Estimating Parameters in Delay Differential Equation Models

Liangliang WANG and Jiguo CAO

Delay differential equations (DDEs) are widely used in ecology, physiology and many other areas of applied science. Although the form of the DDE model is usually proposed based on scientific understanding of the dynamic system, parameters in the DDE model are often unknown. Thus it is of great interest to estimate DDE parameters from noisy data. Since the DDE model does not usually have an analytic solution, and the numeric solution requires knowing the history of the dynamic process, the traditional likelihood method cannot be directly applied. We propose a semiparametric method to estimate DDE parameters. The key feature of the semiparametric method is the use of a flexible nonparametric function to represent the dynamic process. The nonparametric function is estimated by maximizing the DDE-defined penalized likelihood function. Simulation studies show that the semiparametric method gives satisfactory estimates of DDE parameters. The semiparametric method is demonstrated by estimating a DDE model from Nicholson's blowfly population data.

Key Words: Dynamic system; Implicit function theorem; Mackey–Glass equation; Nicholson's data; Penalized spline smoothing.

1. INTRODUCTION

Delay differential equations (DDEs) are popular tools for applied scientists to model their dynamic systems. The forms of the DDE models are usually proposed by investigators based on their understanding of the dynamic system. However, the parameters of DDE models are often unknown. These parameters often have important scientific interpretations, therefore it is important to infer their values. It is also a good calibration of the DDE models if the DDE solutions with certain parameter values fit the measurements of the dynamic systems. The objective of this article is to estimate the DDE parameters from noisy data.

Liangliang Wang is a PhD Candidate, Department of Statistics, University of British Columbia, 6356 Agricultural Road, Vancouver, BC, V6T1Z2, Canada. E-mail: *l.wang@stat.ubc.ca*. Jiguo Cao (✉) is an Assistant Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A1S6, Canada. E-mail: *jca76@sfu.ca*.

Journal of Agricultural, Biological, and Environmental Statistics, ***Accepted for publication***
DOI: 10.1007/s13253-011-0066-6

Published online: 05 October 2011

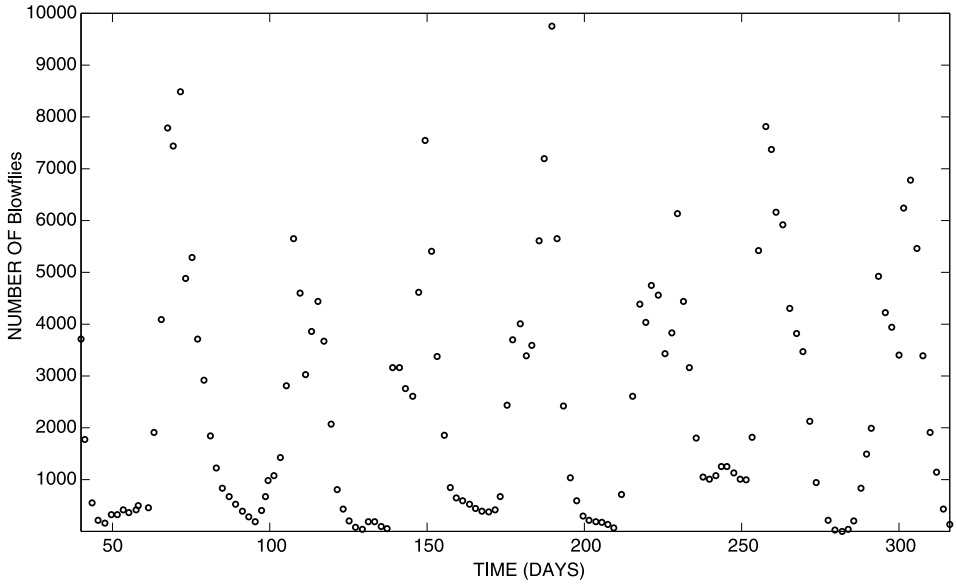

Figure 1. Blowfly population in one experiment published in Nicholson (1954b).

Nicholson (1954a, 1954b, 1957) published his classic studies on the competition for food in laboratory populations of Australian sheep blowflies, *Lucilia cuprina*. The blowflies were cultured in a room maintained at 25°C. Nicholson measured the population of blowflies every day for approximately one year. Figure 1 displays the blowfly population measured in one such experiment. Nicholson's classic work stimulated wide interest in ecology, and the dynamic system of the blowfly population was investigated intensively thereafter (Hassell 1978; Readshaw and Cuff 1980; Gurney, Blythe, and Nisbet 1980; Kendall et al. 1999).

In the dynamic system of the blowfly population, resource limitation acts with a time delay, roughly equal to the time for a larva to grow up to an adult. Thus May (1976) proposed to model the population dynamics of blowflies with a DDE

$$\frac{dX(t)}{dt} = rX(t)\big[1 - X(t-\tau)/(1000 \cdot K)\big], \tag{1.1}$$

where $X(t)$ is the blowfly population, r is the rate of increase of the blowfly population, K is a resource limitation parameter set by the supply of food, and τ is the time delay, roughly equal to the time for a larva to grow up to an adult. One goal of this article is to estimate the three parameters r, K, and τ from the noisy Nicholson's blowfly data.

In addition to the above DDE model, many other DDE models have been widely used in physiology (Mackey and Glass 1977), ecology (Gurney, Blythe, and Nisbet 1980), and many other areas. A typical DDE model may be expressed as

$$\frac{dX(t)}{dt} = f\big(X(t), X(t-\tau)|\boldsymbol{\beta}\big), \tag{1.2}$$

where $X(t)$ is the dynamic process on $[t_1, t_n]$, τ is a constant delay parameter, and $\boldsymbol{\beta}$ is a p-dimensional vector of parameters describing the system. The above DDE usually has

no analytic solutions and can only be solved numerically. The DDE solutions will not only depend on the parameters $\boldsymbol{\beta}$ and τ, but will also rely on the history of the dynamic process $\mathcal{H}_\tau = \{X(t), t \in [t_1 - \tau, t_1]\}$ which is an infinite-dimensional set.

The form of the DDE model is usually proposed based on the scientific understanding of the dynamic system, in other words, the parametric form of $f(\cdot)$ in (1.2) is often known. However, the parameters $\boldsymbol{\beta}$ and τ in the DDE model are typically unknown. These DDE parameters often have scientific interpretations. Therefore, it is of great interest to estimate the DDE parameters from noisy real measurements or observations on the dynamic system. Moreover, obtaining estimates of DDE parameters can help us to show whether a particular DDE model is able to produce the observed dynamic system by checking the fit of the DDE numeric solution to the observed data. After a suitable DDE model is calibrated with the real data, it will then be reliable to make predictions.

Several methods have been proposed to estimate parameters in DDE models when the data are assumed to be collected with little or no errors. Fowler and Kember (1993) estimated the derivative $X'(t)$ using the finite difference $X(t) - X(t - \delta)$, where δ is a sufficiently small delay. Then, they identified the delay parameter by embedding the dynamic process in the low-dimensional space $(X(t), X(t - \delta), X(t - \Delta))$, where Δ is selected as the location having a sharp change of the volume of the low-dimensional space. Kaplan and Glass (1993) estimated the time lag parameter by studying when trajectories passing through a small region of an embedding space are parallel. Bunner et al. (1996) studied the special DDE model $f(X(t), X(t - \tau)|\boldsymbol{\beta}) = -X(t) + g(X(t - \tau))$. They take advantage of the fact that at the time point t_i when $X'(t_i) = 0$, we have $X(t_i) = g(X(t_i - \tau))$. Next, they infer the time delay using the value that gives the smoothest graph of $X(t_i)$ versus $X(t_i - \tau)$. Ellner et al. (1997) unified the above methods and extended them to more realistic cases, in which the data have measurement errors. They first estimate the derivative $X'(t)$ from the noisy data using nonparametric smoothing methods (Hastie and Tibshirani 1990), and then inferred $f(\cdot)$ in the framework of the generalized additive model. In this article, this approach is called the two-step method. It is not difficult to implement and the computation is efficient. However, the estimates for $X'(t)$ from noisy data with classic nonparametric methods are often inaccurate. As a result, the estimation for the DDE parameters are obtained with large errors. Wood (2001) fitted the DDE models to several real noisy data sets by minimizing the distance of the DDE numeric solutions to the noisy data. This is a very difficult optimization problem since the DDE numeric solutions not only depend on the parameter values, but also on the history of the dynamic process $\mathcal{H}_\tau$, which is essentially an infinite-dimensional optimization problem. This problem is more difficult than estimating parameters in ordinary differential equations (ODEs), which is only a finite-dimensional optimization problem because the ODEs solutions only depend on the parameter values and a finite-dimensional set of initial conditions. Bihorel (2011) developed an R package, “scaRabee", which estimated DDE parameters by maximizing the likelihood function of the observed data. This R package requires users to provide the history of the dynamic process and solve DDEs numerically to calculate the likelihood function. Calderhead, Girolami, and Lawrence (2009) proposed a Bayesian approach to estimate DDE parameters via the novel use of Gaussian processes, which also requires repeatedly solving DDEs numerically when sampling for values of DDE parameters.

We propose the semiparametric method to estimate DDE parameters from noisy data. Our method avoids using DDE numeric solutions, releasing us from specifying the history of the dynamic process $\mathcal{H}_\tau$, which greatly decreases the difficulty of the optimization problem. The optimization procedure is easy to converge and the computation load is greatly decreased. In the semiparametric method, a flexible nonparametric function is estimated to approximate the dynamic process by maximizing the penalized likelihood function. The penalty term is defined with the model, and penalizes the infidelity of the nonparametric function to the DDE model. This DDE-defined penalized likelihood criterion fully uses both the information obtained from the data and the structure of the DDE models. Therefore, this semiparametric method is able to obtain more accurate parameter estimates than the two-step method of Ellner et al. (1997), because the latter ignores information from the structure of the DDE model when estimating $X'(t)$ in the first step.

The remainder of this article is organized as follows. Section 2 introduces the semiparametric method for estimating parameters in the DDE model from noisy data. The semiparametric method is demonstrated in Section 3 by estimating three parameters in DDE (1.1) from real data of the blowfly population. Simulation studies are presented in Section 4 to evaluate the finite sample performance of the semiparametric method, which is compared with some alternative approaches. Conclusions are given in Section 5.

2. METHOD

2.1. THE LIKELIHOOD FUNCTION

The dynamic process $X(t)$ is often measured or observed at multiple time points. Let y_i denote the observation or measurement for the dynamic process at the time point t_i, $i = 1, \ldots, n$. Suppose the dynamic process is modeled with the DDE (1.2). Denote the vector of DDE parameters by $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \tau)^T$, and the DDE solution by $X(t|\boldsymbol{\theta})$. The noisy data y_i are assumed to follow some probability distribution with the log likelihood function, $\log L(y_i|X(t_i|\boldsymbol{\theta}))$. For example, it is often assumed that the noisy data y_i are distributed as $\texttt{Normal}(X(t_i|\boldsymbol{\theta}), \sigma^2)$, then the log likelihood function is equivalent to $(y_i - X(t_i|\boldsymbol{\theta}))^2$ up to a constant. Our objective is to estimate the parameter $\boldsymbol{\theta}$ in the DDE model (1.2) from the noisy data.

The DDE parameter $\boldsymbol{\theta}$ can be estimated by maximizing the log likelihood function

$$H^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log L\big(y_i|X(t_i|\boldsymbol{\theta})\big). \tag{2.1}$$

Unfortunately, the above maximum likelihood method is difficult to directly implement in this context, because the DDE solution $X(t|\boldsymbol{\theta})$ is an implicit function of $\boldsymbol{\theta}$, as the DDE (1.2) usually does not have an analytic solution. The DDE can only be solved with some numeric methods. In addition to the high computation load associated with the numeric solver, the log likelihood function $H(\boldsymbol{\theta})$ typically has many local optima because the DDE solution $X(t|\boldsymbol{\theta})$ is usually not a smooth function of $\boldsymbol{\theta}$. Moreover, the history of the dynamic process $\mathcal{H}_\tau$ has to be specified in order to solve the DDE numerically, but $\mathcal{H}_\tau$ is usually

unknown. Given that $\mathcal{H}_\tau$ is an infinite-dimensional set, it is hard to guess by grid searching methods.

2.2. DDE-DEFINED PENALIZED MAXIMUM LIKELIHOOD

To overcome the above mentioned difficulties associated with directly using numeric DDE solutions, we use a flexible nonparametric function to approximate the DDE solution at any given value of the DDE parameter $\boldsymbol{\theta}$. The nonparametric function is expressed as a linear combination of basis functions,

$$x(t) = \sum_{k=1}^{K} \phi_k(t) c_k = \boldsymbol{\phi}^T(t)\mathbf{c}, \tag{2.2}$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_K(t))^T$ is the vector of basis functions at time t, and $\mathbf{c} = (c_1, \ldots, c_K)^T$ is the corresponding vector of basis coefficients. We choose cubic B-splines as the basis functions, since cubic B-splines are non-zero only in short subintervals, a feature called the *compact support* property, which is of the great importance for efficient computation (de Boor 2001).

The number of basis functions must be large enough to adequately approximate the dynamic process and the derivatives appearing in the ODEs. A rule of thumb is to put one knot at each observation point, so users do not have to select the number of knots and their locations. However, when the data are sparse with just a few observations for the dynamic process, it is recommended to use a larger number of knots than the number of observations. When the dynamic process has sharp features such as peaks, valleys, and high-frequency oscillations, a more dense set of knots in these areas will be required. When the dynamic process or its derivatives exhibits discontinuities, the dynamic process can be approximated by B-splines through putting multiple coincident knots in the discontinuity locations (Ramsay and Silverman 2005).

The vector of basis coefficients, $\mathbf{c}$, is estimated, given the value of $\boldsymbol{\theta}$, by maximizing the penalized log likelihood function

$$J(\mathbf{c}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \log L\big(y_i|x(t_i)\big) - \lambda \int_{t_1+\tau}^{t_n} \left\{ \frac{dx(t)}{dt} - f\big(x(t), x(t-\tau)|\boldsymbol{\beta}\big) \right\}^2 dt, \tag{2.3}$$

where the first term quantifies the fit of the nonparametric function $x(t)$ to the data, and the second term defines the infidelity of the nonparametric function $x(t)$ to the DDE model (1.2). The smoothing parameter λ controls the trade-off between fit to the data and fidelity to the DDE model. Since the nonparametric function $x(t)$ is forced to satisfy the DDE model as well as possible, it must be sufficiently smooth; otherwise, the derivative $x'(t)$ would be too large and far away from $f(x(t), x(t-\tau)|\boldsymbol{\beta})$. In other words, the second term also serves as the roughness penalty on the nonparametric function $x(t)$. In fact, in the special case when $f(x(t), x(t-\tau)|\boldsymbol{\beta}) \equiv 0$, the above penalized log likelihood function $J(\mathbf{c}|\boldsymbol{\theta})$ is exactly the standard criterion used in the penalized nonparametric smoothing method, where the roughness penalty term is defined with the first derivative $x'(t)$.

The integral in (2.3) does not usually have a closed form expression and needs to be evaluated using numerical quadrature. We use the composite Simpson's rule, which provides a

good approximation to the exact integral (Burden and Douglas 2000). For a function $g(t)$, the composite Simpson's rule is

$$\int_{t_1+\tau}^{t_n} g(t)\,dt \approx \frac{h}{3}\left\{g(s_0)+2\sum_{q=1}^{Q/2-1} g(s_{2q})+4\sum_{q=1}^{Q/2} g(s_{2q-1})+g(s_Q)\right\},$$

with the quadrature points $s_q = t_1+\tau+qh$, $q=0,\ldots,Q$, Q is an even integer, and $h=(t_n-t_1-\tau)/Q$. To make the approximation accurate, Q needs to be reasonably large such as $Q=5K$, where K is the number of basis functions used in (2.2) for approximating the DDE solution.

2.3. THE APPROXIMATED LIKELIHOOD FUNCTION

Given any value of $\boldsymbol{\theta}$, we obtain the estimate $\widehat{\mathbf{c}}$ by maximizing the penalized log likelihood function. Therefore, $\widehat{\mathbf{c}}$ may be treated as a function of $\boldsymbol{\theta}$, which is denoted as $\widehat{\mathbf{c}}(\boldsymbol{\theta})$. Then, the estimate for the nonparametric function is also a function of $\boldsymbol{\theta}$:

$$\hat{x}(t|\boldsymbol{\theta}) = \boldsymbol{\phi}^T(t)\widehat{\mathbf{c}}(\boldsymbol{\theta}).$$

The above estimated nonparametric function is an approximation of the DDE solution for any given value of the DDE parameter $\boldsymbol{\theta}$. The log likelihood function (2.1) can be approximated as

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log L\big(y_i|\hat{x}(t_i|\boldsymbol{\theta})\big). \tag{2.4}$$

The DDE parameter $\boldsymbol{\theta}$ is estimated by maximizing the approximated log likelihood function. The approximated log likelihood function $H(\boldsymbol{\theta})$ is much easier to maximize than the exact log likelihood function $H^*(\boldsymbol{\theta})$ because it does not contain the DDE numeric solution and hence does not depend on the history of the dynamic process. Moreover, the gradient for the optimization can be obtained analytically, as shown in the next subsection. Hence, the optimization converges more readily with the supplied analytic gradient.

2.4. OPTIMIZATION METHODS

The spline coefficient $\mathbf{c}$ and the DDE parameter $\boldsymbol{\theta}$ are estimated in two nested levels of optimization. In the inner level of optimization, the spline coefficient $\mathbf{c}$ is estimated by maximizing the DDE-defined penalized likelihood function $J(\mathbf{c}|\boldsymbol{\theta})$ for any given value of $\boldsymbol{\theta}$. In the outer level of optimization, the DDE parameter $\boldsymbol{\theta}$ is estimated by maximizing the approximated log likelihood function $H(\boldsymbol{\theta})$. Below we introduce the optimization methods used in each level of optimization. First, we will focus on the special case that the data follow a normal distribution, and then extend this framework to the general case in which the data follow any distributions.

When the noisy data y_i follow the normal distribution with the mean equal to the DDE solution, then the approximated log likelihood function is

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log L\big(y_i|\hat{x}(t_i|\boldsymbol{\theta})\big) \propto \sum_{i=1}^{n} \big(y_i-\hat{x}(t_i|\boldsymbol{\theta})\big)^2.$$

The above criterion is expressed in the form of the sum of squared function values; hence, the Gauss–Newton algorithm can be used to find the estimate for $\boldsymbol{\theta}$ by minimizing the sum of squared errors. As a modification of Newton's algorithm, the Gauss–Newton algorithm has the advantage that the Hessian matrix, which is challenging to compute, is not required, though this method can only be used to minimize functions in the form of the sum of squared function values. As pointed out by one reviewer of this manuscript, the above approximated log likelihood function may be multimodal, so it is suggested to try different initial guesses to check whether the optimization procedure converges to the same point.

Starting with the initial guess $\boldsymbol{\theta}^{(0)}$ for the DDE parameter $\boldsymbol{\theta}$, the Gauss–Newton algorithm proceeds by the iterations

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - \left\{ \left(\frac{d\mathbf{r}}{d\boldsymbol{\theta}} \right)^T \left(\frac{d\mathbf{r}}{d\boldsymbol{\theta}} \right) \right\}^{-1} \left(\frac{d\mathbf{r}}{d\boldsymbol{\theta}} \right)^T \mathbf{r},$$

where $\mathbf{r}$ is a vector of length n with the ith entry $r_i = y_i - \hat{x}(t_i|\boldsymbol{\theta}^{(j)})$, and

$$\frac{d\mathbf{r}}{d\boldsymbol{\theta}} = -\boldsymbol{\Phi} \frac{d\widehat{\mathbf{c}}^T}{d\boldsymbol{\theta}}.$$

Here $\boldsymbol{\Phi}$ is the $n \times K$ basis matrix with the ikth entry $\phi_k(t_i)$. The function $\widehat{\mathbf{c}}(\boldsymbol{\theta})$ is an implicit function and cannot be expressed in an analytic form, since one boundary point of the integral in (2.3) is defined by the delay parameter τ. However, the derivative $d\widehat{\mathbf{c}}/d\boldsymbol{\theta}$ can be obtained analytically using the implicit function theorem as follows.

We take advantage of the fact that the estimate $\widehat{\mathbf{c}}$ satisfies

$$\left. \frac{\partial J}{\partial \mathbf{c}^T} \right|_{\widehat{\mathbf{c}}} \equiv 0.$$

Taking the $\boldsymbol{\theta}$-derivative on both sides of the above identity, we obtain

$$\frac{d}{d\boldsymbol{\theta}} \left. \frac{\partial J}{\partial \mathbf{c}^T} \right|_{\widehat{\mathbf{c}}} = \left\{ \left. \frac{\partial^2 J}{\partial \mathbf{c}^T \partial \boldsymbol{\theta}} \right|_{\widehat{\mathbf{c}}} \right\} + \left\{ \left. \frac{\partial^2 J}{\partial \mathbf{c}^T \partial \mathbf{c}} \right|_{\widehat{\mathbf{c}}} \right\} \left\{ \frac{d\widehat{\mathbf{c}}^T}{\partial \boldsymbol{\theta}} \right\} \equiv 0,$$

which yields

$$\frac{d\widehat{\mathbf{c}}^T}{\partial \boldsymbol{\theta}} = \left\{ \left. \frac{\partial^2 J}{\partial \mathbf{c}^T \partial \mathbf{c}} \right|_{\widehat{\mathbf{c}}} \right\}^{-1} \left\{ \left. \frac{\partial^2 J}{\partial \mathbf{c}^T \partial \boldsymbol{\theta}} \right|_{\widehat{\mathbf{c}}} \right\},$$

provided that $\partial^2 J / \partial \mathbf{c}^T \partial \mathbf{c}$ is non-singular at $\mathbf{c} = \widehat{\mathbf{c}}$.

The criterion $J(\mathbf{c})$ in the inner level of optimization can also be expressed in the form of the sum of squared function values by using numerical quadrature methods to calculate the integral. Therefore, it can also be optimized by the Gauss–Newton algorithm, which only requires the analytic gradient.

One advantage of approximating the DDE solution with the nonparametric function is that the log likelihood function $H(\boldsymbol{\theta})$ is a smoother function of $\boldsymbol{\theta}$ than using the DDE solution directly. As a result, the optimization procedure is less sensitive to the initial values and is easier to converge. Nevertheless, $H(\boldsymbol{\theta})$ is not always convex so the local optimum may not be the global optimum. The Gauss–Newton algorithm is only able to find a local optimum. Consequently, it is advised to try different initial values when implementing the optimization procedure.

When the data follow any other distributions, the criterion $H(\boldsymbol{\theta})$ cannot usually be expressed in the form of a sum of squared function values. Therefore, the optimization cannot be handled with the Gauss–Newton algorithm. Instead, we can use the Newton algorithm. The Newton algorithm starts with the initial guess $\boldsymbol{\theta}^{(0)}$ for the DDE parameter $\boldsymbol{\theta}$, and proceeds by the iterations

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - \left(\frac{d^2H}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}^T}\right)^{-1}\left(\frac{dH}{d\boldsymbol{\theta}}\right),$$

where the gradient $dH/d\boldsymbol{\theta}$ and the Hessian matrix $d^2H/d\boldsymbol{\theta}\, d\boldsymbol{\theta}^T$ can be derived analytically using the implicit function theorem similar to the one above.

2.5. SMOOTHING PARAMETER SELECTION

The smoothing parameter λ in (2.3) controls the trade-off between fitting the data and maintaining fidelity to the DDE model. The semiparametric method is appealing for its flexibility, which allows users to try different values of the smoothing parameter λ based on their confidence in their data or the DDE model. Nevertheless, we propose an objective criterion, K-fold cross validation, to choose the optimal value for the smoothing parameter.

K-fold cross validation is defined as follows. The data are partitioned into K folds. In each of the K experiments, we use $K-1$ folds for training and the remaining one for testing. Let $\boldsymbol{\theta}^{(-j)}$ denote the parameter estimates from the $K-1$ folds of data after removing the jth fold, and $\mathcal{A}^{(-j)}$ denote the set of indices such that all y_i, $i \in \mathcal{A}^{(-j)}$, do not belong to the jth fold of data. The K-fold cross validation is defined as

$$CV = \frac{1}{K}\sum_{j=1}^{K}\sum_{i\in\mathcal{A}^{(-j)}} \log L\left(y_i|\hat{x}\left(t_i|\boldsymbol{\theta}^{(-j)}\right)\right).$$

All simulations and applications in this paper choose $K = 10$, which is a common practice in the statistics literature.

2.6. STANDARD ERROR ESTIMATION

Standard errors of the DDE parameter estimates can be estimated using the parametric bootstrap as follows. Denote $\widehat{\boldsymbol{\theta}}$ as the estimate for the DDE parameters. The DDE is solved numerically with $\widehat{\boldsymbol{\theta}}$ as the value of the parameter, and the DDE solution is denoted as $X(t|\widehat{\boldsymbol{\theta}})$. The DDE solution, $X(t|\widehat{\boldsymbol{\theta}})$, is plugged into the assumed probability distribution of the noisy data, and the simulated data are drawn from the likelihood. Then, the semiparametric method is used to obtain the estimate for the DDE parameter, which is denoted by $\widetilde{\boldsymbol{\theta}}^{(s)}$, $s = 1, \ldots, S$, where s is the index for the simulation, and S is the total number of simulations. The standard error of $\widehat{\boldsymbol{\theta}}$ is calculated as the experimental standard deviation of $\widetilde{\boldsymbol{\theta}}^{(s)}$, $s = 1, \ldots, S$. The standard error for the DDE solution, denoted as $\mathtt{SE}(X(t|\widehat{\boldsymbol{\theta}}))$, is calculated as the experimental standard deviation of $X(t|\widetilde{\boldsymbol{\theta}}^{(s)})$. The 95% confidence interval for the DDE solution is given as $[X(t|\widehat{\boldsymbol{\theta}}) - 1.96 * \mathtt{SE}(X(t|\widehat{\boldsymbol{\theta}})), X(t|\widehat{\boldsymbol{\theta}}) + 1.96 * \mathtt{SE}(X(t|\widehat{\boldsymbol{\theta}}))]$, which is based on the assumption of asymptotic normality.

3. APPLICATION

As introduced in Section 1, the population dynamics of blowflies can be modeled with the delay differential equation (1.1). The semiparametric method is used to estimate the three DDE parameters, r, K and τ, from the real data shown in Figure 1.

Since the real data y_i, $i = 1, \ldots, n$, are counts of blowflies, they are assumed in the log-normal distribution, $\ln \texttt{Normal}(X(t_i|\boldsymbol{\theta}), \sigma^2)$, where $X(t|\boldsymbol{\theta})$ is the solution of the DDE model (1.1) with the parameter $\boldsymbol{\theta} = (r, K, \tau)^T$. The semiparametric method approximates the DDE solution with cubic B-splines with 201 equally spaced knots on the whole time interval. The estimates for the three parameters are displayed in Table 1. The DDE (1.1) is solved numerically with the parameter estimates, and the DDE solution fits the real data well. This suggests that the time-delayed regulatory mechanism is indeed the basic feature of the dynamics of the blowfly population.

Table 1 also shows that the standard errors of the parameter estimates are obtained using the parametric bootstrap method. The parametric bootstrap method is implemented as follows. The simulated data, y_i^*, are generated as $\ln y_i^* = \ln X(t_i|\widehat{\boldsymbol{\theta}}) + \epsilon_i^*$, where $\widehat{\boldsymbol{\theta}}$ is the estimate for the DDE parameter from the real data, $\epsilon_i^* \sim \texttt{Normal}(0, \hat{\sigma}^2)$, and $\hat{\sigma}^2 = 1.87$ is the estimate for σ^2 from the real data. Then, the DDE parameters are estimated from the simulated data, which are denoted as $\widetilde{\boldsymbol{\theta}}$. The above procedure is implemented with 1000 replicates. The experimental standard deviation of $\widetilde{\boldsymbol{\theta}}$ is the parametric bootstrap estimate for the standard error of the parameter estimates.

Figure 2 displays the 95% two-dimensional confidence regions for the DDE parameters using the parametric bootstrap method. They are constructed as $\{\mathbf{z} : (\mathbf{z} - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z} - \widehat{\boldsymbol{\mu}}) \leq \chi^2(2; 0.95)\}$, where $\mathbf{z}$ is a 2-dimensional vector, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ is the experimental mean and covariance matrix of the pair-wise DDE parameter estimates using the parametric bootstrap method, and $\chi^2(2; 0.95)$ is the 95% quantile of the Chi-square distribution with two degrees of freedom. Notice that the upper left panel in Figure 2 shows the negative correlation between r and τ. Recall that r is the rate of increase of the blowfly population, and τ is the time delay, roughly equal to the time for a larva to grow up to an adult. After knowing the meanings of r and τ, we may understand the negative correlation between r and τ: if the blowfly grows faster, it will take a larva shorter lime to grow up to an adult, and the rate of increase of the blowfly population will also be higher.

Table 1. Estimation for the three DDE parameters from the real blowfly data.

Parameters	τ	K	r
Point Estimates	8.781	2.226	0.222
Standard Errors	0.039	0.085	0.003

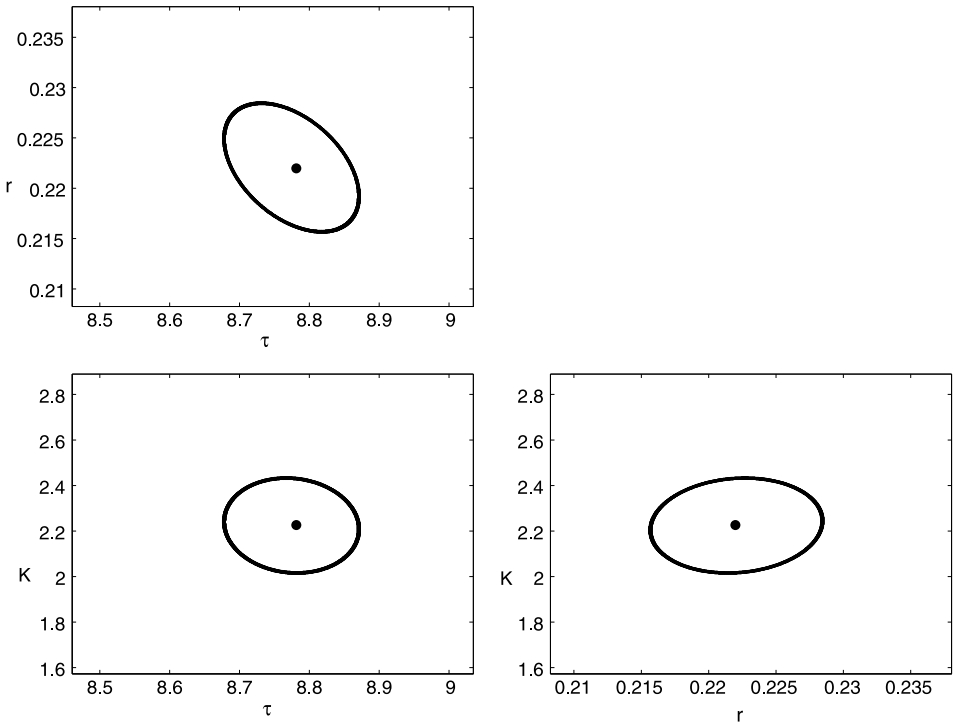

Figure 2. The 95% two-dimensional confidence regions for the DDE parameters using the parametric bootstrap method. The dots are the point estimates of the DDE parameters from the real blowfly data.

4. SIMULATIONS

Some simulations are implemented to evaluate the finite sample performance of the semiparametric method. The semiparametric method is also compared with some alternative approaches in terms of estimation precision.

4.1. HUTCHINSON'S EQUATION

Hutchinson (1948) considered a single species obeying a logistic growth law, and modeled its population dynamics with a DDE model:

$$\frac{dX(t)}{dt} = rX(t)\big[1 - X(t-\tau)/(1000 \cdot K)\big],$$

where r is the growth rate of the species, $1000 \cdot K$ is the maximum population sustainable by the environmental resources, and τ is the time lag built into the operation of this regulatory mechanism. The properties of the above DDE have been studied extensively in the mathematical literature. For example, the population of the species remains at a stable equilibrium point equal to $1000K$ when $r\tau < 0.5\pi$. When $r\tau > 0.5\pi$, the population dynamics exist as a limit cycle. The three parameters τ, r, and K are estimated from the simulated noisy data with the semiparametric method, the two-step method,

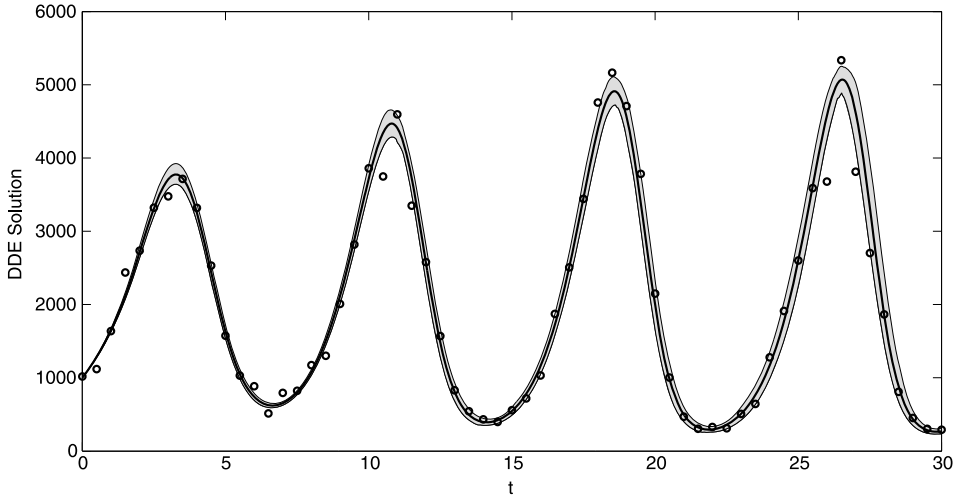

Figure 3. The numeric solution of the DDE $dX(t)/dt = rX(t)[1 - X(t-\tau)/(1000 \cdot K)]$ using the estimated parameter values. The circles are the simulated noisy data by adding noises to the DDE solutions at 61 equally spaced time points in $[0, 30]$. The gray band is the 95% confidence interval for the DDE solution.

and the scaRabee package (Bihorel 2011). The scaRabee package requires users to provide the history of the dynamic process, and we set the history of the dynamic process $\mathcal{H}_\tau = \{X(t) = y_1, t \in [-\tau, 0]\}$, where y_1 is the first data point at $t = 0$.

The simulated noisy data are generated as follows. First, the DDE (1.1) is first solved numerically with the true parameter values in n equally spaced time points $[0, 30]$. The true values of parameters are set to $\tau = 3\pi/5$, $r = 1.0$, and $K = 2$, and the initial history is set to $X(t) = 1000$ when $t \le 0$. The numeric solution of the DDE is shown in Figure 3, which displays the dynamics of a limit cycle. Then, the simulated noisy data are generated using the log-normal distribution, $\ln \texttt{Normal}(X(t_i), \sigma^2)$, $i = 1, \ldots, n$, where $X(t)$ is the numeric solution of the DDE model using the true parameter values. The standard deviation of the noise is set to $\sigma = 0.088$, which is 10% of the standard deviation of the logarithm of the DDE solution. The sample size n is set to 61, 121, 241 in three scenarios, and the simulation is implemented with 100 replicates in each scenario. Figure 3 shows one simulated data set when the sample size is $n = 61$.

The parameter estimates are summarized in Table 2. The parameter estimates with the semiparametric method have much smaller biases, standard deviations (STDs), and root mean squared errors (RMSEs) than the two-step method under any sample size. When the sample size is small ($n = 61$), the RMSEs for $\hat{\tau}$, $\hat{K}$, and $\hat{r}$ using the semiparametric method are 11%, 41%, 14% of those using the two-step method, respectively. When the sample size increases, the gains using the semiparametric method are even larger. For example, when the sample size is $n = 241$, the RMSEs for $\hat{\tau}$, $\hat{K}$, and $\hat{r}$ using the semiparametric method are only 5%, 25%, 5% of those using the two-step method, respectively. Moreover, the biases of the parameter estimates using the semiparametric method are also negligible compared with their standard deviations. On the other hand, the parameter estimates obtained using

Table 2. The biases, standard deviations (STDs), and root mean squared errors (RMSEs) of the parameter estimates for the Hutchinson's DDE model using the semiparametric method, the scaRabee package in R, and the two-step method. The sample size n is set as 61, 121, 241 in three scenarios, and the simulation is implemented with 100 replicates in each scenario. "SEMI" represents the semiparametric method.

Sample Size		$n=61$			$n=121$			$n=241$		
DDE Parameter		τ	K	r	τ	K	r	τ	K	r
True Parameter Value		1.9	2.0	1.0	1.9	2.0	1.0	1.9	2.0	1.0
BIAS$*10^2$	SEMI	0.1	−0.2	−0.1	0.2	−0.3	−0.1	0.1	−5e-2	−6e-4
	scaRabee	0.2	0.3	−0.1	0.3	0.4	−0.1	0.1	0.2	4e-2
	2-Step	−0.2	0.7	−4.0	2.3	−1.7	−1.7	4.5	−4.0	−2.9
STD$*10^2$	SEMI	0.7	2.7	0.8	0.4	1.9	0.3	0.2	1.3	0.2
	scaRabee	1.3	2.9	0.9	1.2	2.7	0.7	1.1	2.2	0.3
	2-Step	6.3	6.5	4.2	5.6	5.8	3.6	3.2	3.6	2.8
RMSE$*10^2$	SEMI	0.7	2.7	0.8	0.4	1.9	0.3	0.3	1.3	0.2
	scaRabee	1.3	2.9	0.9	1.2	2.7	0.7	1.1	2.2	0.3
	2-Step	6.3	6.6	5.7	6.0	6.0	4.0	5.5	5.4	4.0

Table 3. The coverage probabilities of the confidence intervals for $\hat{\tau}$, $\hat{K}$, $\hat{r}$, and the two-dimensional confidence regions for $(\hat{\tau}, \hat{K})$, $(\hat{K}, \hat{r})$, $(\hat{\tau}, \hat{r})$ constructed using the parametric bootstrap method. The sample size n is set as 61, 121, 241 in three scenarios.

	$\hat{\tau}$	$\hat{K}$	$\hat{r}$	$(\hat{\tau}, \hat{K})$	$(\hat{K}, \hat{r})$	$(\hat{\tau}, \hat{r})$
$n=61$	96%	97%	97%	97%	96%	95%
$n=121$	96%	96%	96%	93%	95%	95%
$n=241$	95%	97%	94%	95%	93%	94%

the two-step method have large biases, which are caused by the poor estimates for the derivative $X'(t)$ in the first step.

The scaRabee package obtains more accurate parameter estimates than the two-step method; however, it is not as good as the semiparametric method. In particular, the semiparametric method has much smaller RMSEs for the estimates of the delay parameter, $\hat{\tau}$, than the scaRabee package. One reason for the unsatisfactory performance of the scaRabee package is that it requires providing the history of the dynamic process, but the history of the dynamic process is hard to estimate beforehand.

The 95% confidence interval for $\hat{\tau}$ is constructed as $[\hat{\tau} - 1.96 * \mathtt{SE}(\hat{\tau}), \hat{\tau} + 1.96 * \mathtt{SE}(\hat{\tau})]$, where $\mathtt{SE}(\hat{\tau})$ is the standard error of $\hat{\tau}$ estimated using the parametric bootstrap method as introduced in Subsection 2.6. The 95% confidence intervals for $\hat{K}$ and $\hat{r}$ are constructed in the same way. The 95% two-dimensional confidence region for $(\hat{\tau}, \hat{K})$ is constructed as $\{\mathbf{z} : (\mathbf{z} - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z} - \widehat{\boldsymbol{\mu}}) \leq \chi^2(2; 0.95)\}$, where $\mathbf{z}$ is a 2-dimensional vector, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ is the estimated mean and covariance matrix of $(\hat{\tau}, \hat{K})$ using the parametric bootstrap method, and $\chi^2(2; 0.95)$ is the 95% quantile of the Chi-square distribution with two degrees of freedom. The 95% two-dimensional confidence regions for $(\hat{K}, \hat{r})$ and $(\hat{\tau}, \hat{r})$ are constructed in the same way. Table 3 displays the coverage probabilities of the confidence intervals for

$\hat{\tau}$, $\hat{K}$, $\hat{r}$, and the two-dimensional confidence intervals for $(\hat{\tau}, \hat{K})$, $(\hat{K}, \hat{r})$, $(\hat{\tau}, \hat{r})$ when the sample size $n = 61, 121, 241$. These coverage probabilities are all close to 95%.

4.2. MACKEY–GLASS EQUATION

Mackey and Glass (1977) proposed a delay differential equation to model the nonlinear feedback control in physiology:

$$\frac{dX(t)}{dt} = \frac{aX(t-\tau)}{1 + X(t-\tau)^c} - bX(t). \tag{4.1}$$

The above equation displays several interesting dynamic behaviors including limit cycle oscillations and aperiodic solutions. They showed that the Mackey–Glass equation could reproduce the qualitative features of dynamical respiratory and hematopoietic diseases. The semiparametric method is used to estimate the four parameters a, b, c, and τ from the simulated noisy data, and is compared with the two-step method in terms of estimation accuracy.

The noisy data are simulated using the normal distribution $\texttt{Normal}(X(t_i), \sigma^2)$, where $X(t)$ is the numeric solution of the DDE model (4.1). The true values of parameters are set to $\tau = 2$, $a = 2$, $b = 1$, and $c = 10$, and the initial history is set to $X(t) = 0.5$ when $t \le 0$. The standard deviation of the noise is set to $\sigma = 0.025$, which is 10% of the standard deviation of the DDE solution. The sample size n is chosen to be 20, 50, 100 in three scenarios, and the simulation is implemented with 100 replicates in each scenario. Figure 4 displays one simulated data set, along with the numeric solution of the Mackey–Glass equation.

Table 4 compares the parameter estimates using the semiparametric method and the two-step method. The semiparametric method obtains much smaller biases and root mean

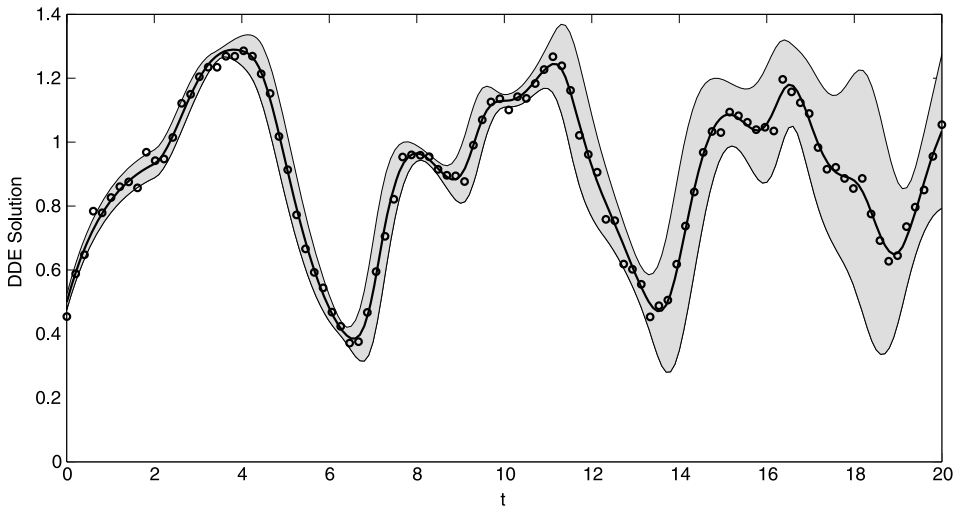

Figure 4. One noisy data set simulated based on the Mackey–Glass equation $dX(t)/dt = \{aX(t-\tau)\}/\{1 + X(t-\tau)^c\} - bX(t)$. The solid line is the numeric solution of the Mackey–Glass equation using the estimated parameter values. The gray band is the 95% confidence interval for the DDE solution.

Table 4. The biases, standard deviations (STDs), and root mean squared errors (RMSEs) of the parameter estimates for the Mackey–Glass equation using the semiparametric method and the two-step method. The sample size n is set as 20, 50, 100 in three scenarios, and the simulation is implemented with 100 replicates in each scenario. "SEMI" represents the semiparametric method.

			BIAS∗10		STD∗10		RMSE∗10	
n	Parameters	True	SEMI	2-Step	SEMI	2-Step	SEMI	2-Step
20	a	2.00	−0.12	−5.85	1.23	1.07	1.24	5.95
	b	1.00	−0.07	−3.19	0.62	0.58	0.63	3.24
	c	10.00	3.04	12.89	4.90	7.70	5.76	15.01
	τ	2.00	−0.01	−1.59	0.18	0.41	0.18	1.64
50	a	2.00	−0.03	−2.51	0.46	0.86	0.46	2.65
	b	1.00	−0.05	−1.29	0.25	0.43	0.26	1.36
	c	10.00	1.62	3.43	2.41	5.02	2.90	6.08
	τ	2.00	0.01	−0.31	0.08	0.19	0.08	0.36
100	a	2.00	−0.02	−1.50	0.28	0.65	0.28	1.63
	b	1.00	−0.03	−0.77	0.16	0.34	0.16	0.84
	c	10.00	0.87	2.08	1.49	3.63	1.73	4.19
	τ	2.00	0.02	−0.18	0.05	0.14	0.06	0.23

squared errors (RMSEs) of the parameter estimates for any sample size. The standard deviations (STDs) of the parameter estimates using the semiparametric method are mostly smaller than those obtained using the two-step method except those of a and b with sample size $n = 20$. The RMSEs of the semiparametric estimates for the delay parameter τ are only 11%, 22%, and 25% of those obtained using the two-step method when the sample size is $n = 20, 50$, and 100, respectively. The gains of using the semiparametric method are also substantial for the estimation of the other three DDE parameters. For instance, the semiparametric method has the RMSEs of the estimates for a reduced 79%, 83%, and 83% over the two-step method when the sample size is $n = 20, 50$, and 100, respectively.

5. CONCLUSION

Delay differential equations (DDEs) usually do not have analytic solutions and can only be solved by some numeric methods. The numeric solutions of DDEs not only depend on the DDE parameters, but also on the history of the dynamic system. Therefore, it is difficult to apply the traditional likelihood method to estimate DDE parameters.

We propose the semiparametric method to address this problem. In order to avoid solving DDEs numerically, a flexible nonparametric function is used to approximate the dynamic process. The nonparametric function is estimated by maximizing the penalized likelihood function. The penalty term is defined using the DDE model, which forces the nonparametric function to satisfy the DDE model and penalizes the roughness of the nonparametric function. Simulation studies show that the semiparametric method obtains more accurate estimates for the DDE parameters than some alternative approaches.

As pointed out by one reviewer, the penalized log likelihood function (2.3) implies a hierarchical model where the prior distribution for the dynamic process is defined with the

DDE model. From this point of view, our method is similar to the Bayesian hierarchical implementations (Wikle and Hooten 2010). Lele, Dennis, and Lutscher (2007) introduced a data cloning method to calculate maximum likelihood estimates and their standard errors for complex statistical models in the Bayesian framework by implementing the Markov chain Monte Carlo (MCMC) algorithm. The data cloning method is promising in estimating the DDE models, and will be investigated in our future research.

Our method can be applied to estimate DDE parameters from data in any distribution by choosing the appropriate distribution and likelihood function for data. For example, when data contain zeros, (Cangelosi and Hooten 2009) suggested using the truncated normal distribution for describing the data.

ACKNOWLEDGEMENTS

The authors greatly appreciate all the very constructive suggestions and comments from Dr. Montse Fuentes, the Associate Editor, and two referees, who were very helpful for improving the quality of this work and manuscript. This research was supported by a discovery grant from the Natural Science and Engineering Research Council of Canada to J. Cao. Part of the research was done when the second author visited the Statistical and Applied Mathematical Sciences Institute (SAMSI) in the winter term of 2011. He would like to thank SAMSI for the kind hosting.

[Received March 2011. Accepted September 2011.]

REFERENCES

Bihorel, S. (2011), *The scaRabee Package: An R-based Tool for Model Simulation and Optimization in Pharmacometrics*.

Bunner, M. J., Popp, M., Meyer, Th., Kittel, A., Ran, U., and Parisi, J. (1996), "Recovery of Scalar Time-Delay Systems From Time Series," *Physics Letters. A*, 211, 345–349.

Burden, R. L., and Douglas, F. J. (2000), *Numerical Analysis* (7th ed.), Pacific Grove, CA: Brooks/Cole.

Calderhead, B., Girolami, M., and Lawrence, N. D. (2009), "Accelerating Bayesian Inference Over Nonlinear Differential Equations With Gaussian Processes," *Advances in Neural Information Processing Systems*, 21, 217–224.

Cangelosi, A. R., and Hooten, M. B. (2009), "Models for Bounded Systems With Continuous Dynamics," *Biometrics*, 65, 850–856.

de Boor, C. (2001), *A Practical Guide to Splines*, New York: Springer.

Ellner, S. P., Kendall, B. E., Wood, S. N., McCauley, E., and Briggs, C. J. (1997), "Inferring Mechanism From Time-Series Data: Delay-Differential Equations," *Physica D*, 110, 182–194.

Fowler, A. C., and Kember, G. (1993), "Delay Recognition in Chaotic Time Series," *Physics Letters. A*, 175, 402–408.

Gurney, W. S. C., Blythe, S. P., and Nisbet, R. M. (1980), "Nicholson's Blowflies Revisited," *Nature*, 287, 17–21.

Hassell, M. P. (1978), *The Dynamics of Arthropod Predator-Prey Systems*, New York: Princeton University Press.

Hastie, T., and Tibshirani, R. (1990), *Generalized Linear Models*, London: Chapman and Hall.

Hutchinson, G. E. (1948), "Circular Causal Systems in Ecology," *Annals of the New York Academy of Sciences*, 50, 221–246.

Kaplan, D. T., and Glass, L. (1993), "Coarse-Grained Embeddings of Time Series: Random Walks, Gaussian Random Process, and Deterministic Chaos," *Physica D*, 64, 431–454.

Kendall, B. E., Briggs, C. J., Murdoch, W. W., Turchin, P., Ellner, S. P., McCauley, E., Nisbet, R. M., and Wood, S. N. (1999), "Why Do Populations Cycle? A Synthesis of Statistical and Mechanistic Modeling Approaches," *Ecology*, 80, 1789–1805.

Lele, S. R., Dennis, B., and Lutscher, F. (2007), "Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models Using Bayesian Markov Chain Monte Carlo Methods," *Ecology Letters*, 10, 551–563.

Mackey, M. C., and Glass, L. (1977), "Oscillations and Chaos in Physiological Control Systems," *Science*, 197, 287–289.

May, R. M. (1976), "Models for Single Populations," in *Theoretical Ecology: Principles and Applications*, ed. R. M. May, Oxford: Blackwell Scientific Publications, pp. 4–25.

Nicholson, A. J. (1954a), "Compensatory Reactions of Populations to Stresses, and Their Evolutionary Significance," *Australian Journal of Zoology*, 2, 1–8.

——— (1954b), "An Outline of the Dynamics of Animal Populations," *Australian Journal of Zoology*, 2, 9–65.

——— (1957), "The Self-adjustment of Populations to Change," *Cold Spring Harbour Symposia on Quantitative Biology*, 22, 153–172.

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer.

Readshaw, J. L., and Cuff, W. R. (1980), "A Model of Nicholson's Blowfly Cycles and Its Relevance to Predation Theory," *Journal of Animal Ecology*, 49 (3), 1005–1010.

Wikle, C. K., and Hooten, M. B. (2010), "A General Science-Based Framework for Dynamical Spatio-Temporal Models," *Test*, 19, 417–451.

Wood, S. N. (2001), "Partially Specified Ecological Models," *Ecological Monographs*, 71, 1–25.