ELSEVIER

Physica D 110 (1997) 182–194

PHYSICA D

Inferring mechanism from time-series data: Delay-differential equations

Stephen P. Ellner [a,*], Bruce E. Kendall [b], Simon N. Wood [c], Edward McCauley [d], Cheryl J. Briggs [e]

[a] *Biomathematics Graduate Program, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA*
[b] *National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA 93106, USA*
[c] *School of Mathematical and Computational Sciences, University of St Andrews, Fife, Scotland KY16 9SS, UK*
[d] *Ecology Division, Department of Biological Sciences, University of Calgary, Calgary, Alta., Canada T2N 1N4*
[e] *Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106, USA*

Received 7 February 1997; received in revised form 2 May 1997; accepted 7 May 1997
Communicated by A.M. Albano

Abstract

When there is *qualitative* information about the underlying processes and structure of a dynamical system, it may be possible to infer very accurate *quantitative* information about these processes using only an output time series from the system. We illustrate how this can be accomplished for time series data from a delay-differential equation with a single fixed delay. Our approach exploits modern techniques for non-parametric function estimation, is robust to fairly high levels of dynamic noise and measurement error, and can be extended straightforwardly to more general delay-differential systems and multivariate systems.

Keywords: Time series analysis; delay-differential equations; population dynamics

1. Introduction

The dynamics of many systems are affected not only by the current state of the system but also by past states. In some cases there is a single fixed delay which is important: these systems can be described by $\dot{x}(t) = F(x(t), x(t-\tau))$, where τ is a constant. These delay-differential equation (DDE) models have been successfully used to describe a wide variety of phenomena in ecology [1], physiology [2], chemistry [3], and physics [4].

In many cases the qualitative nature of the problem suggests a general structure for F even if the quantitative details are unknown. In ecology, for example, we can sometimes determine that the equation should be of the form

$$\dot{x}(t) = f(x(t-\tau)) + g(x(t)), \tag{1}$$

where x is adult population density, τ the development time of juveniles, f the recruitment (number of offsprings born at time $t-\tau$ less juvenile mortality), and g is the number of adult deaths. In these circumstances we would like to be able to estimate τ, f, and g.

Bünner et al. [5] recently introduced a method to estimate τ from a time series of the process (1) in the

* Corresponding author.

PII S0167-2789(97)00123-1

special case of $g(x(t)) = -x(t)$. This method takes advantage of the fact that at times t_i when $\dot{x}(t) = 0$, $x(t_i) = f(x(t_i - \tau))$. A plot of $x(t_i)$ versus $x(t_i - \tau)$ therefore gives a set of points on the graph of f. It is then straightforward to estimate τ by trying different values to find the value that gives the smoothest graph. A major attraction of this method is that there is no need to directly fit f (which may not be given by any simple formula). The method is surprisingly robust: we have found that it performs well on simulated datasets considerably shorter than those analyzed in [5]. A similar method was proposed earlier by Fowler and Kember [6]. Their method uses all values of $x(t)$, and estimates the derivative from the finite difference $(x(t) - x(t - \delta))$. Otherwise the approach is similar to [5] in that the correct delay is identified by finding the value that gives a smooth, low-dimensional embedding in $x(t)$, $x(t-\delta)$, $x(t-\tau)$ phase space. Essentially the same idea also underlies the method of Kaplan and Glass [7], in which the time lag is inferred from the alignment of vector fields in a time-delay embedding of the data. Indeed, if the data obey Eq. (1) and are measured with little or no error, simple visual inspection of two-dimensional embeddings of the time series may be all that is needed to identify the value of τ.

Our goal here is to unify and extend these approaches, in order to accommodate some features typical to biological applications. First, even if g is linear that fact would not be known a priori, so it cannot be an assumption of the method. Second, the data are often sparse relative to what can be obtained in controlled physical experiments (e.g. 100–1000 data values, with only a dozen or so extrema). Third, the data are noisy in two senses: the dynamics are not completely deterministic, and the measurement errors may be appreciable (1–10%). Measurement error in particular makes it difficult to reach unambiguous conclusions by visual inspection of a reconstructed vector field. Finally, despite these data limitations, it is desirable to estimate the rate equations f and g, as well as the delay τ, in order to obtain insight into the underlying mechanisms. Each of the approaches cited above can cope with some of these features, but none can cope with all of them simultaneously.

We proceed by noting that Eq. (1) is in the form of a generalized additive model (GAM), for which there is a well-developed set of statistical tools [8]. In brief, our method consists of:

(i) estimate $\dot{x}(t)$ from the time series $x(t)$ (we denote this estimate $y(t)$);
(ii) for each of a range of τ's, do a GAM fit of Eq. (1), using splines to estimate f and g;
(iii) pick the τ for which the model explains the most of the variance in $y(t)$.

We describe the method in detail in Section 2, in which we apply it to a case in which g is linear, but the data are short and sparse. This is followed by sections in which we illustrate its performance in the cases of nonlinear g and data with substantial dynamic noise or measurement error.

2. No noise, g known

We start by letting g be linear and assuming that we know this fact. The time series we analyze is generated by a simple Euler integration of Eq. (1), with $f(x) = 10xe^{-x}$, $g(x) = -mx$, $\tau = 25$, and $m = 0.27$. This generates a chaotic time series (Fig. 1). This system is qualitatively very similar to the Mackey–Glass equation [2]. We used the exponential form of f because it has been used successfully in models for laboratory insect populations with discrete life-stages, which lead in some cases to Eq. (1) with $x(t)$ the number of reproductively mature females [1,9].

The first step is to estimate $\dot{x}(t)$ from the time series. This is done by smoothing the data and numerically estimating the slope at each value of t. (We use $\dot{x}(t)$ to denote the true but unknown rate of change of x, and $y(t)$ to denote the estimate of $\dot{x}(t)$ obtained from the data.) The choice of smoothing method depends on the nature of the data: there is usually a trade-off between an ability to fit at sharp peaks or troughs where the second derivative is large, and a tendency to fit any noise that might be in the data. For these data we used a method that performs well at sharp peaks, the locally weighted regression smoother LOESS [10,11].

Next we fit Eq. (1), using a generalized additive model, for a variety of values of τ. Generalized

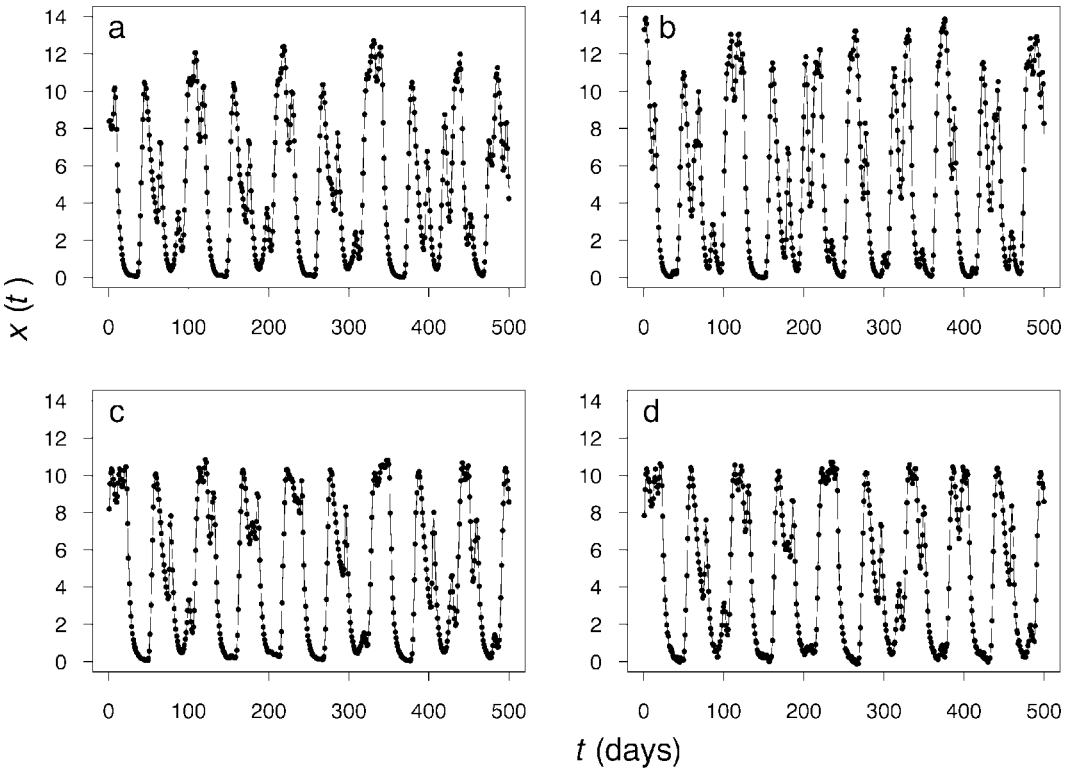

Fig. 1. Model time series analyzed in this paper. $\tau = 25$ in all cases. (a) $f(x) = 10xe^{-x}$; $g(x) = -mx$; $m = 0.27$ (Eq. (1)). (b) As in (a) but with dynamic noise (Eq. (3)) and $f(x) = 12xe^{-x}$. (c) As in (b) but $g(x) = -(0.21 + 0.018x)x$. (d) As in (c) but with measurement error added.

additive models (GAM) are a recent extension of multivariate linear regression which allow nonlinear functional relationships [8]. The form of the model is

$$y = F(\boldsymbol{X}) = \alpha + f_1(x_1) + f_2(x_2) + \cdots + f_d(x_d), \quad (2)$$

where $\boldsymbol{X} = (x_1, x_2, \ldots, x_d)$ is the set of independent variables and y is the dependent variable. In linear regression, each f_i is linear. In GAM each f_i is arbitrary: the user can assume a specific form (e.g., quadratic, sigmoid, etc.) or can specify a general flexible form, such as a spline, neural net, radial basis function, limited only by the availability of numerical software for the specific form. To make α well-defined, we adopt the convention $f_i(0) = 0$.

Here we take, for convenience, each f_i to be an extended family of B-splines [12]. The convenient aspects are that a GAM with this family is easy to fit using the `gam()` procedure in the statistical language S–PLUS [11], that the way in which the splines are parametrized reduces the fitting of (2) to a computationally trivial linear regression, and that this space of functions encompasses complexities ranging from constant functions up to complex piecewise polynomials. The potential complexity of the curve is determined by the "degrees of freedom" (df) which is the number of free parameters to be fitted. When $\mathrm{df} < 3$ the functions are polynomials of order df; because of our convention that $f_i(0) = 0$, the constant term is not included. When $\mathrm{df} \geq 3$ the functions are piecewise cubic splines with continuous second derivatives. The parameters of these spline curves are their values at a finite number of points (called "knots") within the range of the independent variable. The model fitting process then includes selection of an appropriate value of df_i for each of the f_i based on the data.

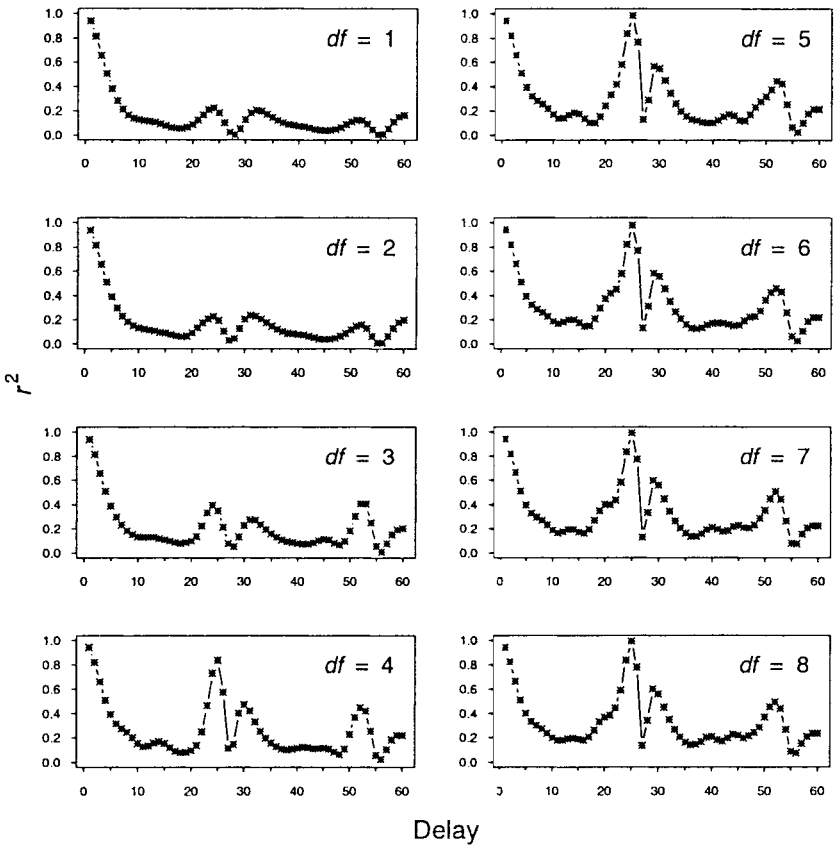

Fig. 2. The r^2 for different values of τ and df.

To quantify the "goodness of fit" we first generate predicted values of the time derivative at each time, $\hat{y}(t)$, by evaluating the fitted model at $(x(t), x(t-\tau))$. We then calculate the proportion of the variation in $y(t)$ that is explained by the model,

$$r^2 = 1 - \frac{\sum(\hat{y}(t) - y(t))^2}{\sum(y(t) - \bar{y})^2}, \tag{3}$$

where $\bar{y}$ is the mean value of the derivative values $y(t)$ estimated from the smoothed data. Fig. 2 shows the r^2 for several values of df. The r^2 is always very high at $\tau = 1$. This is to be expected, since $\dot{x}(t) \approx x(t) - x(t-1)$. Thus, we seek a peak for a value of τ greater than 1. For df sufficiently large, there is a strong peak at $\tau = 25$, the correct value. The minor peak at $\tau = 52$ appears to result from approximate symmetries in the data: the peaks and troughs are about equally wide and the rising and falling data segments are similar. If these symmetries were perfect, and the data were perfectly periodic with period T, then $x(t + T/2)$ would be a linear transformation of $x(t)$ and we would get equally good predictions using delays τ or $\tau + T/2$. In our case $T/2$ is about 27, giving the minor peak at delay 52.

How can we objectively choose the "best" value of df? Increasing df always increases the number of parameters, so larger and larger dfs eventually produce a larger r^2. Standard measures of the "value" of additional parameters, such as the Bayes Information Criterion, typically picked the largest value of df tried, in part because the overall fit is so good. However, there is useful information for selecting df in the structure of the residual errors from the model. If the model is correct, the errors should be iid (independent and identically distributed). Evaluating a model by testing the residuals for departures from iid is a standard approach in time series analysis [13]. We

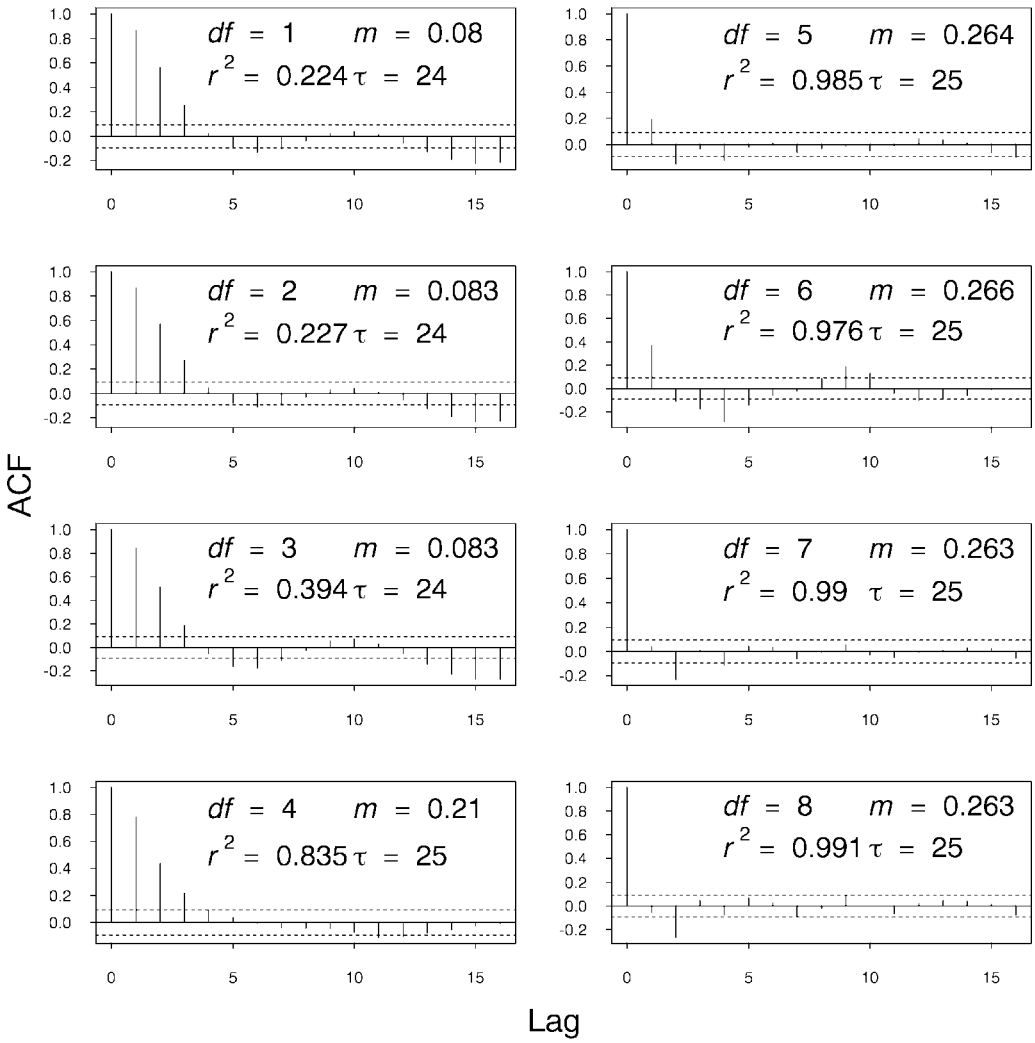

Fig. 3. ACF of the residual errors of the fitted model at the best τ's for a range of df. m is the value estimated by the fit. The dotted lines indicate the 95% confidence bands: autocorrelations exceeding them differ significantly from zero.

have found it useful to focus on the independence aspect, using the autocorrelation function (ACF) as the diagnostic for non-independence (Fig. 3). When the model is too simple, the time series of residuals has significant positive autocorrelations at low lags, which are diminished by increasing the model complexity. In this case, when df < 7 there is a significant positive autocorrelation at lag 1. We thus choose df $= 7$ as the appropriate model. Notice however the persistent negative correlation at lag 2 when df is 7 or 8. This remains for all more complex models, and is a result of autocorrelations created by the smoothing process used to estimate the time derivative of $x(t)$. Thus our real criterion for model selection is to find the simplest model such that further increases in model complexity do not bring the residual time-series closer to the ideal of lacking any significant autocorrelations.

The final estimates of f and g are both quite good (Fig. 4; $\hat{m} = 0.261$). Although the errors in f at df $=$ 5 are small, they are systematic; df $= 7$ corrects most of them. There is still a systematic bias in the tail of f, however. If we have a little more qualitative

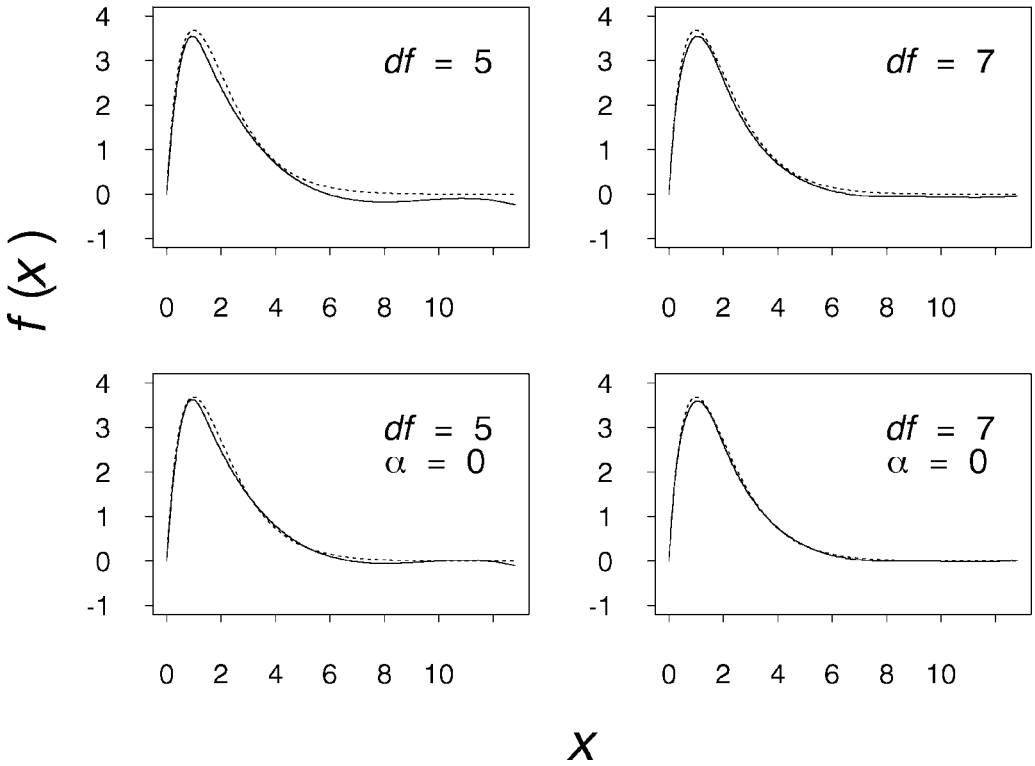

Fig. 4. The estimated f from the fitted model with $\tau = 25$ and df = 5 or 7, α forced to zero or not. The dotted curves are the true value of f. The estimate of m is 0.26 in all four cases.

information about the system, we can do even better. Suppose we know that $f(0) = g(0) = 0$ – that is, $\alpha = 0$ (this is often to be expected if the system under study is closed). Then we can eliminate the constant term in the GAM, and get a slightly better fit. We maintain the assumption of $\alpha = 0$ in the analyses that follow, but it makes little qualitative difference in the results.

The above analysis is for 500 days of data (approximately 10 oscillations), sampled once per day. The method can successfully select τ for as few as 70 days of data (less than two oscillations; Table 1).

If there is an error in the choice of τ, then the estimates of f and m are severely compromised. τ will not typically be a multiple of the sampling interval. We have simulated this situation and found that we can recover the true value of τ by interpolating the smoothed time series to a finer sampling interval. Thus, we can recover $\tau = 25$ from 100 days of data sampled every 2 days, or 200 days of data sampled every 4 days; and we can recover the true value of τ from 300 daily samples of a time series generated by the model with $\tau = 24.7$. In the latter example the residuals are always autocorrelated at lag 1, even for models with very high values of df, as a result of autocorrelations in the interpolation errors. However, there is a point where the autocorrelations essentially stop changing with increasing df, and we use this to

Table 1
Fits for varying numbers of data points (N) with df = 7

N	r^2	$\hat{m}$
500	0.989	0.261
400	0.987	0.261
300	0.987	0.261
200	0.983	0.258
100	0.988	0.236
70	0.986	0.245

In all cases $\hat{\tau} = 25$ and the true value of m is 0.27. At $N = 50$ the method fails to identify an optimal value of τ.

choose the model complexity. A very good estimate of m is obtained (0.255), and the estimate of f is only marginally less accurate than those shown in Fig. 4.

3. Dynamic noise, unknown g

We next examine how our method performs in less idealized settings. First, in many biological applications we would not know a priori that $g(x)$ is linear. We now drop the assumption of linear g and assume only that the data were generated by Eq. (1), with f and g both unknown smooth functions. A second inevitable complication in many biological applications is dynamic noise (random perturbations to the state variable or parameters), even in the laboratory. Motivated by the interpretation of (1) as a population model with delayed recruitment to the class of reproductively active individuals, we modify Eq. (1) to include random variation in the recruitment term. The model is then

$$\dot{x}(t) = a(t) f(x(t-\tau)) + g(x(t)), \tag{4}$$

where $a(t)$ has random fluctuations. We assume first that $a(t)$ is effectively a white noise process, with short correlation time compared to the time scale of changes in $x(t)$.

The procedures described above should in principle require only one modification, namely that the selection of τ and the model-fitting require a scan over df for the estimates of both g and f. Dynamic noise adds some scatter to the observed "map" from $(x(t), x(t-\tau))$ to $\dot{x}(t)$, but it does not fundamentally change the problem.

As a trial of this approach we analyzed data generated by Euler integration of Eq. (4) with $\tau = 25$, $f(x) = 12x\mathrm{e}^{-x}$, and either $g(x) = -0.3x$ or $g(x) = -(0.21 + 0.018x)x$ (Figs. 1(b) and (c)). Our procedures should be able to estimate f and g in both cases, and in the former identify that g is actually linear. We used $a(t)$ lognormal with parameters $\mu = 0$, $\sigma = 0.1$, which is a substantial amount of noise. For example, a typical run of 100 $a(t)$ values will include some that are 20% or more above or below the mean.

As before, we first estimate $\dot{x}(t)$ by smoothing $x(t)$. To estimate τ we fitted GAMs with f and g in our extended spline family (with degrees of freedom df_f and df_g, respectively), for $\tau = 6$–60 and all df pairs with $1 \le \mathrm{df}_g, \mathrm{df}_f \le 8$. The optimal value of τ for each df pair is the value at which r^2 is maximized.

For both the linear-g and the quadratic-g data, the r^2 was maximized at the correct value of τ for all $\mathrm{df}_f \ge 5$, $\mathrm{df}_g \ge 1$ for linear g and $\mathrm{df}_g \ge 2$ for quadratic g. Then to estimate f and g we fitted GAMs at each df pair using this value of τ, and examined the goodness of fit (r^2), and the autocorrelation function (ACF) of the residual time series.

When g is linear, the ACF plots (Fig. 5) indicate that there is no improvement from increasing df_g above 1. That is, if we scan down any column in Fig. 5, which corresponds to fits with the same value of df_f, the ACFs in lower rows ($\mathrm{df}_g > 1$) are no better than those in the top row ($\mathrm{df}_g = 1$). Thus, the linearity of g is identified correctly. Then scanning in the same way across the top row, df_f is picked to be 7, which is complex enough to approximate closely the true f. As a result, the estimates of f and g are both very accurate (Fig. 6).

When g is quadratic the ACF plots clearly indicate a need for $\mathrm{df}_g > 1$ to eliminate autocorrelation (Fig. 7). The ACFs in the second row are markedly better than those in the first, but no worse than those in lower rows, so df_g is identified as being 2, which is correct. Scanning across the second row to choose df_f, the choice between $\mathrm{df}_f = 5$ versus $\mathrm{df}_f = 7$ is less clear. However, with either choice the estimates remain fairly accurate (Fig. 6). As the high r^2 values indicate, the estimates of f and g are robust to reasonable changes in the df values for both linear and quadratic g.

The successful fitting in this case depends on the short autocorrelation time of $a(t)$. If on the other hand the autocorrelation time of $a(t)$ is comparable to or longer than the sampling interval, then the situation is really one of random parameter drift. This can create problems at the initial step of estimating the $\dot{x}(t)$ time series. If $a(t)$ changes slowly, there will be stretches of data during which $\dot{x}(t)$ has a consistent tendency to be higher than its expectation conditional on x,

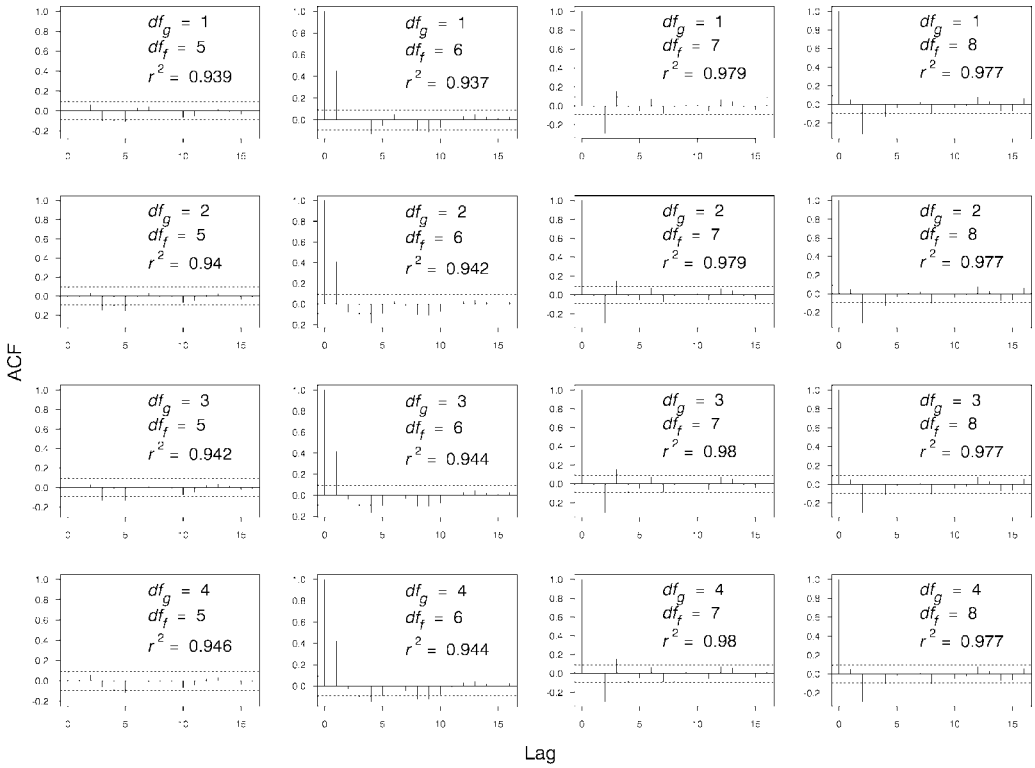

Fig. 5. ACF of residual errors from the fitted model for the linear-g system with dynamic noise, at a range of df values.

and other stretches during which $\dot{x}(t)$ will tend to be lower than its conditional expectation. This means that residuals from fitting model (1), which specifies the conditional expectation, will be autocorrelated even if the model is exactly correct, and the model therefore would be rejected. That would be technically correct, but it would be wrong about the mechanistic validity of the model structure. This problem occurs in our example system even if the autocorrelation in $a(t)$ is fairly weak, with correlation coefficient $\rho = 0.1$ between values at successive sampling times.

4. Measurement errors

Small measurement errors can be ignored; large measurement errors can be reduced by smoothing the data (e.g., low-pass filtering). The former is apparently safe (we have in fact been using 1% measurement error throughout the last two sections), but the latter is not because of the autocorrelations that it induces. Before smoothing, successive measurement errors are independent. After smoothing they are smaller but autocorrelated, and this creates structure in the residuals that would cause the model to be rejected even if f and g are estimated accurately. Therefore, we use the raw data to estimate model complexity, and only afterwards use smoothed data to estimate f and g.

Because measurement errors affect both the independent and dependent variables at the model fitting step, they can have more drastic effects than dynamic noise, which only affects the dependent variable. For the dynamical noise model of Section 3, adding simulated measurement errors to the output (Fig. 1(d)) did not affect the identification of τ, but the goodness of fit (r^2) is greatly reduced, the ACF plots are less informative (Fig. 8), and the function estimates are poorer (Fig. 9). The peak in the estimated f is rounded off because of the random jitter in the $x(t-\tau)$ values. This is a general property of measurement error:

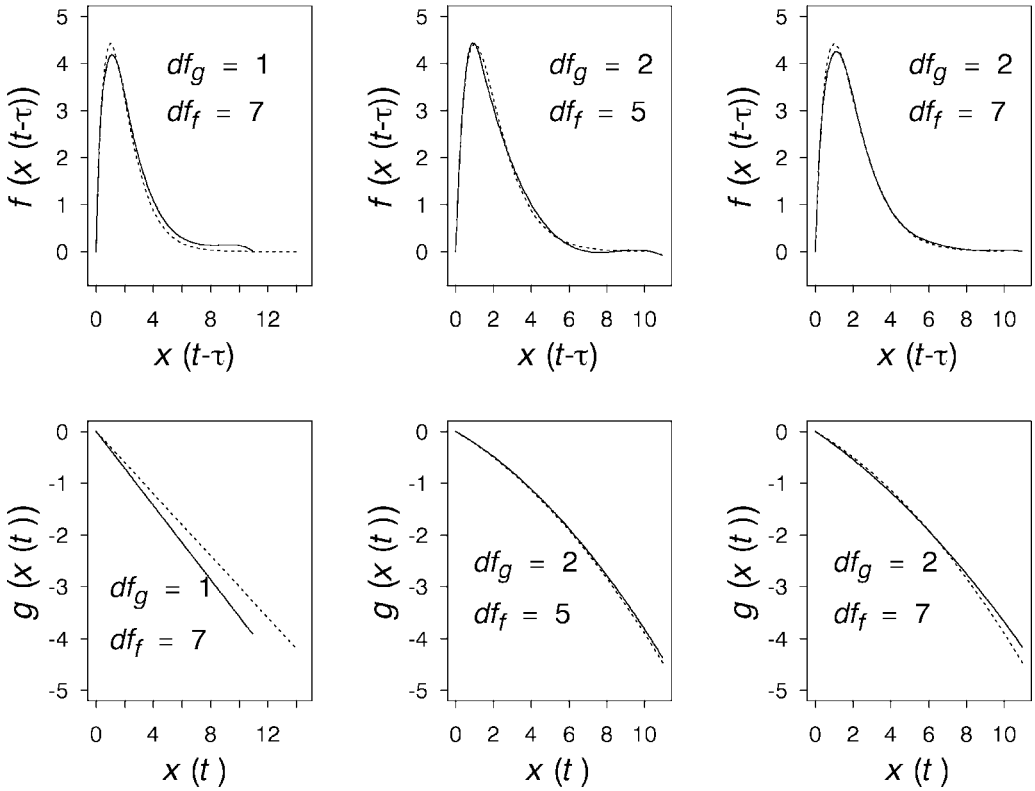

Fig. 6. Estimates of g and f for the linear and quadratic-g systems with dynamic noise.

because it obscures fine structure in the data, such as sharp peaks, there is a bias towards too-simple models. Still, we find these results encouraging because they involve quite large measurement errors: lognormal with $\sigma = 0.1$. As noted above, this distribution implies that a run of 100 data points typically includes errors as large as 20% of the mean.

5. Discussion

We have shown that it is possible to reconstruct the form of the underlying equations for a scalar delay-differential equation from a moderate amount of data (100–500 data points). With 500 fairly accurate data points (1% measurement error) the estimated rate equations were highly accurate even in the presence of substantial dynamic noise. Our approach has two essential ingredients: expressing the problem so that it fits into the framework of generalized additive models, and using the ACF of the residual time series to choose appropriate degrees of freedom in the fitted rate equations. In comparison with previous methods for delay recognition in differential-delay systems [5–7], our approach offers two advantages: (1) it remains accurate on short data series, even at relatively high levels of dynamic noise; (2) it provides quantitative estimates of the individual rate equations, which may be informative about underlying mechanisms.

There are two extensions that increase the generality of the approach, but also increase the computational complexity somewhat. The first extension explicitly addresses the fact that the smoothing process used to estimate derivatives and lagged values of x will often introduce substantial autocorrelations into the derivative estimates. For many smoothing methods (including locally weighted quadratic regression) the covariance matrix for the estimated derivatives y can

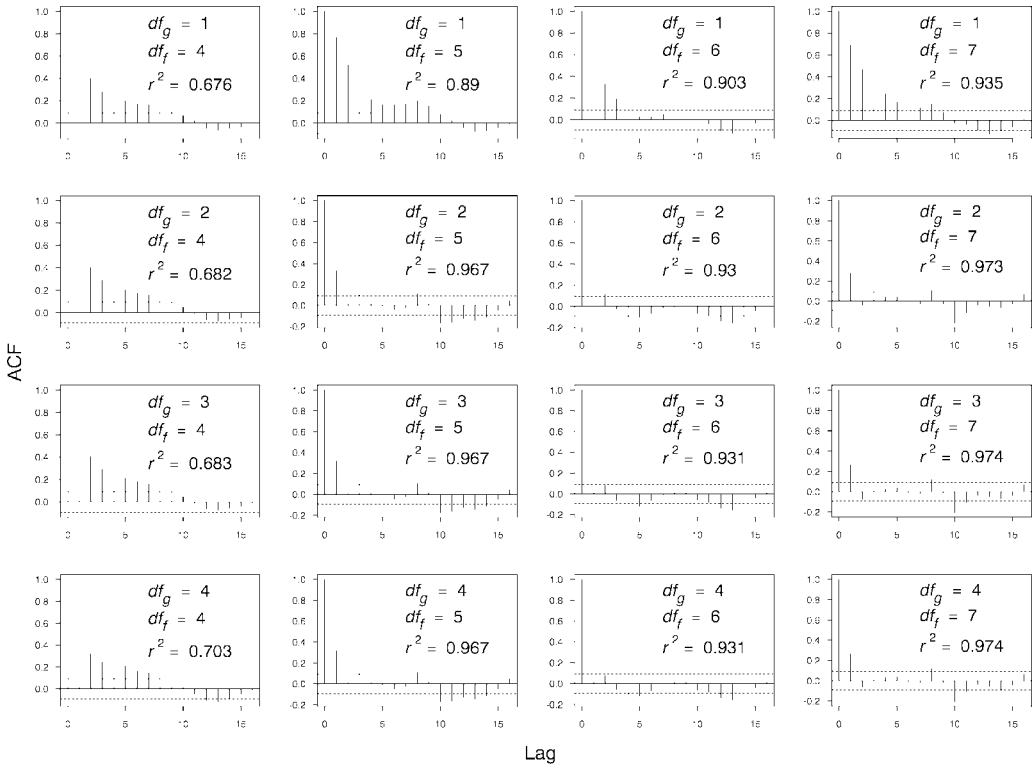

Fig. 7. ACF of residual errors from the fitted model for the quadratic-g system with dynamic noise, at a range of df values.

be calculated or estimated. The covariances can be accounted for by an appropriate weighting in the least squares minimization. The fitting criterion we used in this paper (maximize r^2) is equivalent to replacing the inverse covariance matrix with the identity matrix, an approximation that is reasonable for well sampled data subject to little measurement error. For the data sets studied here, estimates were not improved noticeably by weighting to account for the covariance. However, weighting will be more important with noisier data that require a greater degree of smoothing.

The second extension involves a general multivariate model,

$$\dot{X} = F(X(t), X(t-\tau_1), X(t-\tau_2), \ldots, X(t-\tau_d)),$$

in which F is not additive, but the complete state vector $X(t)$ is again observed. Again the model may contain unknown functions and again the lag(s) will generally be unknown. In this case the model may be fitted by non-linear least squares, which involves far more computational effort than that required for the model in this paper, and which is not guaranteed to converge to the global best fit. Nevertheless such techniques, if properly implemented, generally have better convergence properties than the alternative methods for the same model based on matching observed and predicted trajectories (Wood, in preparation).

Although our results and procedures are specific to the differential-delay system (1), we believe that this case study illustrates some points of general importance for model selection and fitting in the context of nonlinear time series analysis.

First, criteria based solely on overall prediction accuracy were not successful at selecting model complexity. In all of the cases considered here, very high r^2 values were achieved by models that were revealed as too simple by calculating the ACF of the residuals. The Bayes information criterion (BIC) often preferred overly complex models, or overly simple models, depending on the levels of measurement and dynamic

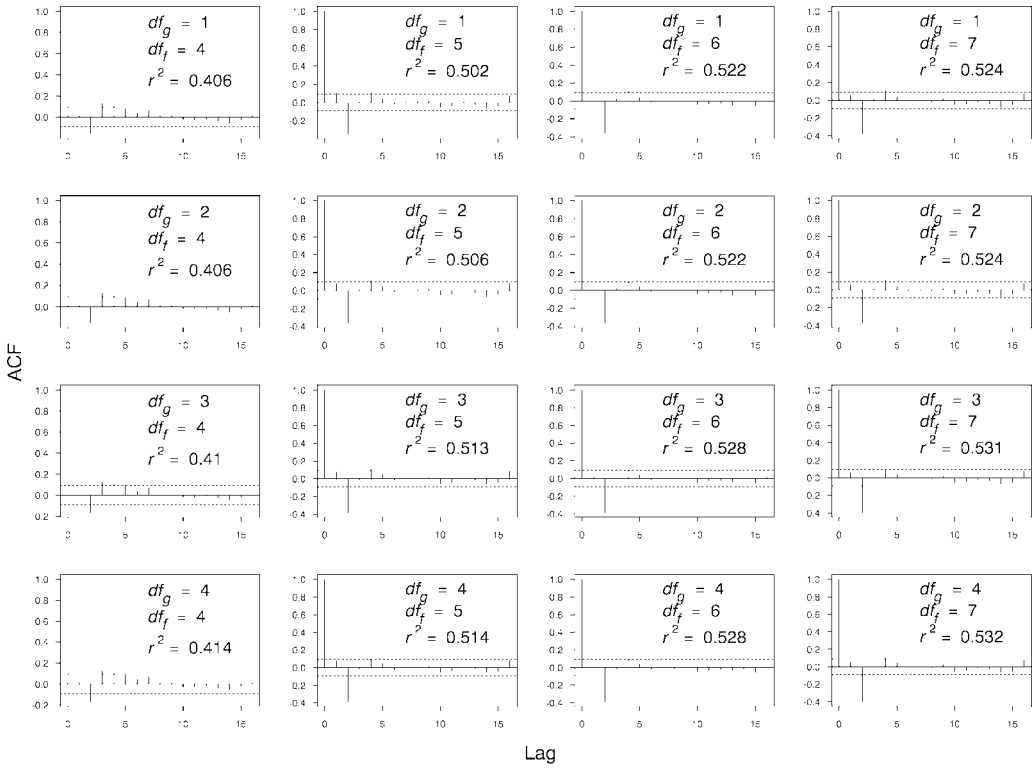

Fig. 8. ACF of residual errors from the fitted model for the quadratic-g system with dynamic noise and measurement error. Inaccuracies at the smoothing stage lead to persistent autocorrelations in the residuals, so the choice of df is less clear-cut than with more accurate data.

noise. It is easy to understand why overall prediction accuracy is not sufficient for model selection, once we realize how much information is discarded [14]. Any candidate model produces a time-series of residuals $e(t)$. Both the raw prediction accuracy and complexity-penalized criteria (such as BIC or the minimum description length criterion [15]) are based on the mean value of $e(t)^2$, a single number. By reducing the residual time series to this single number we lose all information on spatial and temporal structure in the residuals, and on correlations between the residuals and state variables, that can be very informative for evaluating the model.

Residual ACF plots are only one of many ways to examine the time series of residuals for structure that signals a flawed model. Spatial autocorrelation (e.g., Moran's I statistic [16]) might circumvent some of the difficulties that we had in choosing model complexity (where "space" refers to the $(x(t), x(t-\tau))$ state space). The temporal autocorrelations introduced by the process of smoothing $x(t)$ to estimate its derivative would introduce artifactual spatial autocorrelations, but these could be eliminated by the standard device of omitting temporal neighbors from the calculation. However, spatial autocorrelation appears to present difficulties of interpretation, and in our experience it commonly rejects even the correct model if its parameters were estimated from data. The cause of this problem is that even if the (unobserved) errors between the true equation and the data were independent in time and space, the residuals from a fitted estimate of the equation are correlated [16]. However, if the variance–covariance matrix of the residuals is known, then a linear transformation can be used to obtain uncorrelated "pseudo-residuals" that are suitable for a spatial autocorrelation analysis [16]. As

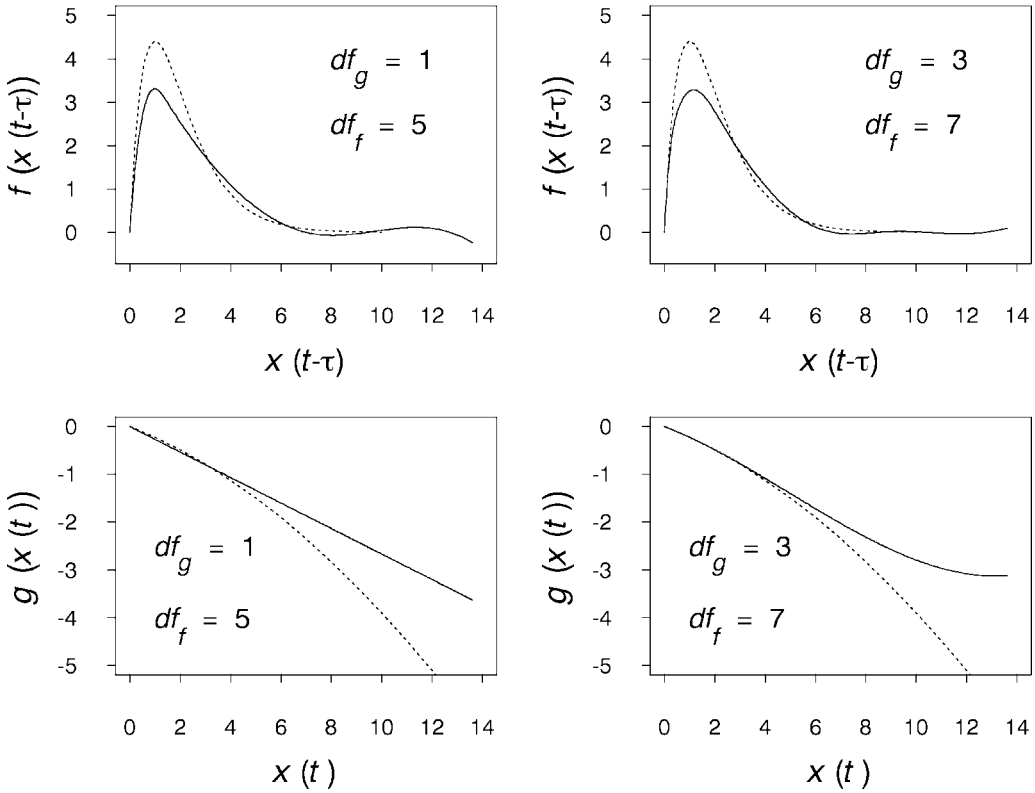

Fig. 9. Estimates of g and f for the quadratic-g system with dynamic noise and measurement error. df $= (1, 5)$ is really too simple, but would be selected based on the ACF plots, because the inaccurate data do not allow an increase in accuracy by fitting more complex models. Increasing df to adequate values (right column) hardly changes the estimates.

noted above, smoothers very similar to the one used here have this property. We are currently implementing and testing this approach (Seifu and Ellner, in preparation). In practice, however, ACFs appear to be adequate for selecting model complexity, if we accept that there will be persistent temporal autocorrelations at small lags due to the initial smoothing stage.

Another type of test for model mis-specification is to check for correlations between residuals and measured system variables, which should be absent if the model is correct. Statisticians usually do this visually. In our case one might plot the residuals versus present or lagged values of $x(t)$. Manuca and Savit [17] have recently presented a way of looking for such dependencies in high-dimensional models, by testing for an increase in apparent determinism when an extra variable is added to the information used for nonlinear forecasting.

The second general point is that it can be extremely valuable to have qualitative information on the structure of the dynamics, to constrain the search in "model space". It is particularly useful, in our experience, to have some guidance on the choice of state variables to include in the model. Model-free approaches, such as attractor reconstruction by time-delay embedding, can provide qualitative and quantitative information about the dynamics, but they do not provide any insights into the processes underlying those dynamics. A model developed using our approach might not have greater predictive power than a conventional delay-embedding model and might not estimate τ more accurately, but it gives explicit approximations of the underlying functional relationships. This allows us to

include qualitative constraints (for example if f represents a birth rate it must be non-negative), which can further ease the task of model selection. On the other hand, if the best-fit unconstrained model grossly violates an expected constraint, this would lead us to re-evaluate our qualitative understanding of the underlying processes.

Acknowledgements

This work was conducted as part of the Complex Population Dynamics Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (grant no. DEB-94-21535), the University of California at Santa Barbara, and the State of California. We thank the remaining members of the Working Group – W.W. Murdoch, R.M. Nisbet and P. Turchin – for discussions that inspired this work and D. Kaplan, G.A. Fox, A. Bohanok, N.G. Hairston Jr. and L. Glass for comments on the manuscript. SE was partially supported by a Japan Society for Promotion of Science Visiting Research Fellowship at Kyushu University.

References

[1] W.S.C. Gurney, S.P. Blythe and R.M. Nisbet, Nicholson's blowflies revisited, Nature 287 (1980) 17–21.
[2] M.C. Mackey and L. Glass, Oscillations and chaos in physiological control systems, Science 197 (1977) 287–289.
[3] K.W. Smith and R.M. Noyes, Gas evolution oscillators. 3. A computational model of the Morgan reaction, J. Phys. Chem. 87 (1983) 1520–1524.
[4] K. Ikeda, Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system, Opt. Commun. 30 (1979) 257–261.
[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] A.C. Fowler and G. Kember, Delay recognition in chaotic time series, Phys. Lett. A 175 (1993) 402–408.
[7] D.T. Kaplan and L. Glass, Coarse-grained embeddings of time series: Random walks, gaussian random process, and deterministic chaos, Physica D 64 (1993) 431–454.
[8] T.J. Hastie and R.J. Tibshirani, Generalized Additive Models (Chapman and Hall, London, 1990).
[9] W.S.C. Gurney, R.M. Nisbet and J.H. Lawton, The systematic formulation of tractable single-species population models incorporating age structure, J. Animal Ecology 52 (1983) 479–495.
[10] S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Statist. Assoc. 74 (1979) 829–836.
[11] Statistical Sciences, S-PLUS Guide to Statistical and Mathematical Analysis, Version 3.2, StatSci, a division of MathSoft, Inc., Seattle (1993).
[12] G. Wahba, Spline Models for Observational Data (SIAM, Philadelphia, PA, 1990).
[13] P. Kennedy, A Guide to Econometrics, 3rd Ed. (MIT Press, Cambridge, MA, 1992).
[14] L.A. Smith, Locally optimized prediction of nonlinear systems: Stochastic and deterministic, in: Chaos and Forecasting, ed. H. Tong (World Scientific, Singapore, 1995) pp. 87–108.
[15] K. Judd and A. Mees, On selecting models for nonlinear time series, Physica D 82 (1995) 426–444.
[16] A.D. Cliff and J.K. Ord, Spatial Autocorrelation, Pion, London (1973).
[17] R. Manuca and R. Savit, Model misspecification tests, model building and predictability in complex systems, Physica D 93 (1996) 78–100.