely non trivial and deserves a discussion by its own. Here, without pretending to derive asymptotic estimates on the dependence of $\bar{\sigma}_{\mathcal{A}}(1, m_2, \tau_0)$ on m_2 and Δ , we limit ourselves to consider two limit cases. The first one consists in assuming that the Fourier modes above a certain frequency ω_c are slaved modes, *i.e.* they are uniquely determined by the amplitude of the lower-frequency modes. In this case, if the sampling time $\Delta < 2\pi/\omega_c$, we

³ In this section we always assume that the delay τ_0 is perfectly known and the uncertainty is entirely due to a model mismatch.

expect the residual uncertainty on $x(t + \Delta)$ to vanish for increasing m_2 , although it is not obvious to determine how rapidly. In the opposite limit, we can assume that the amplitude of each high-frequency mode is an independent variable (as in a stochastic process). In this case, the uncertainty on $x(t + \Delta)$ would depend on the “power” contained in the Fourier spectrum above the sampling frequency $\omega_s = 2\pi/\Delta$ and would not decrease for increasing m_2 . Were this the typical condition generated by DDEs, one should conclude that the model mismatch is so severe that one can never reproduce a continuous-time dynamics with arbitrary accuracy (with the exception of m_2 larger than the minimal dimension required by the embedding theorems for a faithful reconstruction).

An alternative approach for constructing a DM consists in approximating the first-order time derivative with linear combinations of the observable x in neighbouring points along the time lattice. It is well known that one can write

$$\dot{x}\left(t - \left(\frac{m_1}{2} - 1\right)\Delta\right) = \frac{1}{\Delta} \sum_{i=-m_1+1}^1 a_i x(t + i\Delta) + \mathcal{O}(\Delta^{m_1+2}) \quad (18)$$

for a suitable choice of the coefficients a_i . Upon substituting the above expression in the initial DDE and solving for $x(t + \Delta)$, we find that $x(t + \Delta)$ can be expressed as a function of the m preceding values and 1 value one-delay unit back in time. In other words, we have arrived at a DM of type $(m_1, 1)$, which involves an unavoidable error $\bar{\sigma}_{\mathcal{A}}(m_1, 1, \tau_0) \simeq \Delta^{m_1+1}$. This is again a purely perturbative result which is valid only for moderately large m 's.

In both the above discussed cases, we have seen that a discrete-time model can reproduce only approximately the dynamics of the original continuous-time system. In comparison to low-dimensional dynamical systems, for which we know that a generic ODE can be exactly transformed into a discrete mapping (even with the additional advantage of reducing the phase space dimension, if a Poincaré section is taken), the above results look very modest. The main reason for such a difference is that when a DDE is turned into a DM, the phase-space is necessarily “compressed” from an infinite- to a finite-dimensional one. The compression may be practically harmless, but necessarily involves the loss of small but nonzero interaction terms.

The two pairs $(1, m_2)$ and $(m_1, 1)$ are the limit cases of the more general combination (m_1, m_2) . We have been unable to estimate directly the uncertainty in this general case, because we failed to find an interpretation of the corresponding model in terms of derivatives and/or integrals. Nevertheless, with the help of a recursive argument we conjecture that

$$\bar{\sigma}_{\mathcal{A}}(m_1, m_2, \tau_0) \simeq \Delta^{m_1+m_2}. \quad (19)$$

We show this by starting from a DM model of the type (m_1, m_2) , namely

$$x(n + 1) = F^{(1)}(\mathbf{v}(n; m_1, m_2, \tau_0)) \quad (20)$$

which we assume to be accurate up to order $\Delta^{m_1+m_2}$. Moreover, we can claim that, as long as the distance between the two windows of a given pair is Δ -close to the true delay τ_0 , the error of the corresponding model does not change significantly. Accordingly, the dynamics is equally well described by the model

$$x(n+1) = F^{(2)}(\mathbf{v}(n; m_1, m_2, \tau_0 - 1)). \quad (21)$$

By now solving this latter equation with respect to $x(n+1 - m_1)$ (assuming that no problems connected with the invertibility of the nonlinear expression arise) and substituting in equation (20), we can write

$$x(n+1) = F^{(3)}(\mathbf{v}(n; m_1 - 1, m_2 + 1, \tau_0)) \quad (22)$$

for some function $F^{(3)}$. In other words, the value of x at time $n+1$ can be predicted on the basis of a window-pair of the type $(m_1 - 1, m_2 + 1)$ with the same order of magnitude for the uncertainty, *i.e.* $\Delta^{m_1+m_2}$ (in fact, the additional factor due to the error propagation in the inversion of the second equation is a finite correction term, independent of Δ). By iterating the same argument, one can eventually convince oneself that the accuracy of the model depends only on the total number of points in the two windows, as long as each single window is larger than m_d , the minimal number of degrees of freedom to be reconstructed.

Finally, we briefly discuss the general case of how to approximate multicomponent DDEs with multicomponent DM-models. In order to avoid technicalities, we limit ourselves to summarize the main steps, the whole derivation being straightforward. The generalization of the first approach (leading to $(1, m)$ -models in the scalar case) confirms the naive expectation based on the knowledge of the scalar case, *i.e.*, window pairs of the type $(m_d, m_d + l)$ guarantee an error of the order Δ^{l+2} . In fact, in full analogy with the discussion of equation (17), one can conclude that the formal integration of all equations allows approximating the original DDE equation up to order Δ^{l+2} with a “generalized” DM, where $l + 1$ past values of the scalar feedback variable are required. A simple repetition of the arguments presented in the first part of this section shows that this vector map can be turned into a scalar one of the type $(m_d, m_d + l)$.

In the complementary case that has led to the development of $(m, 1)$ -models in the scalar case, the scenario is much worse, since the derivative of each of the d variables must be determined with the prescribed level of accuracy. In order to fulfill this requirement, one must transform the original DDE into a DM involving dm_d variables in the first window. A by far larger number of variables is required as soon as $d > 1$.

4 Local indicators

Due to the discrete sampling rate and the well known numerical instabilities in the computation of derivatives from noisy data, the determination of a DDE model from experimental data is unreliable. Therefore, it is more reasonable

to reconstruct DM-rather than DDE-models, since the input information for a DM-model consists in the raw data, without the need of computing derivatives. Anyway, our discussion on how to determine the optimal model can nevertheless be extended with minor modifications to the determination of DDE-models as well.

There are several ways to quantify how well the observed data fulfill a deterministic rule in a given reconstruction space, such as the filling factor [20,21], the ACE-method [22], and others [23]. For the sake of simplicity, here we restrict our investigations to (m, m) -DMs,

$$y(n+1) = h(\mathbf{v}(n; m, \tau)). \quad (23)$$

In the following, we shall use the one-step forecast error $\sigma(h; m, \tau) = \bar{\sigma}(h; m, m, \tau)$, as a tool both to identify the correct delay time and to construct a meaningful model. In practice, one cannot deal with such a large space like that of analytic functions \mathcal{A} considered in the previous section. Accordingly, one must first identify a proper class of parameterized functions h to work with:

$$\mathcal{M}(m) = \{h(\cdot; \mathbf{a}): R^{2m} \rightarrow R\}, \quad (24)$$

where the parameter \mathbf{a} is varied to minimize the FCE. The optimal choice of a specific class $\mathcal{M}(m)$ depends on the problem under consideration. In practice, however, local linear models or global models built by radial basis functions [33] are generally quite successful. Here, we stick to the former class. The average required by the definition (13) is obviously performed along the available time series.

In practice, besides the fundamental problems discussed in the previous section, there are several additional limiting factors like a finite sampling time Δ , measurement noise, finite number L of data, and the mismatch between delay time and the actual sampling time, *i.e.* $\text{mod}(\tau_0, \Delta) \neq 0$ (note that so far in the literature only the case of no mismatch has been discussed). While the effect of noise will be considered in the second part of the paper with reference to a truly experimental system [31], here we shall investigate whether the other limitations may actually obstruct the model reconstruction.

The approach adopted in this section consists in the simultaneous identification of the model and determination of the optimal m -value as well as of the delay time. This is achieved by minimizing the forecast error σ which is nothing but the “distance” of the reconstructed model (23) from the true dynamics.

It is important to notice that local closeness between the model and the true dynamics does not necessarily imply closeness of the global dynamics. We can see this by discussing the case of a grossly wrong τ -value wherein we can expect that $y(n+1)$ is almost totally uncorrelated with the y -values belonging to the second (delayed) window. Accordingly, the information content of the second window is totally irrelevant in the minimization procedure of FCE. By invoking, as in the previous section, the analyticity properties of the underlying signal $y(t)$, we can estimate that $\sigma_{\mathcal{M}} \simeq \Delta^m$ (since knowing the value

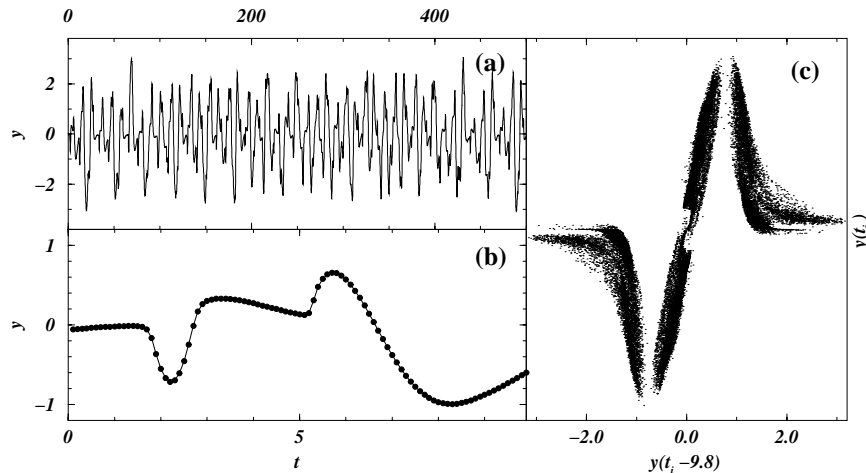


Fig. 3. (a-b) Time series of the generalized Mackey-Glass system; (c) delay plot of an extremal section. The values of extremal points $y(t_i), \dot{y}(t_i) = 0$ are plotted versus their retarded values.

of y in m points is tantamount to knowing the first m derivatives). This means that the FCE can be made very small even in the absence of relevant information about the force field (we recall that the delay is assumed to be far from the correct value), a conclusion that appears utterly illogical. In fact, this result tells us that in the small sampling-time limit, it is possible to perform reasonable short term predictions by simply exploiting the smoothness of the signal itself. In particular, it is not even possible to distinguish the true dynamics from that of the naive model $y^{(m)} = 0$, which corresponds to polynomial dependence on time, *i.e.* a dynamics which is neither stationary nor even limited. The conclusion to be drawn from this observation is that the smallness of $\sigma_{\mathcal{M}}$ alone is not enough to conclude that a meaningful model can be extracted from the raw data. This is the reason why we devote the next section to the discussion of other, global, indicators which do not suffer the same problems.

For $m_d = 1$ (scalar DDEs), the FCE proved to be a very effective and numerically inexpensive strategy to detect the unknown delay time τ_0 , since $\sigma_{\mathcal{M}}$ displays a pronounced local minimum for $\tau = \tau_0$ [26]. Before presenting the numerical data, we will focus on the expected FCE obtained for the “true” $\tau = \tau_0$. From the discussion carried on in the previous section, if $m \geq m_d$, the FCE is at least of the order of $\Delta^{2(m-m_d+1)}$. By comparing this estimate, derived for the correct value of τ , with the typical error expected for a generic delay, we find that the latter one is smaller, if $m < 2(m_d - 1)$, which is a clear nonsense. Since the FCE is defined as the *optimal* error, whenever some prior information is given, we can conclude that whenever both mechanisms do apply (*i.e.* when $\tau = \tau_0$), it is the most efficient one which determines the actual FCE. In other words, we do not expect any sensible dependence of $\sigma_{\mathcal{M}}(m, \tau)$ on τ if $m < 2(m_d - 1)$ (preventing the detection of the delay time with the help of the FCE), while a clear minimum should be seen in the opposite case. We can explain that behavior of the FCE by noticing that the two estimates of $\sigma_{\mathcal{M}}$ have been derived by invoking different mechanisms: (1) continuity of the evolution, (2) effective approximation of the delayed feedback model.

Since both mechanisms allow for high-quality short-term predictions, both will lower the FCE and, supposedly, any local indicator. Therefore, local indicators are not appropriate tools to distinguish between the two mechanisms. Global indicators (as discussed in the next section) are good candidates to also detect the delay times for $m < 2(m_d - 1)$. Anyway, the above inequality represents a necessary condition to be satisfied by a local indicator (such as the FCE) for a correct identification of the delay in the worst possible case.

In the following we discuss the problem of model reconstruction with reference to the generalized Mackey-Glass system [30]

$$\begin{aligned} \dot{y}(t) &= \frac{ay(t - \tau_0)}{1 + y^{10}(t - \tau_0)} + x(t), \\ \dot{x}(t) &= -\omega^2 y(t) - \rho x(t) \end{aligned} \quad (25)$$

which can be easily transformed into a second order ($d = 2$), neutral DDE

$$\begin{aligned} \ddot{y}(t) &= -\omega^2 y(t) - \rho \dot{y}(t) + \omega^2 f(y(t - \tau_0)) \\ &\quad + \frac{df(y(t - \tau_0))}{dy(t - \tau_0)} \dot{y}(t - \tau_0). \end{aligned} \quad (26)$$

The parameters are chosen as $a = 3.0, \rho = 1.5, \omega = 1.0$, and $\tau_0 = 9.83$, for which the Kaplan-Yorke dimension of the attractor is $D_{KY} = 7.2$. For $\rho = \omega^2$ and $\rho \rightarrow \infty$, the above equation reduces to the standard Mackey-Glass system by eliminating adiabatically the variable $x(t)$.

For the analysis, we use a time series of the variable that is fed back, $y(t)$, with a sampling time $\Delta = 0.1$. Notice that with this choice of Δ , the retarded values $y(t - \tau_0)$ lie outside the time lattice if $y(t)$ corresponds to one of the sampled values. In Figure 3 portions of the time series and a delay plot of an extremal section $\dot{y}(t_i) = 0$ are presented. The effect of the second component x in the dynamical equation (25) can be clearly visualized in the delay plot (with the delay being close to the delay time τ_0), since the intersection points of a scalar system have to lie on a curve in such a representation [20].

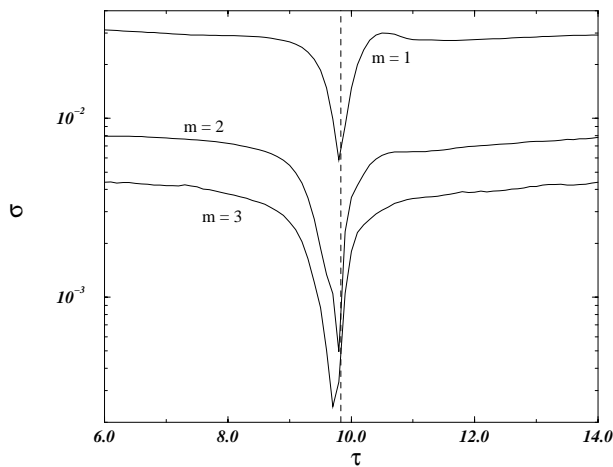


Fig. 4. One-step forecast error of an (m, m) -model as function of τ . From top to bottom, the curves refer to $m = 1, 2$, and 3 , respectively.

Table 1. Estimate of the delay time for different embedding dimensions.

	$m = 1$	$m = 2$	$m = 3$
$\hat{\tau}(m)$	9.8 ± 0.1	9.9 ± 0.1	9.7 ± 0.1

We have numerically investigated $\sigma_{\mathcal{M}}(m, \tau)$ for different choices of m and τ in a local linear model. The FCE is computed taking into account time series of length $L(m = 1) = 50\,000$; $L(m = 2) = 100\,000$; $L(m = 3) = 200\,000$. From the data reported in Figure 4, it is interesting to notice that a pronounced minimum of $\sigma_{\mathcal{M}}$ is observed even for $m = 1$, when, *a priori*, there is no reason to expect a faithful reproduction of the original dynamics. Such a result is the consequence of a general feature of dissipative systems: the various components are not equally “active”. Indeed, as long as the attractor is highly dimensional, the feedback term can be viewed as a noise term. In the absence of this “noise” source [34], the original system reduces to an ordinary differential equation, whose attractor fills a manifold of dimension smaller than m_d . The addition of the noise “thickens” the distribution along all directions in the state space, the width of the distribution depending on both transverse stability and noise amplitude. Accordingly, it may happen that the role of some components (corresponding to rather stable directions) in a multidimensional DDE is just to blur the distribution generated by a suitable DDE with less components. This is, to some extent, what happens in our system as clearly seen in Figure 3c, where the points cluster around a smooth curve which is the expected shape for the scalar Mackey-Glass system. A $(m = 1)$ -model will detect this curve, leading to a local minimum in the FCE.

The position of the minimum is an estimate $\hat{\tau}(m)$ of the delay time. A parabolic approximation of the FCE around the minimum yields the data reported in Table 1.

The estimated values agree with the correct value $\tau_0 = 9.83$ within the errors due to the finiteness of the sampling time.

Some comments are in order about the behaviour of the FCE. First of all, let us notice that a local minimum is observed also for zero delay. This minimum is due to the fact that we pass from a system of two windows of length m to a single window of double length. According to the arguments put forward in the first part of this section, we have to expect an accuracy of the same order as for the leading minimum. However, this accuracy cannot correspond to an equivalent accuracy of the global dynamics as the information about the feedback is missing. Moreover, we notice that the plateaus of the various curves decrease for increasing m . This is qualitatively in agreement with the considerations presented in the first part of this section about grossly wrong delays. However, there is no quantitative agreement about the scaling dependence on m : we attribute this to the existence of residual correlations between y values even when they are some time units apart.

Analogous considerations can be made for the height of the minimum that decreases less than expected on the basis of the general considerations discussed in the previous section. In this case, we have identified in the accuracy of the local linear model and in the finiteness of the number of points the main limiting factors which prevent $\sigma_{\mathcal{M}}$ from being smaller for $m = 3$.

5 Global indicators

We have seen that the FCE is a useful tool to decide whether a reconstructed model \hat{h} is locally close to the observed dynamics. Nevertheless, there is neither guarantee that the model dynamics remains confined to the region where it has been originally defined nor guarantee that the dynamics does not converge to a smaller subset (*e.g.*, a fixed point or a limit cycle). In other words, the smallness of the FCE $\sigma_{\mathcal{M}}$ is a necessary but not sufficient condition to establish whether a given model provides a globally faithful reconstruction. To test this, we iterate the models \hat{h} for different m -values and the optimal choice of the delay time, $\tau = \hat{\tau}_0(m)$ to generate some typical time series $\{\hat{y}_m\}$. We consider a model as valid, if the resulting attractor is “close” to the original one. To this aim, we introduce and utilize the cross forecast error, compute the power spectrum, the probability distribution of the sampled variable, and the Lyapunov spectrum as tools to establish altogether the validity of a given model.

However, before discussing all such indicators, it is instructive to perform a qualitative analysis of the generalized Mackey-Glass system (25) for $m = 1, 2, 3$ and the optimal choice of the delay time (as identified in the previous section). The resulting time series $\{\hat{y}_m\}$ (of length $L = 100\,000$) reveal a qualitative good agreement with the original series only if $m \geq 2$. Indeed, for $m = 1$, the time series \hat{y}_1 is asymptotically attracted to either a strictly positive or strictly negative region (see Fig. 5). Indeed, attractors with a specific sign exist in the standard Mackey-Glass system, where the unstable fixed point $y = 0$ acts as an “impenetrable” domain boundary separating the two coexisting attractors (changes of sign can exist only if they

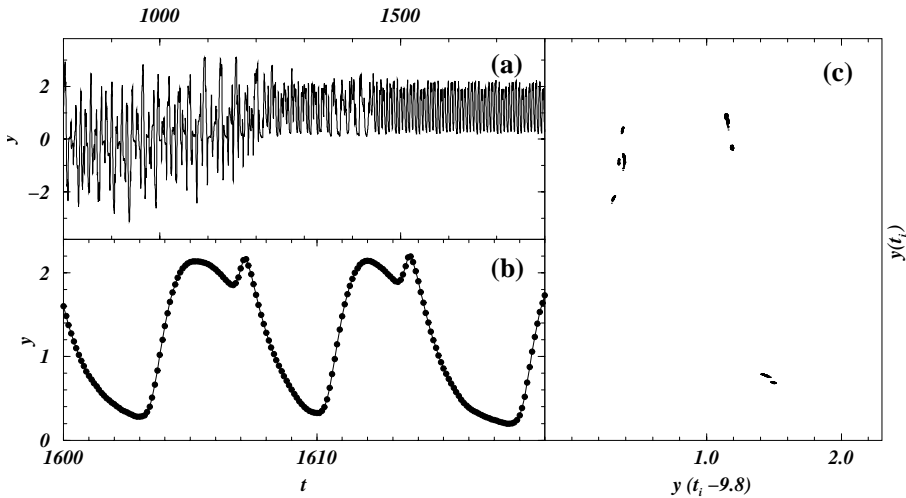


Fig. 5. Iterated time series $\{\hat{y}_m\}$ of the $(m = 1)$ -model: (a) convergence to an attractor with purely positive values; (b) blow-up of the attractor dynamics; (c) delay plot of an extremal section.

Table 2. The global distance defined in equation (27) for different values of m .

m	1	2	3
$\chi(m)$	1.197	0.032	0.026

are present in the initial condition; during the evolution, once disappeared, they cannot be generated again). Since, for $m = 1$, the second variable is obviously absent, it is not surprising that the reconstructed dynamics exhibits typical features of the standard Mackey-Glass system.

As this eventual convergence towards either positive or negative values persists, independently of the accuracy used in the model reconstruction and of the number of data points, we must rule out the possibility that $m_d = 1$, *i.e.* that a minimal approximate model can be constructed with just one component.

The problem of quantifying the “closeness” between the original time series y and the iterated time series \hat{y}_m cannot be faced by measuring to what extent the model-generated time-series remains all the way close to the original time-series. In fact, because of the chaotic properties of the evolution, an exponential separation always occurs which can hide the statistical equivalence of the two time-series. The most appropriate approach would consist in defining and measuring the distance between the two probability distributions. The natural space where this question should be formulated is the $(2m + 1)$ -dimensional space \mathcal{S} introduced in Section 3, *i.e.* the same space where the FCE is estimated and the dynamical rule reconstructed. There are various ways to define a distance, such as the Kullback-Leibler information [35] or the cross correlation sum [36]. Unfortunately, a meaningful implementation is a rather delicate matter. For instance, in the case of the Kullback-Leibler information, one needs sufficiently many data to get rid of statistical fluctuations in the local probabilities. Therefore, we have preferred to introduce a more robust geometrical indicator which, although carrying less information, can be satisfactorily complemented with the implementation of other tools.

For any point $P \in \{y\}$, determined by following the original trajectory, we identify the closest template used to construct the local linear model along the iterated time series $\{\hat{y}_m\}$ (after a suitable transient) and measure the distance $d(n)$ of P from such a $2m$ -dimensional surface. By averaging the square distances over all points P , we finally obtain the global indicator

$$\chi(m) = \sqrt{\frac{1}{N - n_0} \frac{\sum_{n=n_0}^N d^2(n)}{\langle y^2 \rangle - \langle y \rangle^2}} \quad (27)$$

where n_0 is such that the components of $\mathbf{v}(n; m, \tau)$ are all in $\{y(n)\}$. The definition of $\chi(m)$ is essentially the average forecast error along the time series $\{y\}$ on the basis of a model of the time series $\{y_m\}$, the latter being restricted to the attractor region (cross forecast error [37]). The results are presented in Table 2.

As a result of the confinement of the dynamics to a region with strictly positive values, the distance $\chi(m)$ is large for $m = 1$ (actually, so large that it compares with the standard deviation of the data). For $m \geq 2$, $\chi(m)$ decreases substantially (and could be further reduced by increasing the number of data points). Accordingly, the minimal choice $m_d = 1$ does not yield a faithful reconstruction, while the hypothesis $m_d = 2$ is already sufficiently good to be almost indistinguishable from further refinements (with the reasonable amount of data points adopted in our simulations).

Nevertheless, as already anticipated, such an indicator does not necessarily give a definite answer. In fact, we can imagine two distributions with the same support but grossly different densities. A geometrical indicator such as $\chi(m)$ would likely fail to identify at once such important differences since small distances would be found for all points (only a finer analysis could possibly allow detecting an insufficient quality of the reconstruction).

Therefore, we have decided to compute other quantities which have also a direct physical meaning. In Figures 6 and 7, we compare the spectra and the histograms of the original time series and of the iterated time series \hat{y}_2, \hat{y}_3 . A good agreement is achieved in both cases.

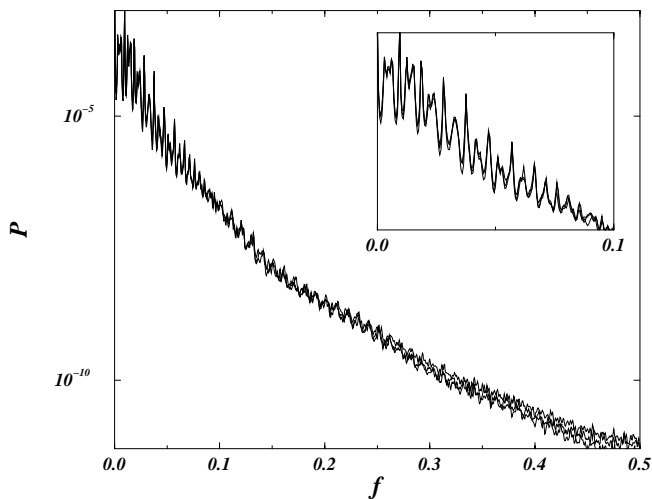


Fig. 6. Power spectra of the original time series of the generalized Mackey-Glass system, the ($m = 2$)-model, and the ($m = 3$)-model. The inset shows a blow-up for low frequencies.

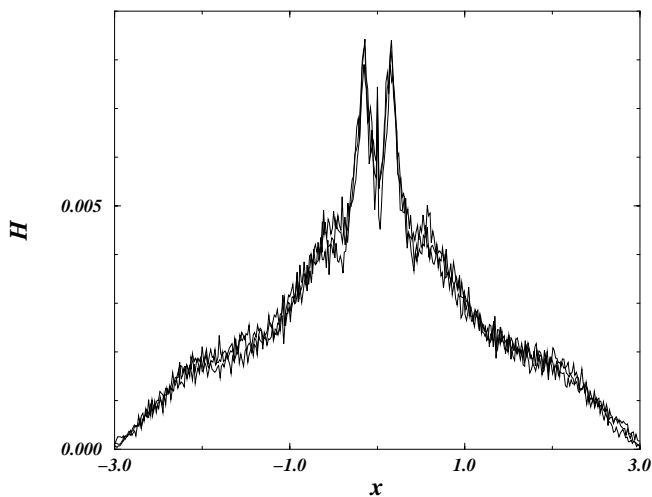


Fig. 7. Histograms of the original time series of the generalized Mackey-Glass system, the ($m = 2$)-model, and the ($m = 3$)-model.

Since no significant improvements are found in going from $m = 2$ to $m = 3$, we can confirm the previous conjecture that $m_d = 2$ is the minimal number of components necessary for a good reconstruction.

As a consequence of a successful modelling, it is not only possible to forecast the evolution in the real space, but also to extract information about the tangent space. In particular, one can compute the Lyapunov spectrum (LS) [24] for different choices of m (in correspondence of the optimal value of the delay). We expect that the LS grossly differs from the correct one whenever m is chosen too small, so that we can use Lyapunov exponents as a further global indicator to judge the quality of the reconstructed model.

In our example of the generalized Mackey-Glass equation, we estimated the LS for $m = 1, 2$, and 3 . The results are compared with the estimation of the spectrum obtained by direct integration of the equations (see Fig. 8). Again we observe large deviations for $m = 1$, while for

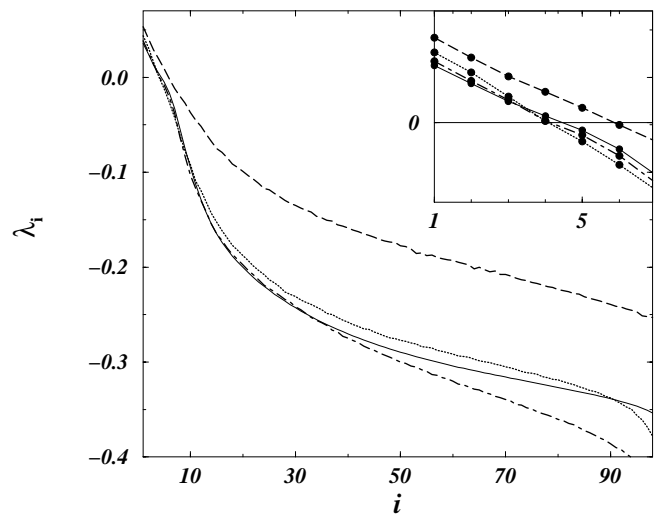


Fig. 8. Lyapunov spectra of the generalized Mackey-Glass system as estimated from the equations (solid line); from the ($m = 1$)-model (dashed line); from the ($m = 2$)-model (dotted line); from the ($m = 3$)-model (dot-dashed line). The inset shows a blow-up of the largest Lyapunov exponents.

$m \geq 2$, the LS is rather close to the true spectrum, thus confirming once more the scenario suggested by the other indicators.

While it is not the goal of this paper to derive general quantitative estimates of the various sources of errors⁴, we would finally like to draw the attention of the reader to some possible effects of dynamical noise. In fact, besides downgrading the quality of the model predictions, dynamical noise can mask the presence of additional degrees of freedom. Once more, the Mackey-Glass equation is a simple system to illustrate this phenomenon, since additional noise in the scalar model can induce jumps from positive to negative values of the variable y (as we have numerically verified), thus making the evolution practically indistinguishable from that of a noisy generalized Mackey-Glass equation. In other words, the problem of a correct identification of the deterministic components depends on the possible/required accuracy of the modellization. In some cases, it might even be desirable to model only the gross nonlinear features in terms of a few variables while assimilating all the others to a sort of background noise indistinguishable from a true noise.

We conclude with a remark about the length of the time series. It has been emphasized by some authors that the number of points necessary to estimate the delay time in a scalar system series can be quite small (say 500–1000) compared to the number of data points required in conventional nonlinear time series analysis. In principle, we can confirm this result for multi-component systems and a single time-delay feedback. On the one hand, the discovery of hidden variables requires embedding the data in spaces of increasingly higher dimensions (though much smaller than the dimension of the entire phase space) and

⁴ In the second part paper [31] we shall see that dynamical and measurement noise do not prevent a meaningful extraction of information from experimental data.

thus an increasingly larger number of data points. On the other hand, since we do not aim at detecting scale invariant properties, the number of points required to obtain statistically significant results for the estimation of the delay time can be comparably small.

6 Open problems

In this paper we have shown that time-delayed feedback systems can be investigated on the basis of a single valued time series. In particular, we have seen that the dynamics can be reconstructed in low-dimensional state-spaces even when the attractor dimension can be arbitrarily large. This result has been obtained by restricting ourselves to the case where only one variable is fed back with a fixed specific delay time τ_0 . Two possible generalizations of this setup can be conceived that might be of interest in practical applications.

First, one can assume that the single feedback variable acts with several different delays. As long as the number of such interactions is finite, no dramatic changes are expected from a theoretical point of view: instead of working with a “two-window” embedding, it should suffice to use an $(n + 1)$ -window embedding, where n is the number of delays. This is a straightforward generalization, as long as the windows do not overlap. Of course, the advantage of a low dimensionality of the state space is lost as soon as n becomes large, but for only a few delays it might still work reasonably well. Completely different is the situation when we have to deal with a continuous spectrum of delays. In this case we expect this method to fail, as it is no longer possible to reconstruct the equations of motion in a low-dimensional manifold.

A second possible generalization consists in sticking to a single delay time τ_0 , but admitting that several variables are fed back. This is similar to the case where we measure the *wrong* variable, as only a scalar variable is used to reconstruct the dynamics. This suggests that the length of the two windows should be increased by some factor in this case.

Another open problem concerns the uncertainty affecting a DM model that arises from the model mismatch due to the supposedly continuous-time dynamics. In this paper, we have employed perturbative arguments to estimate the order of magnitude of the FCE, when the windows are not too long. However, this is still insufficient to draw mathematically rigorous conclusions about the convergence properties of DM models towards the expected continuous-time limit. In fact, a non-perturbative approach is presumably necessary to deal with large window-lengths, besides the inclusion of additional information about the dynamical behaviour of the process under investigation. This is a hard task that extends a general and still unsolved problem: that of estimating the indeterminacy of an optimal prediction (on the basis of the standard embedding approach) for a high-dimensional deterministic process.

M.J.B. is supported by a Marie-Curie-Fellowship of the EU with the contract number: ERBFMBICT972305. R.H.

is partly supported by the EU with the contract number: ERBFMRXCT96.0010 and wants to thank the colleagues at the INOA for their kind hospitality.

References

1. N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
2. F. Takens, in *Dynamical Systems and Turbulence* (Warwick, 1980), *Lecture Notes in Mathematics*, edited by D.A. Rand, L.-S. Young (Springer-Verlag, Berlin, 1980), Vol. 898, pp. 366–381.
3. P. Grassberger, I. Procaccia, *Phys. Rev. Lett.* **50**, 346 (1983).
4. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 1997).
5. M.C. Mackey, L. Glass, *Science* **197**, 287 (1977).
6. J.K. Hale, *Theory of functional differential equations* (Springer-Verlag, New York, 1977).
7. R. Bellman, K.L. Cooke, *Differential-Difference Equations* (Academic Press, New York, 1963).
8. F. Arecchi, W. Gadomski, R. Meucci, *Phys. Rev. A* **34**, 1617 (1986).
9. G. Giacomelli, R. Meucci, A. Politi, F. Arecchi, *Phys. Rev. Lett.* **1099**, 2686 (1996).
10. J.D. Farmer, *Physica D* **4**, 366 (1982).
11. G. Giacomelli, A. Politi, *Phys. Rev. Lett.* **76**, 2686 (1996).
12. K. Pyragas, *Phys. Lett. A* **170**, 421 (1992).
13. S. Bielawski, D. Derozier, P. Glorieux, *Phys. Rev. E* **49**, R971 (1994).
14. W. Just *et al.*, *Phys. Rev. Lett.* **78**, 203 (1997).
15. D. Sukow, M. Bleich, D. Gauthier, J. Socolar, *Chaos* **7**, 560 (1997).
16. G. Giacomelli, S. Lepri, A. Politi, *Phys. Rev. E* **51**, 3131 (1995).
17. E. Olbrich, H. Kantz, *Phys. Lett. A* **232**, 63 (1997).
18. A.C. Fowler, G. Kember, *Phys. Lett. A* **125**, 402 (1993).
19. Y. Tian, F. Gao, *Physica D* **108**, 113 (1997).
20. M.J. Bünner *et al.*, *Phys. Lett. A* **211**, 345 (1996).
21. M.J. Bünner, T. Meyer, A. Kittel, J. Parisi, *Phys. Rev. E* **54**, R3082 (1996).
22. H. Voss, J. Kurths, *Phys. Lett. A* **234**, 336 (1997).
23. S. Ellner *et al.*, *Physica D* **110**, 182 (1997).
24. R. Hegger, *Phys. Rev. E* **60**, 1563 (1999).
25. M.J. Bünner *et al.*, *Europhys. Lett.* **42**, 353 (1998).
26. R. Hegger, M. Bünner, H. Kantz, A. Giaquinta, *Phys. Rev. Lett.* **211**, 345 (1998).
27. T. Sauer, J.A. Yorke, M. Casdagli, *J. Stat. Phys.* **65**, 579 (1991).
28. M. Casdagli, in *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity* (Addison-Wesley, Reading, 1992).
29. J. Stark, D. Broomhead, M. Davies, J. Huke, *Nonlin. Anal. Meth. Appl.* **30**, 5303 (1997).
30. M.J. Bünner, T. Meyer, A. Kittel, J. Parisi, *Phys. Rev. E* **56**, 5083 (1997).
31. M. Bünner *et al.*, *Eur. Phys. J. D* **10**, 177 (2000).
32. K. Ikeda, K. Matsumoto, *Physica D* **29**, 223 (1987).
33. X. He, A. Lapedes, *Physica D* **70**, 289 (1993).
34. B. Dorizzi *et al.*, *Phys. Rev. A* **35**, 328 (1987).
35. R. Badii, A. Politi, *Complexity, hierarchical structures and scaling in physics, Cambridge Nonlinear Science Series* (Cambridge University Press, Cambridge, 1997).
36. H. Kantz, *Phys. Rev. E* **49**, 5091 (1994).
37. T. Schreiber, A. Schmitz, *Phys. Rev. Lett.* **79**, 1475 (1997).