BIOMETRICS 67, 1305–1313
December 2011
DOI: 10.1111/j.1541-0420.2011.01577.x

Robust Estimation for Ordinary Differential Equation Models

J. Cao,[1,*] L. Wang,[2] and J. Xu[1]
[1]Department of Statistics & Actuarial Science, Simon Fraser University, Burnaby, British Columbia V5A1S6, Canada
[2]Department of Statistics, University of British Columbia, Vancouver, British Columbia V6T1Z2, Canada
*email: jiguo_cao@sfu.ca

SUMMARY. Applied scientists often like to use ordinary differential equations (ODEs) to model complex dynamic processes that arise in biology, engineering, medicine, and many other areas. It is interesting but challenging to estimate ODE parameters from noisy data, especially when the data have some outliers. We propose a robust method to address this problem. The dynamic process is represented with a nonparametric function, which is a linear combination of basis functions. The nonparametric function is estimated by a robust penalized smoothing method. The penalty term is defined with the parametric ODE model, which controls the roughness of the nonparametric function and maintains the fidelity of the nonparametric function to the ODE model. The basis coefficients and ODE parameters are estimated in two nested levels of optimization. The coefficient estimates are treated as an implicit function of ODE parameters, which enables one to derive the analytic gradients for optimization using the implicit function theorem. Simulation studies show that the robust method gives satisfactory estimates for the ODE parameters from noisy data with outliers. The robust method is demonstrated by estimating a predator–prey ODE model from real ecological data.

KEY WORDS: Dynamic model; Predator–prey system; Robust penalized smoothing; System identification.

1. Introduction

Ordinary differential equations (ODEs) are popular models for describing complex dynamic systems in biology, engineering, physics, and many other areas of applied science and technology. For instance, ODEs display some dynamic behaviors similar to those observed in ecological populations in the field and laboratory, such as coexistence at an equilibrium, a limit cycle, or a chaotic attractor (e.g., Becks et al., 2005). Therefore, ODEs are widely used for understanding and predicting the dynamics of interacting populations (Rosenzweig and MacArthur, 1963; Murdoch, Briggs, and Nisbet, 2003). The Lotka–Volterra model, which refers to two coupled ODEs, is the pioneering and simplest possible predator–prey dynamic model. It has been modified in many ways since its original formulation in the 1920s. Some of the modifications include inducible defenses in the prey (Vos et al., 2004) and adaptive foraging by the predator (Kondoh, 2003).

Although an ODE model can be proposed by applied scientists based on their expert knowledge about the dynamic process, the parameters in the ODE models are often unknown. Estimation of ODE parameters from real data is of great interest, but is a difficult problem, because most ODEs do not have analytic solutions, and numerically solving ODEs is computationally expensive.

A host of methods have been proposed to estimate ODE parameters from noisy data. Bard (1974) and Biegler, Damiano, and Blau (1986) introduced the nonlinear least squares method, which searches for the optimal values of ODE parameters by optimizing the fitting of the numerical solutions of the ODEs to the data. However, the numeric solution is very sensitive to the parameter values, resulting in the optimization surface having many local optimizers. Hence the nonlinear least squares method sometimes does not converge. A referee points out that the roughness of the optimization surface due to the numeric ODE solutions may cause difficulties when finite difference approximation is used to calculate gradients for nonlinear least squares. This numeric problem can be solved if gradients are calculated by solving the variational equations, making the optimization much more stable. On the other hand, the roughness of the optimization surface due to the ODE model structure itself, e.g., from oscillatory behavior, is a more difficult problem for the standard nonlinear least squares method. A possible remedy in this case is the multiple shooting approach (Bock, 1983). Gelman, Bois, and Jiang (1996) solved this optimization problem by using a Bayesian approach. Both methods repeatedly solve ODEs numerically over thousands of parameter values, and the computation is very intensive.

A two-step approach was suggested by Varah (1982) and was later developed by Ramsay and Silverman (2005), Poyton (2005), Chen and Wu (2008), and Brunel (2008). The first derivative of the dynamic process is estimated in the first step, and is treated as a response variable in the second step. The ODE parameters are then estimated with a standard nonlinear least squares method in the second step. The bottleneck for the two-step approach is the difficulty in estimating the first derivative accurately.

Ramsay et al. (2007) develop the generalized profiling method, which uses a nonparametric function to represent the dynamic process with the penalized spline smoothing method. This method has been shown to obtain good estimates of the ODE parameters with low computational load. Qi and Zhao

(2010) establish the asymptotic efficiency of the generalized profiling method.

All of the above methods do not take into account that the data may have outliers, which is often a serious problem in practice. We propose a robust method for estimating parameters in ODE models from real data. The dynamic process is represented as a nonparametric function, which is a linear combination of some basis functions. The nonparametric function is estimated with the robust (M-type) smoothing method. We define some robust measurement of the fitted residuals to ensure that the estimated nonparametric function is robust to the outliers in the data.

Robust smoothing has been studied by Huber (1979), Cox (1983), Härdle and Gasser (1984), Silverman (1985), and Hall and Jones (1990). We extend their work by defining a penalty term incorporating the ODEs. Defining a penalty term with the ODEs has two advantages. First, the penalty term contains derivatives coming from the ODEs, which can penalize the roughness of the nonparametric function. This allows us to use a saturated number of basis functions to construct the nonparametric function, and the nonparametric function will not overfit the data. Second, the penalty term penalizes the infidelity of the nonparametric function to the ODEs, which forces the nonparametric function to represent the dynamic process that is modeled with the ODEs.

The basis coefficients and ODE parameters are estimated in two nested levels of optimization. In the lower level optimization, the robust penalized smoothing method estimates the basis coefficients, conditional on the ODE parameters. So the estimator for the basis coefficients can be treated as an implicit function of the ODE parameters. In the upper level optimization, the ODE parameters are then estimated by minimizing the summation of the robust norm of fitting errors of the nonparametric function. The optimization is accelerated by providing the analytic gradients, which are derived using the implicit function theorem. The variances of the ODE parameters are estimated by the sandwich method (Huber, 1967; White, 1982; Kauermann and Carroll, 2001).

The robust method for estimating ODE parameters is motivated by a predator–prey dynamic system described in Fussmann et al. (2000). An aquatic laboratory community containing two microbial species whose dynamic behavior is studied by Fussmann et al. (2000), Shertzer et al. (2002), and Yoshida et al. (2003). The system is a nutrient-based predator–prey food chain, in which unicellular green algae, *Chlorella vulgaris*, are eaten by planktonic rotifers, *Brachionus calyciflorus*. The growth of *Chlorella* is also limited by the supply of nitrogen. *Chlorella* and *Brachionus* are grown together in replicated, experimental flow-through cultures, called chemostats. Nitrogen continuously flows into the system with concentration N^* at the dilution rate δ, and all variables are removed from the chemostats at the same rate δ. Fussmann et al. (2000) mathematically model the system using a set of nonlinear ODEs, coupled by consumer–resource interactions between the planktonic rotifers, green algae, and the nitrogen resource:

$$\begin{aligned}
\frac{dN}{dt} &= \delta(N^* - N) - F_C(N)C\,,\\
\frac{dC}{dt} &= F_C(N)C - F_B(C)B/\epsilon - \delta C\,,\\
\frac{dR}{dt} &= F_B(C)R - (\delta + m + \alpha)R\,,\\
\frac{dB}{dt} &= F_B(C)R - (\delta + m)B\,,
\end{aligned} \tag{1}$$

where N, C, R, B are the concentrations of nitrogen, *Chlorella*, reproducing *Brachionus*, and total *Brachionus*, respectively; $F_C(N) = b_C N/(k_C + N)$, $F_B(C) = b_B C/(k_B + C)$ are two functional responses (with b_C and b_B the maximum birth rates of *Chlorella* and *Brachionus*; k_C and k_B the half-saturation constants of *Chlorella* and *Brachionus*); and ϵ, α, and m are the assimilation efficiency, the decay of fecundity, and the mortality of *Brachionus*, respectively.

The above dynamic model correctly predicts three qualitative types of dynamic behavior of the experimental system: the predator and prey coexist at an equilibrium at low nutrient supply (small δ or small N^*); the system switches to a limit cycle when nutrient supply increases (increasing δ or N^*); very high nutrient supply leads to extreme oscillations that cause the extinction of the predator or both the predator and the prey. However, Fussmann et al. (2000) point out that their model performs poorly at predicting quantitative aspects of the experimental predator–prey dynamics because of the lack of knowledge of the parameter values. Cao, Fussmann, and Ramsay (2008) improve the fitting of the ODE solution to the real data by estimating the ODE parameters using the generalized profiling method.

Figure 1 displays the ODE solutions using the generalized profiling estimates and the parameter values given in Fussmann et al. (2000). The generalized profiling method clearly makes the ODE solutions fit the data better, which is a good validation for the ODE model (1). However, one data point (marked with a circle in Figure 1) is too high for the cyclic trend of the concentration of *Brachionus*, and may be an outlier, but the generalized profiling method does not consider this outlier problem. Our robust method should further improve the fitting of the ODE model by downweighting the impact of outliers, which will be shown in Section 3.

Our article is organized as follows. The robust method for estimating ODE models from noisy data is introduced in Section 2. Simulation studies are presented in Section 3 to evaluate the finite-sample performance of the robust method. Sections 4 demonstrates the robust method by estimating a predator–prey ODE model from the real ecological data. Conclusions and discussions are given in Section 5.

2. The Robust Method

In this section, we introduce the robust method for estimating ODE models from the noisy data. For simplicity of notation, we first assume that the ODE model is composed of one single variable, and later extend the notation to include ODE models with multiple variables. Let $X(t)$ be a dynamic process modeled with one ODE

$$\frac{dX(t)}{dt} = h(X(t)\,|\,\boldsymbol{\theta})\,, \tag{2}$$

where $h(X(t)\,|\,\boldsymbol{\theta})$ has a known parametric form. The parameter vector $\boldsymbol{\theta}$ is unknown and to be estimated with the robust

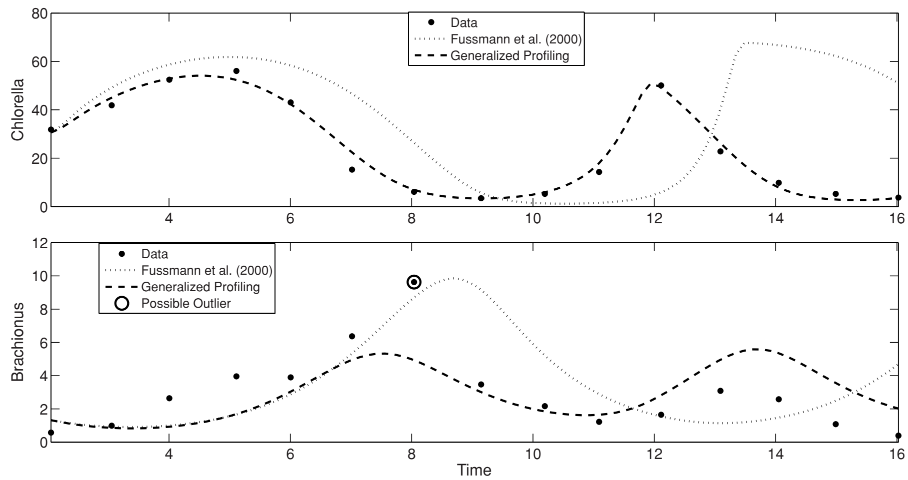

Figure 1. The numeric solutions of the predator–prey ODE (1) using the generalized profiling parameter estimates or the parameter values given in Fussmann et al. (2000). Observed experimental data are from Yoshida et al. (2003; Figure 2), with dilution rates $\delta = 0.68\,\text{day}^{-1}$. The unit of *Chlorella* and *Brachionus* is $\mu mol\,\text{L}^{-1}$, and the unit of time is day.

method from noisy observations $y(t_i)$, $i = 1, \dots, n$, where we assume $y(t_i)$ to have mean $X(t_i)$ and variance σ^2.

The ODE parameter $\boldsymbol{\theta}$ is estimated in two nested levels of optimization. In the lower level optimization, a nonparametric function $x(t)$ is estimated to represent the dynamic process by robust penalized smoothing using an ODE-defined penalty, conditional on $\boldsymbol{\theta}$. Therefore, the estimate $\widehat{x}(t)$ can be treated as a function of $\boldsymbol{\theta}$. In the upper level optimization, $\boldsymbol{\theta}$ is estimated by minimizing the summation of the robust norm of errors, which is a function of $\widehat{x}(t)$, and thus is also a function of $\boldsymbol{\theta}$.

2.1 *Robust Penalized Smoothing with the ODE-defined Penalty*

The dynamic process is represented with a nonparametric function $x(t)$, which is constructed as a linear combination of K basis functions $\phi_k(t)$, $k = 1, \dots, K$,

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

The basis system must have the capacity to represent the dynamic process, as well as derivatives involved in the ODEs. Many ODE solutions have sharp features, such as peaks, valleys, high-frequency oscillations, and discontinuities in derivatives. Therefore we choose the B-spline basis system that can accommodate the discontinuities by assigning multiple knots to the critical locations (Ramsay and Silverman, 2005). Moreover, B-spline basis functions are only positive over a short subinterval and zero elsewhere. This is called the *compact support* property, and is essential for efficient computation using sparse programming techniques. A saturated number of basis functions are used here to make sure that the nonparametric function $x(t)$ can well represent the dynamic process. A rule of thumb is to use the cubic B-spline basis functions with $4n$ equally spaced knots. More knots may be required when the dynamic process has some sharp features. However, this may cause the over-fitting problem, because the number of basis coefficients is larger than the number of observations to fit.

To avoid over-fitting, a penalty term is defined to penalize the roughness of the nonparametric function. For instance, in order to obtain a smooth nonparametric function, the penalty term can be defined in terms of the second derivative of the function, that is,

$$\texttt{PEN}(x) = \int_{t_1}^{t_n} \left\{ \frac{d^2 x(t)}{dt^2} \right\}^2 dt\,.$$

This is called *penalized spline smoothing* (Eilers and Marx, 1996; Ruppert, Wand, and Carroll, 2003). In our case, the nonparametric function must also satisfy the ODE model (2), so it is natural to define the penalty term with the ODE model

$$\texttt{PEN}(x \mid \boldsymbol{\theta}) = \int_{t_1}^{t_n} \left\{ \frac{dx(t)}{dt} - h(x(t) \mid \boldsymbol{\theta}) \right\}^2 dt\,.$$

The vector of spline coefficients, $\mathbf{c}$, is estimated by minimizing the negative penalized log-likelihood function, conditional on the ODE parameter $\boldsymbol{\theta}$,

$$-\frac{1}{n} \sum_{i=1}^{n} \log f(y(t_i) - x(t_i)) + \lambda \int_{t_1}^{t_n} \left\{ \frac{dx(t)}{dt} - h(x(t) \mid \boldsymbol{\theta}) \right\}^2 dt,$$

where the smoothing parameter λ controls the trade-off between fit to the data and fidelity to the ODE model, and $f(\cdot)$ is the density function for the residuals $r_i = y(t_i) - x(t_i)$. If $f(\cdot)$ is the normal density function, the above criterion is equivalent to the penalized sum of squared errors.

In this article, we define some robust measurement of residuals $f(r) = \exp(-\rho(r))$, where $\rho(r)$ is typically convex and symmetric about zero, quadratic in the neighborhood of zero, and increasing at a rate slower than r^2 for large r. So $\rho(r)$ downweights extreme residuals in comparison with squared residuals. One common choice of ρ is the family of Huber functions

$$\rho(r) = \begin{cases} r^2 & \text{if } |r| \leqslant \kappa\,, \\ 2\kappa|r| - \kappa^2 & \text{if } |r| > \kappa\,, \end{cases}$$

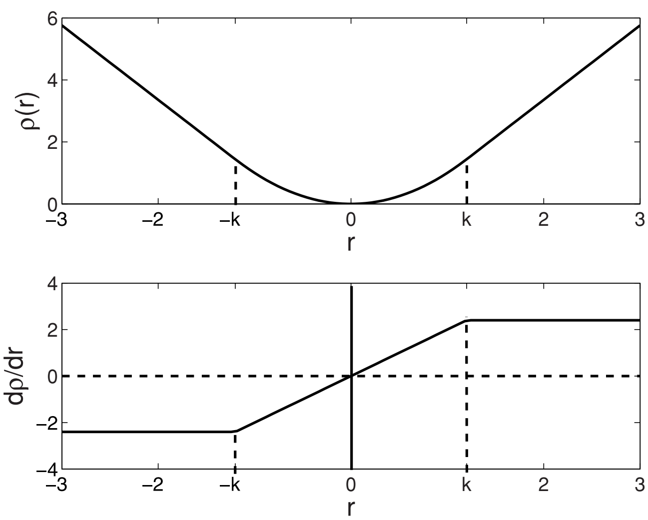

Figure 2. The Huber function $\rho(r)$ and the first derivative $d\rho/dr$.

where κ is the cutoff of the Huber functions. Figure 2 displays one example of the Huber function and its first derivative. Users can also choose other robust functions for $\rho(r)$, such as the bisquare family of functions.

The robust penalized smoothing method can be generalized to a set of ODEs with multiple variables. Denote the set of ODEs as $dX_\ell(t)/dt = f_\ell(\mathbf{X}(t)\,|\,\boldsymbol{\theta})$, $\ell = 1,\ldots,S$, where $\mathbf{X}(t) = (X_1(t),\ldots,X_S(t))^T$, and the parametric form of $f_\ell(\cdot)$ is known. Suppose only M variables are observed, where $M \leqslant S$. Denote $y_j(t_{ij})$ as the observation for the jth variable at t_{ij}, $i = 1,\ldots,n_j$, $j = 1,\ldots,M$. The ℓth variable is represented as a linear combination of basis functions, $x_\ell(t) = \mathbf{c}_\ell^T \boldsymbol{\phi}_\ell(t)$, $\ell = 1,\ldots,S$.

The vector of basis coefficients $\mathbf{c} = (\mathbf{c}_1^T,\ldots,\mathbf{c}_S^T)^T$ is estimated by minimizing

$$J(\mathbf{c}\,|\,\boldsymbol{\theta}) = \sum_{j=1}^{M}\omega_j \sum_{i=1}^{n_j}\rho_j(y_j(t_{ij}) - x_j(t_{ij})) + \sum_{\ell=1}^{S}\lambda_\ell\,\omega_\ell \int_{t_1}^{t_n}\left\{\frac{dx_\ell(t)}{dt} - f_\ell(\mathbf{x}(t)\,|\,\boldsymbol{\theta})\right\}^2 dt, \quad (3)$$

where $\mathbf{x}(t) = (x_1(t),\ldots,x_S(t))^T$ is a vector of nonparametric functions for the S variables. The robust measurement of residuals, $\rho_j(\cdot)$, is chosen as the Huber function with the cutoff parameter κ_j for each variable. The weight, ω_j, is the normalizing weight, which is required in order to maintain comparable scales for different variables. A referee points out that the optimal weights are the inverse of the variance of measurement noises in the case of squared loss. Better estimates may be obtained with the iterated reweighted robust estimation in which the weights are updated in each iteration as the inverse of the estimated variance of the measurement noises.

Some ODE variables may not be observed, i.e., $M < S$. For instance, in our predator–prey ODEs (1), only the concentrations of *Chlorella* (C) and total *Brachionus* (B) are observed, so we have $S = 4$ and $M = 2$. When some variables are not observed, the nonparametric functions representing these missing variables can still be estimated by minimizing (3), because these nonparametric functions are contained in the penalty term. The implementation for optimizing $J(\mathbf{c}\,|\,\boldsymbol{\theta})$ is discussed in Web Appendix A.

2.2 *Estimating ODE Parameters*

The ODE parameters $\boldsymbol{\theta}$ are estimated in the top level of optimization by minimizing the summation of the robust norm of errors for the fitting of the nonparametric functions

$$H(\boldsymbol{\theta}) = \sum_{j=1}^{M}\omega_j \sum_{i=1}^{n}\rho_j(y_j(t_{ij}) - \boldsymbol{\phi}_j(t_{ij})\widehat{\mathbf{c}}_j(\boldsymbol{\theta})),$$

where $\widehat{\mathbf{c}}_j(\boldsymbol{\theta})$ is the conditional estimate of the spline coefficients for the jth variable by minimizing (3), which is treated as a function of $\boldsymbol{\theta}$. This function $\widehat{\mathbf{c}}_j(\boldsymbol{\theta})$ cannot be expressed explicitly, so a lower level of optimization for estimating $\mathbf{c}_j$ conditional on $\boldsymbol{\theta}$, described in Subsection 2.1, is embedded inside the objective function $H(\boldsymbol{\theta})$.

The optimization is faster and more stable if we can provide the analytic gradient, which is expressed using the chain rule

$$\frac{dH}{d\boldsymbol{\theta}} = \left(\frac{d\,\widehat{\mathbf{c}}}{d\boldsymbol{\theta}}\right)^T \frac{dH}{d\,\widehat{\mathbf{c}}}$$

where $\widehat{\mathbf{c}} = (\widehat{\mathbf{c}}_1^T,\ldots,\widehat{\mathbf{c}}_S^T)^T$ is the vector of spline coefficient estimators. The above gradient requires the analytic expression for the first-order derivative of $\widehat{\mathbf{c}}$ with respect to $\boldsymbol{\theta}$, but $\widehat{\mathbf{c}}$ cannot be expressed explicitly as a function of $\boldsymbol{\theta}$. Fortunately, the implicit function theorem can be applied to find the analytic expression for the first-order derivative of $\widehat{\mathbf{c}}$ with respect to $\boldsymbol{\theta}$, which is introduced as follows. Taking the $\boldsymbol{\theta}$-derivative on both sides of the identity $\partial J/\partial\mathbf{c}\,|_{\widehat{\mathbf{c}}} = 0$

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

Assuming that $\partial^2 J/\partial\mathbf{c}^2|_{\widehat{\mathbf{c}}}$ is nonsingular, the analytic expression for the first-order derivative of $\widehat{\mathbf{c}}$ with respect to $\boldsymbol{\theta}$ is obtained

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

2.3 *Variances of the ODE Parameters*

The variances for the ODE parameters are estimated with the sandwich method. The sandwich method provides a consistent estimate for the covariance matrix of $\widehat{\boldsymbol{\theta}}$, without the need to make any distribution assumptions (Carroll et al., 2006). In the top level of optimization, the estimating equation for $\boldsymbol{\theta}$ is

$$\sum_{j=1}^{M}\omega_j \sum_{i=1}^{n_j}\Psi_{ij}(\boldsymbol{\theta}) = 0,$$

where

$$\Psi_{ij}(\boldsymbol{\theta}) = \left(\frac{d\widehat{\mathbf{c}}_j}{d\boldsymbol{\theta}}\right)^T \boldsymbol{\phi}_j(t_{ij})\rho'\{y_j(t_{ij}) - \boldsymbol{\phi}_j^T(t_{ij})\widehat{\mathbf{c}}_j(\boldsymbol{\theta})\}.$$

The sandwich method estimates the covariance matrix of $\widehat{\boldsymbol{\theta}}$ as

$$\widehat{\mathtt{Cov}}(\widehat{\boldsymbol{\theta}}) = \mathbf{A}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{B}(\widehat{\boldsymbol{\theta}})\{\mathbf{A}^{-1}(\widehat{\boldsymbol{\theta}})\}^T ,$$

where the two matrices $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$ are

$$\mathbf{A}(\boldsymbol{\theta}) = \sum_{j=1}^{M} \omega_j \sum_{i=1}^{n_j} \frac{d}{d\boldsymbol{\theta}} \Psi_{ij}(\boldsymbol{\theta}) ,$$

$$\mathbf{B}(\boldsymbol{\theta}) = \sum_{j=1}^{M} \omega_j \sum_{i=1}^{n_j} \Psi_{ij}(\boldsymbol{\theta})\Psi_{ij}(\boldsymbol{\theta})^T ,$$

where the analytic derivative for $d\Psi_{ij}/d\boldsymbol{\theta}$ is given in Web Appendix B. This requires the second-order derivative of $\widehat{\mathbf{c}}$ with respect to $\boldsymbol{\theta}$, which is also derived using the implicit function theorem.

2.4 *Identifiability of ODE Parameters*

Denote $\boldsymbol{\theta}^*$ as the true value of the ODE parameter $\boldsymbol{\theta}$, and $\mathbf{X}^*(t)$ as the solution of ODEs when the parameter value is $\boldsymbol{\theta}^*$. Brunel (2008) proposed a global identifiability criterion for the ODE parameter $\boldsymbol{\theta}$. Under L^2 norm on the space of integrable functions on $[t_1, t_n]$, the identifiability criterion is

$$\forall \nu > 0, \inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}^*|| \geqslant \nu} R(\boldsymbol{\theta}) > R(\boldsymbol{\theta}^*) ,$$

where

$$R(\boldsymbol{\theta}) = \sum_{\ell=1}^{S} \sqrt{\int_{t_1}^{t_n} [f_\ell(\mathbf{X}^*(t) \,|\, \boldsymbol{\theta}^*) - f_\ell(\mathbf{X}^*(t) \,|\, \boldsymbol{\theta})]^2 \, dt} .$$

In practice, Brunel (2008) suggested checking the identifiability of ODE parameters by considering the singularity of the Hessian H^* of $R(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$

$$H^* = \sum_{\ell=1}^{S} \int_{t_1}^{t_n} \left[\frac{d^2 f_\ell(\mathbf{X}^*(t) \,|\, \boldsymbol{\theta}^*)}{d\boldsymbol{\theta} d\boldsymbol{\theta}^T} \right]^2 dt . \qquad (4)$$

If the Hessian H^* is nonsingular, $R(\boldsymbol{\theta})$ behaves like a positive-definite quadratic form on the neighborhood of $\boldsymbol{\theta}^*$, and the ODE parameter $\boldsymbol{\theta}$ is identifiable.

2.5 *Smoothing Parameter Selection*

Our ultimate goal is to obtain an estimate for the ODE parameters $\boldsymbol{\theta}$ such that the solution of the ODEs is close to the observed data. For each value of the smoothing parameter $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_S)^T$, we obtain the ODE parameter estimate $\widehat{\boldsymbol{\theta}}$, so $\widehat{\boldsymbol{\theta}}$ may be treated as a function of $\boldsymbol{\lambda}$. The optimal value of $\boldsymbol{\lambda}$ is chosen by minimizing

$$F(\boldsymbol{\lambda}) = \sum_{j=1}^{M} \omega_j \sum_{i=1}^{n_j} \rho_j(y_j(t_{ij}) - s_j(t_{ij} \,|\, \widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}))) , \qquad (5)$$

where $s_j(t_{ij} \,|\, \widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}))$ is the ODE solution at the point t_{ij} with the parameter $\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda})$ for the jth variable . The criterion (5) can be understood as some measurement of the prediction errors: given the initial values of ODE variables and ODE parameter values, the ODE numeric solution is used as the prediction for all the data points. The criterion (5) chooses the optimal value of λ such that the ODE solution with the parameter estimates is closest to the data.

2.6 *Estimating Initial Values of ODE Variables*

The goodness-of-fit of ODE models to noisy data can be assessed by solving ODEs numerically, and comparing the fit of ODE solutions to data. Solving ODEs requires one to specify the initial values for the ODE variables, which are defined as the values of the ODE variables at the first time point. A tiny change to the initial values may result in a huge difference of the ODE solutions. Therefore, it is very important to use an accurate estimate for the initial values. The first observations for the ODE variables at the first time point often have measurement errors, and thus it is dangerous to use the first observations as the initial values of the ODE variables. Moreover, some ODE variables may not be measurable, and no first observations are available.

A useful byproduct of the robust method is that the initial values of the ODE variables can be estimated after obtaining the estimates for the ODE parameters. The robust method uses a nonparametric function to represent the dynamic process, hence the initial values of the ODE variables can be estimated by evaluating the nonparametric function at the first time point: $\hat{x}_\ell(t_0) = \widehat{\mathbf{c}}_\ell^T \boldsymbol{\phi}(t_0)$, $\ell = 1, \ldots, S$. Our experience indicates that the ODE solution with the estimated initial values tends to fit the data better than using the first observations directly.

3. Simulations

Two simulation studies are carried out to evaluate the finite-sample performance of the robust method, one is a linear ODE and the other is a nonlinear ODE. The robust method is compared with the generalized profiling method in these two simulations. The nonlinear ODE is chosen as the FitzHugh-Nagumo ODE, which is a popular model for describing the behavior of spike potentials in the giant axon of squid neurons (FitzHugh, 1961; Nagumo, Arimoto, and Yoshizawa, 1962):

$$\begin{aligned} \frac{dV(t)}{dt} &= c\left(V(t) - \frac{V(t)^3}{3} + R(t)\right) , \\ \frac{dR(t)}{dt} &= -\frac{1}{c}\left(V(t) - a + bR(t)\right) , \end{aligned} \qquad (6)$$

where $V(t)$ is the voltage across an axon membrane and $R(t)$ is the outward currents. The parameter a, b, c are to be estimated from the simulated noisy data.

The simulated data are generated as follows. The FitzHugh-Nagumo ODEs are solved numerically at 201 equally spaced points on [0,20] with the initial values $V(0) = -1$ and $R(0) = 1$ and the true ODE parameters $a = 0.2$, $b = 0.2$, and $c = 3$. The noisy data are generated by adding noise from $\mathtt{Normal}(0, 1)$ to the ODE solutions. We randomly select m observations using the discrete Uniform distribution on [1,201]. These selected m observations are replaced with randomly generated outliers that are added or subtracted from the original observations with equal probability. Each outlier is generated from the Pareto distribution, Pareto($\upsilon = 3$, $\xi = 3$) .

The ODE parameters, a, b, and c, are estimated from the simulated data using the robust method and the generalized profiling method. Both methods represent the ODE variables, $V(t)$ and $R(t)$, with cubic B-splines using 201 equally spaced knots on [0,20]. The effect of outliers is studied by varying the number of outliers $m = 0, 20, 40, 60$. The ODE

Table 1

The biases, standard deviations (SDs), and root mean squared errors (RMSEs) of parameter estimates on 100 simulation replicates using the robust method, the generalized profiling (GP) method, and the two-step method for the nonlinear ODE (6). The tuning parameter in the Huber function, $\tau = 1.35$, which correspond to 95% asymptotic efficiency at the normal distribution.

			No outliers			20 Outliers		
	True	Method	Bias	SD	RMSE	Bias	SD	RMSE
a	0.2	Robust	−0.009	0.045	0.046	−0.012	0.051	0.052
		GP	−0.010	0.045	0.046	−0.009	0.090	0.090
		2-Step	0.019	0.092	0.094	0.024	0.130	0.131
b	0.2	Robust	-4×10^{-4}	0.185	0.184	−0.010	0.198	0.197
		GP	4×10^{-4}	0.175	0.175	−0.043	0.282	0.284
		2-Step	−0.908	0.199	0.929	−0.884	0.231	0.914
c	3	Robust	0.015	0.187	0.186	0.003	0.213	0.212
		GP	0.017	0.197	0.196	0.065	0.407	0.410
		2-Step	−2.855	0.040	2.856	−2.862	0.046	2.863
			40 Outliers			60 Outliers		
	True	Method	Bias	SD	RMSE	Bias	SD	RMSE
a	0.2	Robust	−0.012	0.069	0.069	−0.017	0.083	0.085
		GP	−0.094	0.524	0.529	−0.119	0.669	0.676
		2-Step	0.014	0.157	0.157	0.012	0.195	0.194
b	0.2	Robust	−0.019	0.271	0.271	−0.034	0.276	0.277
		GP	−0.188	1.069	1.080	−0.159	1.324	1.327
		2-Step	−0.830	0.259	0.869	−0.842	0.387	0.926
c	3	Robust	−0.012	0.265	0.264	0.005	0.263	0.262
		GP	0.644	3.145	3.195	0.887	2.937	3.054
		2-Step	−2.866	0.063	2.867	−2.867	0.067	2.868

parameters are also estimated using the robust two-step method, which is implemented as follows. In the first step, the estimated nonparametric functions, $\hat{V}(t)$ and $\hat{R}(t)$, and their derivatives, $d\hat{V}(t)/dt$ and $d\hat{R}(t)/dt$, are obtained by the robust local polynomial regression (Loader, 1999); in the second step, the three ODE parameters are estimated with the standard linear regression by treating the estimated $d\hat{V}(t)/dt$ and $d\hat{R}(t)/dt$ as the response variables, and treating $\hat{V}(t)$ and $\hat{R}(t)$ as covariates.

Table 1 displays the biases, standard deviations (SDs), and root mean squared errors (RMSEs) of the parameter estimates on 100 simulation replicates using the three methods. When the simulated data have no outliers, the robust method has almost the same biases, SDs and RMSEs as the generalized profiling method for all three parameters. When 20 outliers are present in the simulated data, the generalized profiling method has around double the SDs and RMSEs of $\hat{a}$ and $\hat{c}$ compared to the scenario with no outliers, while the robust method has only a slightly increase. The RMSEs of the estimates for a, b, and c using the robust method are around 58%, 69%, and 52% of those from the generalized profiling method, respectively. When 40 or 60 outliers are present in the simulated data, the RMSEs of the estimates for a, b, and c using the robust method are only around 13%, 25%, and 8% of those using the generalized profiling method. In contrast with the robust method and the generalized profiling method, the robust two-step method produces seriously biased estimates for the parameter b and c. This is caused by the poor estimates, $\hat{V}(t)$ and $\hat{R}(t)$, and their derivatives, $d\hat{V}(t)/dt$ and $d\hat{R}(t)/dt$, in the first step using robust local polynomial regression, which does not incorporate the structure of the ODE model. The second simulation study on the linear ODE is discussed in Web Appendix C, and it also shows the robust method obtains more accurate estimates than the generalized profiling method.

4. Application

The predator–prey ODE model (1) is estimated from real ecological data using the robust method. The ODE parameters to estimate are $\boldsymbol{\theta} = (\epsilon, \alpha, m, b_C, b_B, k_C, k_B)^T$. The biological interpretations are given in Section 1. Four variables, N, C, R, B, in the predator–prey ODE model are each represented with a cubic B-spline with 400 equally spaced knots. We only have the data for *Chlorella* (C) and *Brachionus* (B), and the other two variables, nitrogen (N) and reproducing *Brachionus* (R), are not measurable. Hence, the criterion (3) is used to obtain the estimate of the spline coefficients, given any value of the ODE parameters $\boldsymbol{\theta}$. We have the measurements of the two variables, so $M = 2$ in (3). The predator–prey ODE model (1) has four variables, so $S = 4$ in (3). The weights ω_j, $j = 1, 2, 3, 4$ in (3) are chosen as the reciprocals of variances of the predator–prey ODE solutions using the parameter values given in Fussmann et al. (2000), which are 0.0011, 0.0011, 0.16, 0.094, respectively, such that the normalized sums of squared errors are of roughly comparable sizes. The optimal smoothing parameter is chosen as $\lambda_\ell = 10^4$, $\ell = 1, 2, 3, 4$ by minimizing the criterion (5). The cutoff κ_j for the Huber function ρ_j in (3) is taken as $\kappa_j = \tau_j \hat{\sigma}_j$, where $\hat{\sigma}_j$ is a robust

Table 2

Parameter estimates (PEs) and the standard errors (SEs) for the predator–prey ODE model (1) *from the real ecological data. MSE is defined as the mean squared errors of the ODE solutions to the data excluding outliers. As a comparison, we also give the parameter values given in Fussmann et al.* (2000) *and the generalized profiling estimates.*

Estimates		ϵ	α	m	b_C	b_B	k_C	k_B	MSE
Fussmann	PE	0.25	0.40	0.055	3.3	2.25	4.3	15.0	1.762
Profiling	PE	0.11	0.01	0.152	3.9	1.97	4.3	15.7	0.171
	SE	0.02	0.14	0.073	0.47	0.26	1.95	2.01	
Robust	PE	0.09	7.1×10^{-5}	0.072	3.5	1.74	6.6	17.5	0.122
	SE	0.01	0.08	0.088	0.2	0.07	0.6	0.9	

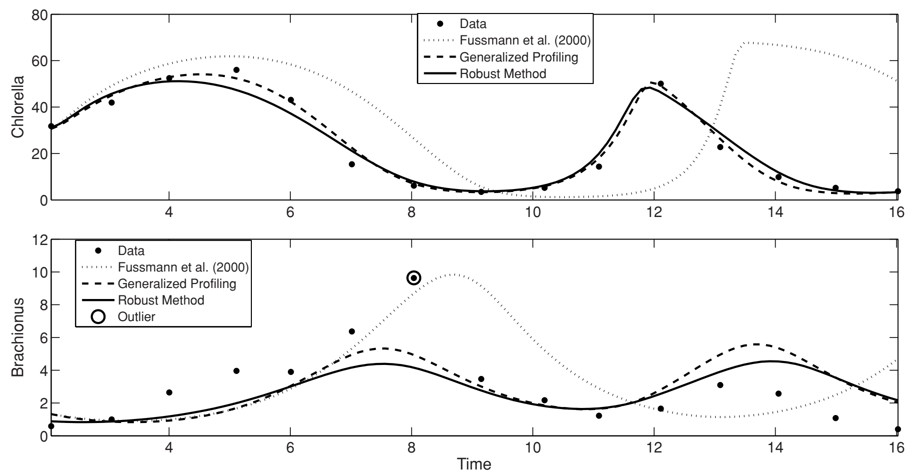

Figure 3. Solutions of the predator–prey ODEs (1) using the parameter values as robust estimates, generalized profiling estimates, and those in Fussmann et al. (2000). Observed experimental data are from Yoshida et al. (2003; Figure 2), with dilution rates $\delta = 0.68\,\text{day}^{-1}$. The circle indicates the outlier identified by robust method. The unit of *Chlorella* and *Brachionus* is $\mu mol\,\text{L}^{-1}$, and the unit of time is day.

estimate of the noise standard deviation, and τ_j is a positive constant usually chosen as $\tau_j = 1.345$ (Huber, 1981).

The parameter estimates from the observed data are shown in Table 2. The robust method gives a smaller assimilation efficiency (ϵ) and decay of fecundity (α) but larger half-saturation constants of *Chlorella* and *Brachionus* (k_C and k_B) when compared with the generalized profiling method and the parameter values given in Fussmann et al. (2000). The Hessian matrix defined in (4) is nonsingular under the robust ODE parameter estimates, indicating that the ODE parameters are identifiable. The robust estimates for the standard deviations are $\hat{\sigma}_C = 1.73$ and $\hat{\sigma}_B = 2.10$. The standard errors for robust estimates are estimated using the sandwich method. We notice that some parameter values are well defined by the data, as indicated by their small standard errors, and others are poorly defined, such as parameters α and m. This suggests that more observations are required in order to estimate these parameters accurately.

The predator–prey ODEs (1) are solved numerically using parameter values equal to robust estimates, generalized profiling estimates, and those given in Fussmann et al. (2000), respectively. The ODE solutions are shown in Figure 3. The two peaks of the ODE solution for *Brachionus* using the robust estimates are lower than those using the generalized profiling estimates, because the Huber function in the robust method downweights the effect of the outlier marked with a circle. We define the outliers as those observations satisfying $y_j(t_{ij}) > s_j(t_{ij}\,|\,\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda})) + 1.96\hat{\sigma}_j$ or $y_j(t_{ij}) < s_j(t_{ij}\,|\,\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda})) - 1.96\hat{\sigma}_j$. The mean squared errors (MSE) of the ODE solution to the observations excluding outliers are calculated to quantify the goodness of fit of the ODE models with the parameter estimates. The MSE with robust estimates is reduced 93% from that with parameter values given in Fussmann et al. (2000). The robust method also has 29% smaller MSE than the generalized profiling method.

5. Conclusions and Discussion

ODEs are popular tools for modeling dynamic process in many areas. However, the values of ODE parameters are rarely known. While it is of great interest to estimate ODE models from noisy observations, there are some limitations with current statistical approaches for estimating such models. For instance, the current estimation methods do not take into account outliers in observations, and hence the estimators are not robust.

We propose a robust method for estimating ODE models from noisy observations. A nonparametric function is used to represent the dynamic process. The nonparametric function is estimated by the robust penalized spline smoothing method. Some robust measurements for fitted residuals are defined, so the estimate for the nonparametric function is robust to any outliers in the data. The parametric ODE models define the penalty term, which controls the roughness of the nonparametric function and maintains the fidelity of the nonparametric function to the ODE models.

The spline coefficients and the ODE parameters are estimated by two nested levels of optimization. The spline coefficients are estimated in the lower level optimization, conditional on the ODE parameters, hence the coefficient estimates can be treated as an implicit function of the ODE parameters. The ODE parameters are then estimated in the upper level optimization. The sandwich method is applied to estimate the variance of the ODE parameters. The functional relationships between the spline coefficients and the ODE parameters are considered, and are used to derive the analytic gradients for optimization with the implicit function theorem.

It is hard to distinguish outliers and model misspecification when no measurement repetitions are available. In our application, our biology collaborators would like to assume their ODE model is correct, and obtain the robust estimate for the ODE parameters when data are suspected to have outliers. Our robust estimation method is useful in this application. However, if more than one model is available to choose from, the first step we recommend is to do model selection or model averaging (Claeskens and Hjort, 2008). It is very important to be sure that the estimated parameters are statistically meaningful and reliable. When some ODE variables cannot be measured or observed, some ODE parameters may not be identifiable (Walter, 1987). A referee also points out that a weighted least squares approach should be preferable if weights for each data point are available based on information about the measurement errors.

The simulation studies show that the robust method provides satisfactory estimates for the ODE parameters when data have outliers. The robust method is applied to estimate a predator–prey ODE model from the real ecological data. The ODE model with the robust estimates fits the data better than the generalized profiling estimates.

6. Supplementary Materials

Web Appendices A, B, C referenced in Sections 2.1, 2.3, and 3 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

This research was supported by a discovery grant from the Natural Science and Engineering Research Council of Canada. The authors would like to thank Dr Geert Verbeke, the associate editor, and three referees for their very constructive suggestions and comments, which greatly improved this manuscript.

References

Bard, Y. (1974). *Nonlinear Parameter Estimation*. New York: Academic Press.

Becks, L., Hilker, F. M., Malchow, H., Jurgens, K., and Arndt, H. (2005). Experimental demonstration of chaos in a microbial food web. *Nature* **435,** 1226.

Biegler, L., Damiano, J. J., and Blau, G. E. (1986). Nonlinear parameter estimation: A case study comparison. *AIChE Journal* **32,** 29–45.

Bock, H. G. (1983). Recent advances in parameter identification techniques for ODE. In *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, P. Deuflhard and E. Harrier (eds), 95–121. Basel: Birkhäuser.

Brunel (2008). Parameter estimation of ODE's via nonparametric estimators. *Electronic Journal of Statistics* **2,** 1242–1267.

Cao, J., Fussmann, G., and Ramsay, J. O. (2008). Estimating a predator-prey dynamical model with the parameter cascades method. *Biometrics* **64,** 959–967.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*, 2nd edition. New York: Chapman & Hall/CRC.

Chen, J. and Wu, H. (2008). Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. *Journal of the American Statistical Association* **103,** 369–383.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge, UK: Cambridge University Press.

Cox, D. D. (1983). Asymptotics for M-type smoothing splines. *Annals of Statistics* **11,** 530–551.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties (with discussion). *Statistical Science* **89,** 89–121.

FitzHugh, R. (1961). Impulses and physiological states in models of nerve membrane. *Biophysical Journal* **1,** 445–466.

Fussmann, G. F., Ellner, S. P., Shertzer, K. W., and Hairston, N. G. J. (2000). Crossing the hopf bifurcation in a live predator-prey system. *Science* **290,** 1358–1360.

Gelman, A., Bois, F., and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* **91,** 1400–1412.

Hall, P. and Jones, M. C. (1990). Adaptive M-estimation in nonparametric regression. *Annals of Statistics* **18,** 1712–1728.

Härdle, W. and Gasser, T. (1984). Robust non-parametric function fitting. *Journal of the Royal Statistical Society B, Series* **46,** 42–51.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the 5th Berkeley Symposium* **1,** 221–233.

Huber, P. J. (1979). Robust smoothing. In *Robustness in Statistics*, R. L. Launer and G. N. Wilkinson (eds), 38–48. New York: Academic Press.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Kauermann, G. and Carroll, R. J. (2001). The sandwich variance estimator: Efficiency properties and coverage probability of confidence intervals. *Journal of the American Statistical Association* **96,** 1387–1396.

Kondoh, M. (2003). Foraging adaptation and the relationship between food-web complexity and stability. *Science* **299,** 1388–1391.

Loader, C. (1999). *Local Regression and Likelihood*. New York: Springer.

Murdoch, W., Briggs, C., and Nisbet, R. (2003). *Consumer-Resource Dynamics*. New York: Princeton University Press.

Nagumo, J. S., Arimoto, S., and Yoshizawa, S. (1962). An active pulse transmission line simulating a nerve axon. *Proceedings of the IRE* **50,** 2061–2070.

Poyton, A. (2005). Application of principal differential analysis to parameter estimation in fundamental dynamics models. Master's Thesis, Queen's University, Kingston, Canada.

Qi, X. and Zhao, H. (2010). Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *Annals of Statistics* **38,** 435–481.

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

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society, Series B* **69,** 741–796.

Rosenzweig, M. L. and MacArthur, R. H. (1963). Graphical representation and stability conditions of predator-prey interactions. *American Naturalist* **97,** 209.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.

Shertzer, K. W., Ellner, S. P., Fussmann, G. F., and Hairston, N. G. (2002). Predator-prey cycles in an aquatic microcosm: Testing hypotheses of 30 mechanism. *Journal of Animal Ecology* **71,** 802–815.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *Journal of the Royal Statistical Society* **47,** 1–52.

Varah, J. M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific Computing* **3,** 28–46.

Vos, M., Kooi, B. W., DeAngelis, D. L., and Mooij, W. M. (2004). Inducible defences and the paradox of enrichment. *Oikos* **105,** 471–480.

Walter, E. (1987). *Identifiability of Parametric Models*. Elmsford, New York: Pergamon Press.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50,** 1–25.

Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F., and Hairston, N. G. (2003). Rapid evolution drives ecological dynamics in a predator-prey system. *Nature* **424,** 303–306.

Received April 2010. *Revised November* 2010.
Accepted December 2010.