Physics Letters A 175 (1993) 402–408
North-Holland

PHYSICS LETTERS A

Delay recognition in chaotic time series

A.C. Fowler and G. Kember
Mathematical Institute, Oxford University, 24–29 St. Giles', Oxford OX1 3LB, UK

Received 26 November 1992; revised manuscript received 22 February 1993; accepted for publication 24 February 1993
Communicated by A.P. Fordy

We show how the use of "smart" embeddings of time series can indicate the presence of (large) delay in a system, and how they can be used to enhance predictions based on nonlinear dynamics methods.

1. Delay equations are of wide relevance to natural dynamical systems, particularly in medicine and physiology. For example, models of respiration [1] and cell maturation [2] naturally include significant delays; other systems which have been modelled with delay differential equations are population dynamics [3], and lasers [4].

Since such systems can display extremely chaotic behaviour, even for a first order equation with a single delay, it is of interest to know whether any of the current nonlinear dynamical methods of time series analysis (e.g. ref. [5]) have the potential to recognise and make use of the delay. In this Letter we will report one idea which seems to have some application in this context.

Many of the systems which incorporate delays are modelled by *delay-recruitment* equations, of the general form

$$\epsilon\dot{x} = -x + f(x_1) , \tag{1}$$

where $x_1 = x(t-1)$. Such models combine an exponential relaxation term with a nonlinear forcing term dependent on the retarded argument x_1. Of particular interest is when the dimensionless parameter ϵ, the ratio of the relaxation time to the delay, is small, for then solutions oscillate on a time scale $t \sim \epsilon$, and as a consequence, the effective dimension of chaotic behaviour is of order $1/\epsilon$ [6].

Chaos in equations such as (1) is associated with chaos in the discrete map $x \to f(x)$, although the precise relationship is not clear. When ϵ is small, one might expect solutions to be close to the singular limit $x = f(x_1)$ [7], but the rapid oscillations make any such easy comparison opaque. In this paper we focus on the Mackey–Glass equation [2], which can be written in the form (1), with

$$f(x) = \frac{\lambda x}{1+x^c} . \tag{2}$$

The equation was studied by Farmer [6], and can be obtained in the form (1) by putting $\lambda = a/b$, $\epsilon = 1/b\tau$, with a, b, τ as in Farmer's paper. He chose values $c = 10$, $\lambda = 2$, and a range of $\epsilon < 1$. He found that the information dimension of the chaotic attractor at small ϵ was $D \approx 1/\epsilon$. A typical time series of the solution is shown in fig. 1, when the attractor dimension is about 20. In this note, we always keep $c = 10$, $\lambda = 2$.

2. In the normal way, a system of large dimension requires an equivalent number of variables for its description. And yet the time series in fig. 1 is generated by a single equation. Is there some way in which this information can be extracted? In particular, can we infer the delay from the time series? The answer, surprisingly, is yes.

The basic idea is this. For small enough δ ($\delta < \epsilon$), we approximate $\dot{x} \approx (x - x_\delta)/\delta$ ($x_\delta = x(t-\delta)$), thus

$$x \approx \frac{\epsilon}{\epsilon+\delta} x_\delta + \frac{\delta}{\epsilon+\delta} f(x_1) . \tag{3}$$

While (3) is not necessarily very accurate, the ideas of embedding techniques suggest that if the time series in fig. 1 is embedded in $\mathbb{R}^3$ as (x, x_δ, x_1), then

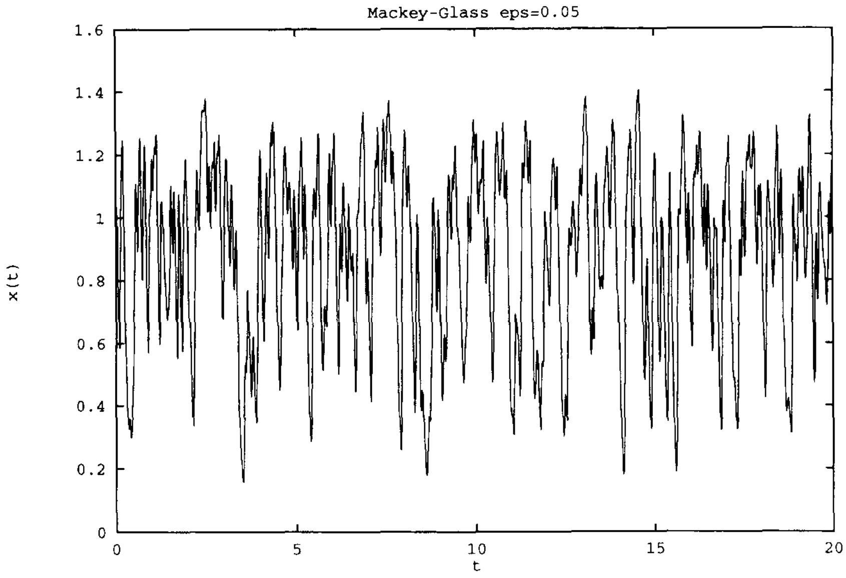

Fig. 1. Time series for Mackey–Glass equation (1), (2) with $\lambda=2$, $c=10$, $\epsilon=0.05$.

the trajectory will lie close to a surface. On the other hand for fixed $\delta\sim\epsilon$ and $\Delta\neq 1$, the trajectory embedded in $\mathbb{R}^3$ as (x, x_δ, x_Δ) should fill a three-dimensional volume, since the attractor dimension is $\sim 1/\epsilon$. Therefore, if we plot some measure of trajectory volume versus Δ, we should see a sharp change near $\Delta=1$, corresponding to a collapse of the volume.

The measure we have chosen is called the *singular value fraction* (SVF), and is constructed as follows. Given a time series, we choose an embedding with one variable time lag Δ. We then normalise the embedded time series so that it has zero mean and unit variance. For each embedding, we use global singular value decomposition (SVD) to determine the principal singular vectors $w_1, ..., w_{d_E}$ and their associated singular values $\sigma_1, ..., \sigma_{d_E}$, where d_E is the embedding dimension, and $\sigma_1\geqslant\sigma_2\geqslant...\geqslant\sigma_{d_E}\geqslant 0$. For any given choice of k, we define

$$F_{SV}(k)=\sum_1^k \sigma_i^2 \Big/ \sum_1^{d_E} \sigma_i^2 \,. \tag{4}$$

Notice that $0<F_{SV}\leqslant 1$ for $1\leqslant k\leqslant d_E$, and F_{SV} is monotone increasing. We can show that in fact $F_{SV}\geqslant k/d_E$, and therefore we define the *singular value fraction* $f_{SV}(k)$ as $(d_E F_{SV}-k)/(d_E-k)$, or equivalently,

$$f_{SV}(k)=1-\frac{1}{(d_E-k)N}\sum_{k+1}^{d_E} \sigma_i^2 \,. \tag{5}$$

Thus $f_{SV}\in[0, 1]$ for $1\leqslant k\leqslant d_E$, and is monotone increasing with k (and $f_{SV}(d_E)=1$). If $f_{SV}=1$ for $k<d_E$, then the attractor resides in a k-dimensional subspace. Now since the attractor dimension $\sim 1/\epsilon$ is large, we assume that $d_E<1/\epsilon$, so that the embedded trajectory will fill the phase space. We then expect that if we plot f_{SV} as a function of Δ, there will be a sharp rise at $\Delta=1$ as the attractor volume collapses. In fig. 2, we show that $f_{SV}(\Delta)$ experiences just a maximum near $\Delta=1$. Further details on the use of singular value fractions in diagnosing delays and selecting optimal embedding lags will be presented elsewhere; here we wish to examine the possible use of so-called "smart" embeddings in making predictions.

It should be emphasised that other methods could be used to establish the attractor collapse at $\Delta=1$. In particular, use of generalised dimension statistics [8] such as the correlation integral $C_2(r)$ (which measures the proportion of pairs of points in the embedded trajectory which are a distance less than r from

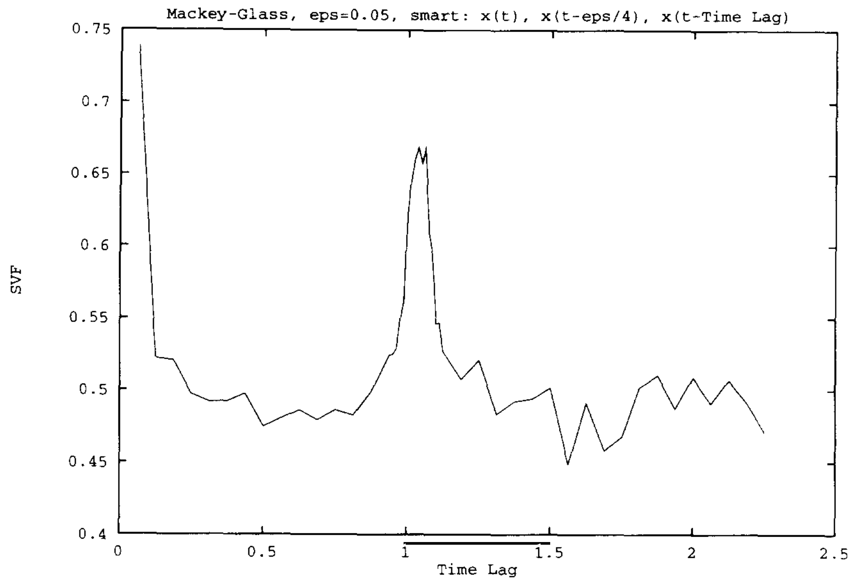

Fig. 2. SVF for Mackey–Glass equation, $k=2$, as a function of Δ for the 3-D embedding (x, x_δ, x_Δ) with $\delta=\epsilon/4$.

each other) might allow for possible folding, which the SVF would not be good for. Specifically, we can expect

$$C_2(r) \sim g(\Delta)(r/D)^\nu \tag{6}$$

for small r, where D is the linear dimension of the embedding trajectory, and ν is the correlation dimension. Reasonably, one can expect the premultiplicative factor g to depend on the variable lag time. Now if the attractor really collapsed to a lower dimension, one might expect $C_2 \approx (r/D)^{\nu-1}$ when $\Delta \approx 1$, and thus $g(\Delta)$ should experience a sharp peak near $\Delta=1$, at fixed (small) r. We have not tested this conjectural idea, though it may have its own problems; the point to be made is that, where folding of the collapsed trajectory does occur, SVD is not likely to be as useful.

3. The standard embedding for a time series such as in fig. 1 would be $x, x_\delta, x_{2\delta}, \ldots$. We designate a *smart* embedding as $(x, x_\delta, x_{2\delta}, \ldots, x_{k\delta}, x_{\Delta_1}, x_{\Delta_2}, \ldots, x_{\Delta_j})$, where δ is a "normal" lag selection, and $\Delta_1, \Delta_2, \ldots$ are one or more lags selected by successive use of SVF, or some similar method. In making predictions based on local linear or nonlinear predictors, we are interested in minimising the average absolute prediction error, P. In general, P will depend on the number of data points, N; the embedding dimension, d_E; as well as the prediction method, noise level, etc. If the attractor dimension is D, then for $d_E < D$, we can expect the trajectory to fill the embedding space – it is being projected onto it, and nearby points in $\mathbb{R}^{d_E}$ may not be close on the attractor. Thus P should decrease until d_E reaches D, *providing N is large enough.* As d_E increases further, P will increase again, since the N data points are being spread around a larger and larger space.

In a delay system with a large delay, there is a serious problem if N is limited. If D is large, one typically expects $\exp[O(D)]$ data points as a requirement to make useful diagnostics or predictions [9]. This makes the use of smart embeddings very attractive, as they provide an effective way of decreasing the dimension of the attractor – or at least squashing it flatter.

Figure 3 shows the mean absolute prediction error P versus Δ for the Mackey–Glass equation, using a three-dimensional smart embedding $(x, x_{\epsilon/4}, x_\Delta)$, and local linear prediction using the average of four nearest neighbours. Prediction is one step $(\epsilon/4)$ ahead. We see that there is a sharp minimum near $\Delta=1$ (remember the normalised time series has unit stan-

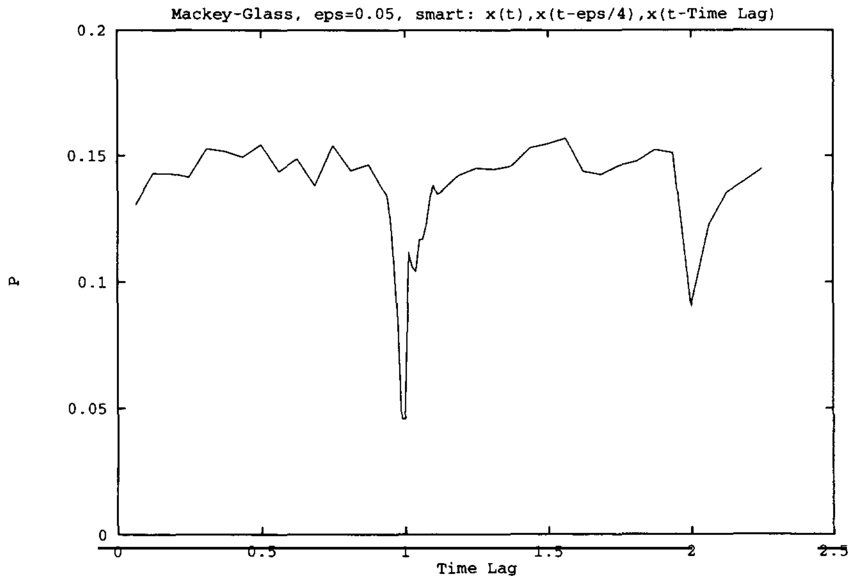

Fig. 3. Average absolute prediction error P versus Δ for Mackey–Glass equation, embedding as for fig. 2.

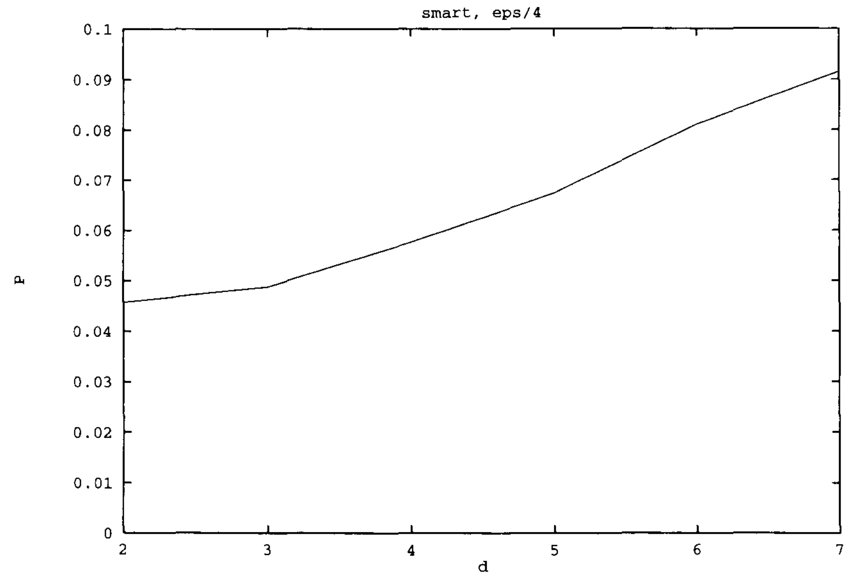

Fig. 4. Prediction error P versus d_E for the normal embedding $(x, x_\delta, ..., x_{(d_E-1)\delta})$.

dard deviation). In fig. 4 we plot P versus d_E for the smart embedding $(x, x_\delta, ..., x_{d_E-2}, x_\Delta)$ with $\delta = \epsilon/4$, $\Delta = 1$. Apparently paradoxically, P *increases* with d_E. This is due to the fact that $d_E < 1/\epsilon$, so that the points continue to spread apart as d_E increases. To obtain P decreasing with d_E, we can change the prediction step ahead to be ≈ 1. This point will be pursued elsewhere.

4. The evidence presented above illustrates our main thesis, that particularly in delay differential equations, smart embeddings provide a useful vehicle for constructing accurate low-dimensional reconstructions. Moreover, this approach may be extended to other systems with, for example, different time scales of behaviour. Here we wish to pursue a feature of (1), as exhibited by fig. 3. This is most strikingly illustrated if we plot phase portraits (x, x_τ) of fig. 1, using in fig. 5a $\tau=1.5$, and in fig. 5b $\tau=1.025$, where also $f(x_\tau)$ is plotted. We see that the trajectory collapses close to $f(x_\tau)$ when $\tau\approx 1$. This suggests that in some average sense, the singular limit $\epsilon=0$ in (1) is attained, although the means by which this occurs is very unclear. In fig. 6, we show the average value of $|x-f(x_\Delta)|$ as a function of Δ. The minimum near $\Delta=1$ is very similar to the minimum displayed by the prediction error.

Some understanding of this comes from rewriting

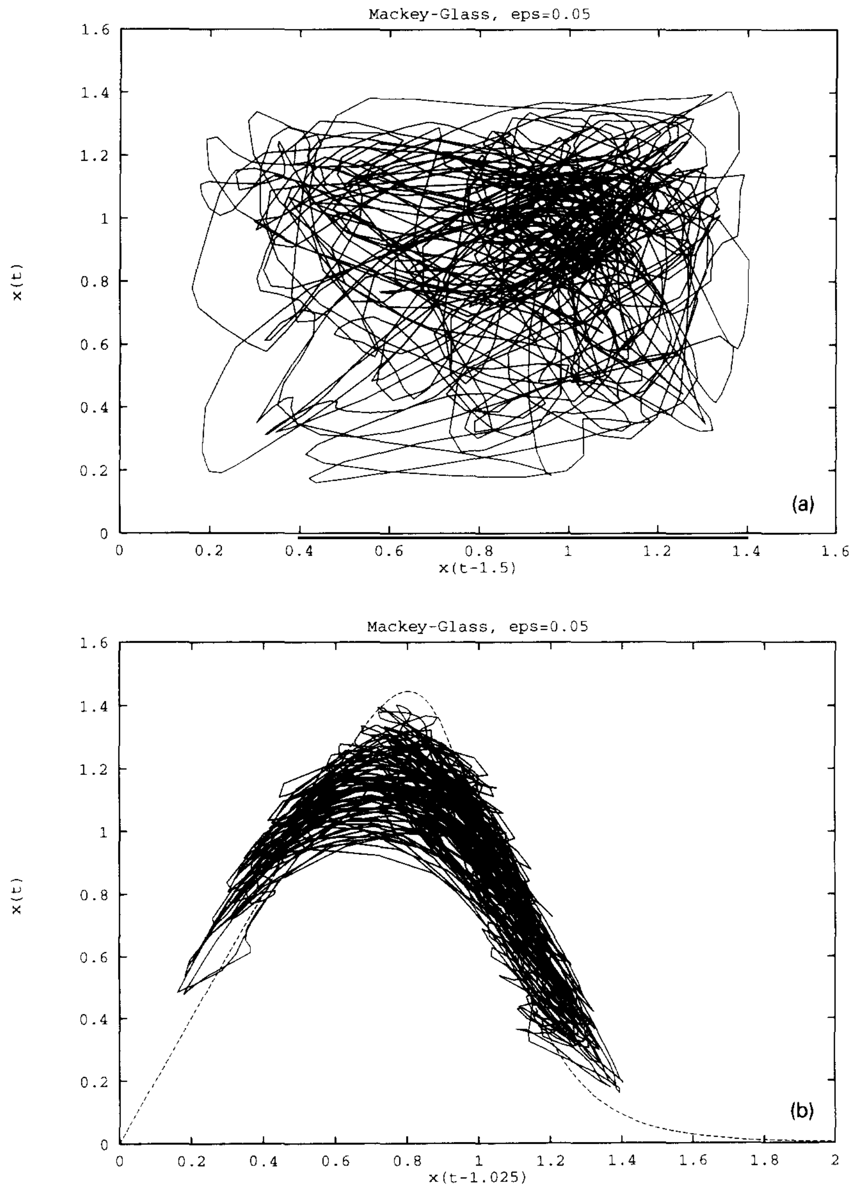

Fig. 5. Phase plot of fig. 1, x versus x_δ: (a) $\delta=1.5$; (b) $\delta=1.025$.

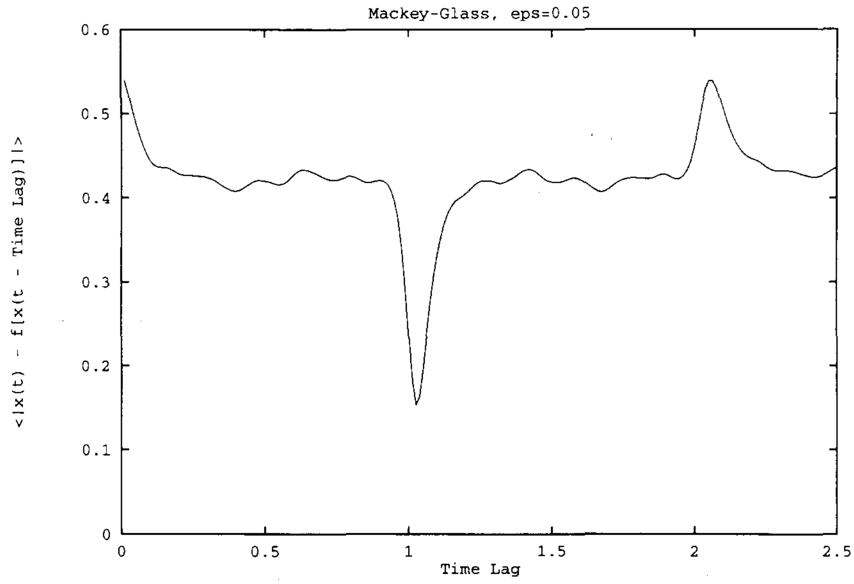

Fig. 6. Mean absolute deviation $\langle |x-f(x_\Delta)| \rangle$ as a function of Δ, Mackey–Glass, $\epsilon=0.05$.

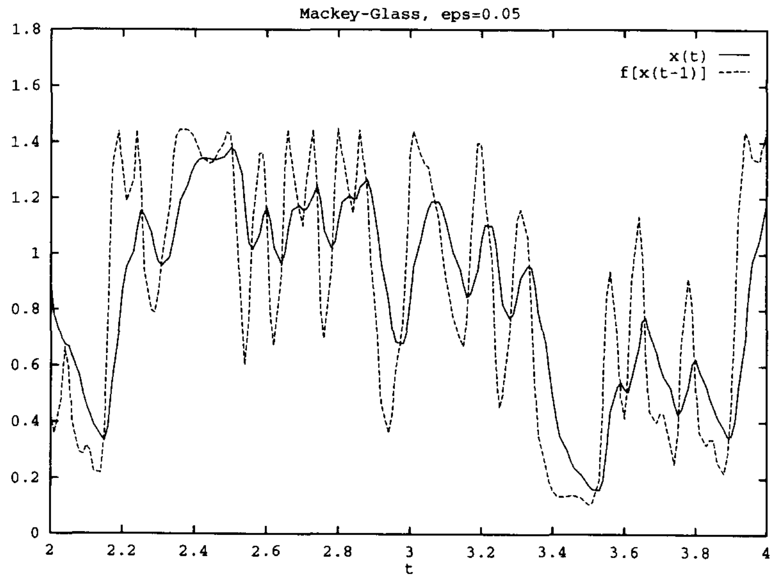

Fig. 7. $x(t)$ (solid line) and $f(x_1)$ (dotted) as functions of t.

the equation as an integral equation, neglecting transients, and expanding in powers of ϵ. Equivalently, we use operational calculus to write

$$x=(1+\epsilon D)^{-1}f_1 = f_1 - \epsilon\dot{f}_1 + \epsilon^2\ddot{f}_1 - \epsilon^3\dddot{f}_1 + \dots\,, \qquad (7)$$

which suggests that $x \approx f_1$ at leading order. However, as we expect $\mathrm{d}/\mathrm{d}t \sim 1/\epsilon$, there is in fact no guarantee that the terms diminish. Nevertheless, (6) is suggestive, as are successive finite difference approximations:

$$x(t+\epsilon)=f_1+\tfrac{1}{2}\epsilon^2\ddot{f}_1+\dots,$$

$$x(t+2\epsilon)-x(t+\epsilon)+x(t)=f_1+\tfrac{7}{6}\epsilon^3\dddot{f}_1+\dots. \quad (8)$$

Qualitatively, we can partly understand fig. 6 by considering fig. 7, which shows x and f_1 versus t. Since x tracks f_1, always tending exponentially towards it, it follows that x has the same shape as f_1, but retarded by a time of $O(\epsilon)$. It is because of this that fig. 6 has a minimum at $\Delta=1+O(\epsilon)$. Nevertheless, it is clear that the resemblance is only qualitative.

5. In conclusion, we have shown that the use of variable, "smart" embeddings can be used to diagnose the presence of delay in chaotic time series, and that they can be used to make improved predictions, particularly in data-poor, high-dimensional series. For this particular delay-recruitment model, there is a close relationship between the map (the singular limit) and the equation, whose exact nature remains opaque, however.

Acknowledgement

This work was partly carried out at the CNLS of Los Alamos National Laboratory, and supported by U.S. DoE contract KC-07-01-01-0. We thank Pat Hagan for hospitality.

References

[1] F.S. Grodins, J. Buell and A.J. Bart, J. Appl. Physiol. 22 (1967) 260.
[2] M.C. Mackey and L. Glass, Science 197 (1977) 287.
[3] R.M. May, J.R. Beddington, C.W. Clark, S.J. Holt and R.M. Laws, Science 205 (1979) 267.
[4] I. Ikeda and K. Matsumoto, Physica D 29 (1987) 223.
[5] M. Casdagli, Chaos and deterministic versus stochastic nonlinear modelling, J.R. Stat. Soc. B 54 (1992) 303.
[6] J.D. Farmer, Physica D 4 (1982) 366.
[7] S.-N. Chows and J. Mallet-Paret, in: North-Holland mathematical studies, Vol. 80. Coupled nonlinear oscillators, eds. J. Chandra and A.C. Scott (North-Holland, Amsterdam, 1983) pp. 7–12.
[8] P. Grassberger, T. Schreiber and C. Schaffrath, Int. J. Bifurcation Chaos 1 (1991) 521.
[9] L.A. Smith, Phys. Lett. A 133 (1988) 283.