Analysis of Variance of Integro-Differential Equations with Application to Population Dynamics of Cotton Aphids

Xueying Wang[1], Jiguo Cao[2*], and Jianhua Z. Huang[3]

Abstract

The population dynamics of cotton aphids are usually described by mechanistic models, in the form of integro-differential equations (IDEs), with the IDE parameters representing some key properties of the dynamics. Investigation of treatment effects on the population dynamics of cotton aphids is a central issue in developing successful chemical and biological controls for cotton aphids. Motivated by this important agricultural problem, we propose a framework of analysis of variance (ANOVA) of IDEs. The main challenge in estimating the IDE-based ANOVA model is that IDEs usually have no analytic solution, and repeatedly solving IDEs numerically leads to a high computational cost. We propose a penalized spline method in which spline functions are used to estimate the IDE solutions and the penalty function is defined by the IDEs. The estimated IDE solutions, as implicit functions of the parameters, are inputs in a nonlinear least squares criterion, which in turn is minimized by a Gauss-Newton algorithm. The proposed method is illustrated using simulation and an observed cotton aphids data set.

KEY WORDS: cotton aphid, generalized mechanistic model, Kindlmann-Prajneshu model, penalized spline

---

[*] Corresponding author.

[1] Xueying Wang is a joint postdoctoral fellow of Applied Mathematics and Computational Science and Department of Mathematics, Texas A&M University, College Station, TX 77843-3143, USA (Email: *xueying@math.tamu.edu*);

[2] Jiguo Cao is Assistant Professor, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada (Email: *jca76@sfu.ca*);

[3] Jianhua Z. Huang is Professor, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA (Email: *jianhua@stat.tamu.edu*).

1 Introduction

Aphids are small sap sucking insects which can draw a large amount of sap out of plants and make their leaves and stems distorted (Blackman and Eastop, 2000). Aphids inflict enormous damage on agricultural crops and forest trees (Dixon, 1998), resulting in worldwide loss of billions of dollars on food and feed grains annually (Metcalf and Metcalf, 1995). The primary economic losses caused by aphids take place on wheat, barley, corn, sorghum, oats, and cotton (Tatchell, 1989). For instance, the cotton aphids, also called *Aphis gossypii Glover*, are renowned as one of the most devastating pests of US cottons (Leclant and Deguine, 1994). In 2007, cotton aphids infected about 64% of US cotton fields, leading to 6.7 million acres of losses of cotton production (Williams, 2008).

In order to develop successful chemical and biological controls for aphids, the dynamics of aphid population have to be studied and understood. Since Barlow and Dixon (1980), there has been a considerable number of work to develop mathematical models for quantifying the aphid population dynamics. A cumulative density-dependent mechanistic model was proposed by Kindlmann (1985) and solved analytically by Prajneshu (1998). This mechanistic model, called the Kindlmann-Prajneshu model (KPM), captures the adverse effect of honeydew accumulation on aphid survival and has gained acceptance in fitting experimental aphid population data (Kindlmann et al., 2004; Matis et al., 2006, 2007a,b). Nevertheless, the solution of the KPM model is reflection symmetric and can not capture the left skewness presented in observed data; see Way (1967), Rabbinge et al. (1979), Mashanova et al. (2008), and Figure 2 of the current paper. To overcome the drawback of KPM, Matis et al. (2007c) proposed a more practical power-law generalization of KPM, called the generalized mechanistic model (GMM). Although GMM has the desired properties in modeling aphid population dynamics, it has not been used to fit experimental data mainly because it has no analytic solution and there is no method available for parameter estimation.

The purpose of this paper is to fill in the gap and develop a penalized spline method for fitting the GMM to the empirical data of cotton aphids. We develop our method in a general framework where the population dynamics are specified by general integro-differential equations (IDEs) for which the GMM is a special case, and the IDE parameters satisfy an analysis of variance (ANOVA) model. The ANOVA model has been widely used in statistics to study treatment effects, but its use in the context of IDEs is novel. To address the issue that the IDEs may not have an analytic solutions, linear combinations of spline basis functions are used to estimate the IDE solutions and penalty functions defined by the IDEs are employed to ensure the closeness of the spline approximations to the IDE solutions. In our method, the IDE solutions are implicit functions of the IDE parameters and such implicit functions are used in the nonlinear least squares fitting criterion. Our method does not rely on application of numeric IDE solvers, and thus can avoid the potentially high computational cost for repeatedly solving the IDEs.

The rest of the article is organized as follows. Section 2 reviews mechanistic models for aphid population dynamics. Section 3 describes the experimental data of aphid population, and their IDE-based ANOVA model. Our penalized spline method for estimating a general ANOVA model of IDEs is presented in Section 4. Simulation studies are conducted in Section 5 to evaluate the finite sample performance of our method. In Section 6, our method is applied on the observed data to examine the significance of the main and interaction effects of nitrogen fertilization and water irrigation treatments on cotton aphid population dynamics. Section 7 concludes the paper. Some technical details are collected in the Appendix.

2 Review of Mechanistic Models

Let $N(t)$ denote the aphid population size at time t, and $F(t) = \int_0^t N(s)\,ds$ denote the cumulative density up to time t. The mechnistic model of Kindlmann (1985) states that

$$\frac{dN(t)}{dt} = \lambda N(t) - \delta F(t) N(t), \tag{1}$$

where λ is the birth rate parameter, and δ is the death rate parameter. This model assumes that the growth of aphid population is determined by the net difference between birth rate, $\lambda N(t)$, and death rate, $\delta F(t)N(t)$. The dependence of aphid death rate on the cumulative density is supported by the facts that honeydew excreted by the aphids degrades the living environment of aphids, and the area covered by honeydew is proportional to the cumulative density of the aphid population in the past (Dixon, 1998). Prajneshu (1998) showed that the analytic solution of (1) takes the form

$$N(t) = a \exp(-bt)\{1 + d \exp(-bt)\}^{-2}. \tag{2}$$

Denote $N_0 = N(0)$ and N_{max} for the unique maximum of $N(t)$. The three parameters, a, b, and d, in (2) satisfy the following relationships: $N_0 = a/(1+d)^2$, $\lambda = b(d-1)/(d+1)$, $\delta = b^2/(2N_{max})$, and $N_0 = 4dN_{max}/(1+d)^2$. These relationships can be used to reparameterize the analytical solution using interpretable parameters, λ, δ, and N_{max}. Model (1) is often referred to as the Kindlmann-Prajneshu Model (KPM).

Figure 1 shows that the analytic solution of the KPM is reflection symmetric. Using observed data, Matis et al. (2007c) argued that the curve of aphid population size as a function of time should be left-skewed, a feature that the KPM model can not capture. Therefore, Matis et al. (2007c) suggested the following power-law generalization of the KPM

$$\frac{dN(t)}{dt} = \lambda N(t) - \delta\{F(t)\}^s N(t), \qquad \text{where } s > 1. \tag{3}$$

Equation (3) is referred to as the generalized mechanistic model (GMM). As one can see from Figure 1, the solution of the GMM is left-skewed. When $s = 1$, the GMM is reduced to the KPM, and the corresponding solution is symmetric. Thus, s is a power parameter controlling the skewness of the solution of the GMM.

Figure 1 also shows that, under the GMM, the population gradually grows before achieving its peak value, and then quickly falls down. Thus GMM captures three key features of the underlying biological principles for aphid population dynamics: (a) offspring produc-

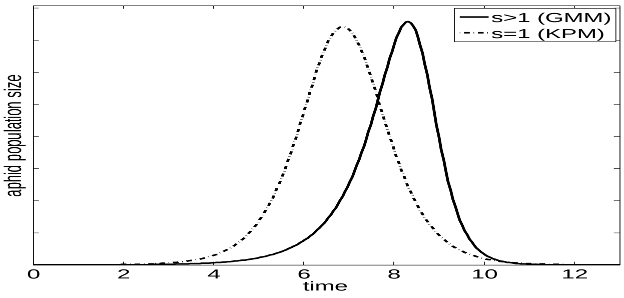

Figure 1: Solutions of the KPM (1) and the GMM (3).

tion is prolific; (b) growth is constrained by the cumulative density of the past generations; (c) the population diminishes rapidly after the population peak, which is due to the rapid departure of the aphids once their local resources are depleted.

3 Aphid experimental data and their IDE-based ANOVA model

Irrigation water and nitrogen fertilization are two primary factors affecting cotton production. To investigate the effects of various water and nitrogen treatments on the population dynamics of the cotton aphids, an experiment was conducted in 2004 at the Texas A&M Agricultural extension center in Lubbock, TX. Three irrigation levels and three nitrogen fertilization treatments are tested in a randomized block split-plot experimental design (Jones and Nachtsheim, 2009). The three irrigation levels are 65% (low), 75% (medium) and 85% (high), which are indexed as $i = 1, 2, 3$ in this paper. The three levels of nitrogen treatments are no nitrogen (zero), variable-rate-nitrogen (variable), and blanket-rate-nitrogen (blanket), which are indexed as $j = 1, 2, 3$. In this design, three blocks are chosen in distinct areas, and they are indexed as $k = 1, 2, 3$. Within each block, three different water treatments are arbitrarily applied to three whole plots. Within each whole plot, three nitrogen treatments are arbitrarily assigned to three split plots. There are a total of 27 experimental units. The number of aphids was counted at seven time points in a nearly weekly pattern between July 26 and September 10 at each experimental unit. More details of the experiment and data can be found in Matis et al. (2008).

Figure 2 displays the number of observed cotton aphids over a nearly 7-week period under the blanket-rate-nitrogen treatment. It shows that the data tend to be skewed to the left: the aphid population appears to slowly increase in the first three or four weeks; then the number of cotton aphids exponentially increases until it reaches a peak value in the next two or three weeks; thereafter, the aphid population declines rapidly.

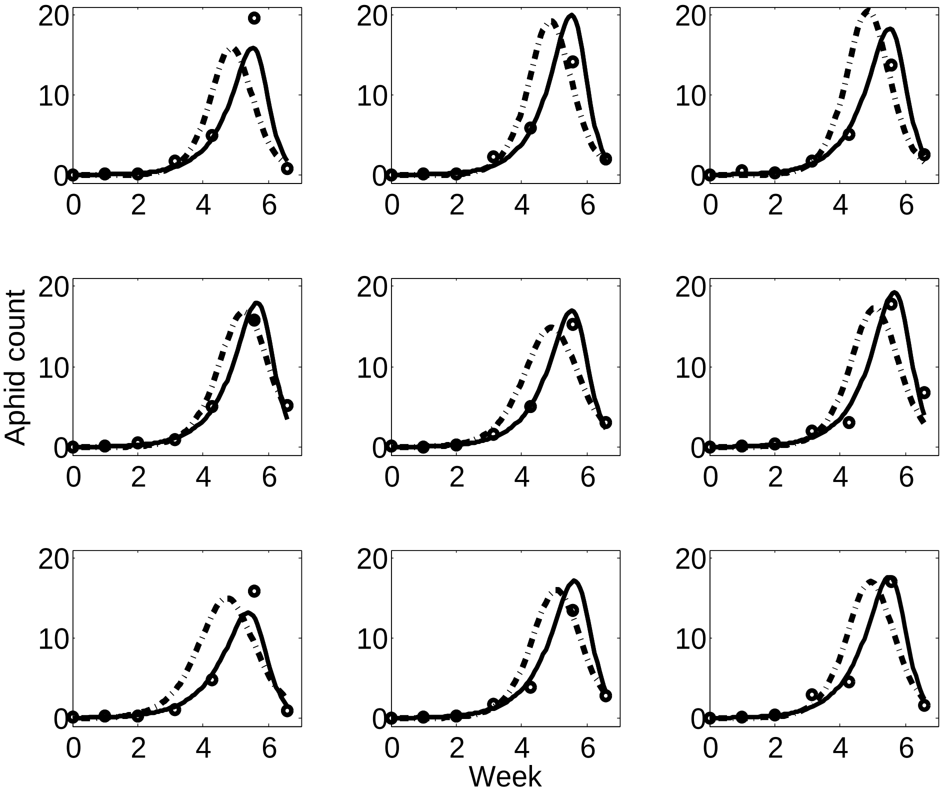

Figure 2: The number of cotton aphids over a nearly 7-week period under the blanket-rate-nitrogen treatment for all combinations of three irrigation levels and the three blocks. The columns correspond to low, medium, and high water treatments from left to right, respectively. The rows correspond to Block $1, 2$, and 3 from top to bottom, respectively. The x-axis is the time in weeks. The y-axis is the aphid count. The observed data are displayed in circles. The solid and dashed curves are the numeric solutions of the GMM (3) and the KPM (1) using the estimated parameter values.

To test the effects of water and nitrogen treatments on the dynamics of the cotton aphid population, Matis et al. (2006, 2007a,b) performed a two-way analysis of variance (ANOVA) using parameters in the fitted KPM as the response variables. The adoption of the KPM was mainly due to its tractability. However, as shown in Figure 2, the solution of the KPM is inconsistent with the skewness of observed data. Since the method that we develop in this paper does not rely on availability of analytical solutions of IDE models, we shall use the more accurate GMM in this paper.

We now state our GMM-based ANOVA method for investigating the treatment effects on the population dynamics of the cotton aphids. Let Y_{ijkn} denote the observed number of aphids at time t_{ijkn} under the i-th level of water treatment and the j-th level of nitrogen treatment at the k-th block, where $i = 1, 2, 3$, $j = 1, 2, 3$, $k = 1, 2, 3$, and $h = 1, \ldots, 7$. Because Y_{ijkn} is always positive, it is assumed to follow a log-normal distribution:

$$\log(Y_{ijkn}) = \log\{N_{ijk}(t_{ijkn})\} + \epsilon_{ijkn}, \tag{4}$$

where $N_{ijk}(t_{ijkn})$ is the true aphid population size at time t_{ijkn}, and ϵ_{ijkn}'s are independent normally distributed errors with mean 0 and variance σ^2. Let $F_{ijk}(t)$ denote the cumulative density of aphid population at time t. The GMM for the dynamics of the aphid population can be written as

$$\begin{aligned} \frac{dN_{ijk}(t)}{dt} &= \lambda_{ijk} N_{ijk}(t) - \delta_{ijk}[F_{ijk}(t)]^s N_{ijk}(t), \\ F_{ijk}(t) &= \int_0^t N_{ijk}(s)ds, \end{aligned} \tag{5}$$

in which λ_{ijk} and δ_{ijk} are birth rate and death rate of the aphid population. The KPM is a special case of (5) with $s = 1$. As in Matis et al. (2007c), these two parameters are assumed to be structured as the following ANOVA model:

$$\begin{aligned} \lambda_{ijk} &= \mu^\lambda + \alpha_i^\lambda + \xi_j^\lambda + \rho_k^\lambda + (\alpha\xi)_{ij}^\lambda\,, \\ \delta_{ijk} &= \mu^\delta + \alpha_i^\delta + \xi_j^\delta + \rho_k^\delta + (\alpha\xi)_{ij}^\delta\,, \end{aligned} \tag{6}$$

where μ^λ and μ^δ are the grand means of the birth and death rates, respectively; α_i^λ and α_i^δ, ξ_j^λ and ξ_j^δ, ρ_k^λ and ρ_k^δ are the main effects of water, nitrogen, and blocks for the birth and death rates, respectively; and $(\alpha\xi)_{ij}^\lambda$ and $(\alpha\xi)_{ij}^\delta$ are interaction effects between water and nitrogen treatments for the birth and death rates, respectively. Some constraints on the parameters are necessary for identifiabilit; details will be given in Section 4 when we introduce the general model framework.

4 Methodology

To present our methodology in a general framework, we first introduce the ANOVA model of IDEs. Without loss of generality, two factors A and B are assumed to possibly have effects on the dynamic system, and they have I and J levels respectively. Suppose each treatment combination is repeated in K blocks. Denote $X_{ijk}(t)$ as the dynamic process defined on the time interval $[0, T]$ at the k-th block under the level i of treatment A and level j of treatment B, where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$. It is assumed that the dynamic process $X_{ijk}(t)$ satisfies the following IDE:

$$\begin{cases} \dfrac{dX_{ijk}(t)}{dt} = g(X_{ijk}(t), F_{ijk}(t)|\theta_{ijk}), \\ F_{ijk}(t) = \displaystyle\int_0^t h(X_{ijk}(s))\, ds, \end{cases} \tag{7}$$

where θ_{ijk} is a vector of IDE parameters at the k-th block under the ij-th treatment combination, and $g(\cdot)$ and $h(\cdot)$ are two smooth function with known parametric forms based on the expert knowledge on the dynamic system.

Now, suppose the IDE parameters θ_{ijk} follow a two-way fixed effect ANOVA model

$$\theta_{ijk} = \mu + \alpha_i + \xi_j + \gamma_{ij}, \tag{8}$$

where μ is the grand mean, α_i and ξ_j represent, respectively, the effect of factor A at level i and of factor B at level j, and γ_{ij} is the interaction between factor A at level i and factor B at level j. For identifiability, we impose the constraints $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \xi_j = \sum_{i=1}^{I} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0$. Because of these constraints, there are only IJ free parameters. We can remove some redundant parameters by using $\alpha_I = -\sum_{i=1}^{I-1} \alpha_i$, $\xi_J = -\sum_{j=1}^{J-1} \xi_j$, $\gamma_{Ij} = -\sum_{i=1}^{I-1} \gamma_{ij}$, and $\gamma_{iJ} = -\sum_{j=1}^{J-1} \gamma_{ij}$. For the rest of this section, we use $\boldsymbol{\beta}$ to denote the column vector of free parameters in the ANOVA model, which we refer to as the ANOVA parameters.

The data are observations of the dynamic processes at discrete time points. Specifically, suppose the dynamic process $X_{ijk}(t)$ is observed at N_{ijk} time points with some measurement errors. Denote Y_{ijkn} as the observation of $X_{ijk}(t)$ at time t_{ijkn}, where $n = 1, \ldots, N_{ijk}$. We assume that $l(Y_{ijkn})$ is normally distributed with the mean $l(X_{ijk}(t_{ijkn}))$, where $l(\cdot)$ is a link function. For instance, we choose $l(\cdot) = \log(\cdot)$ in the example presented in Section 3.

If the IDE (7) had an analytical solution, denoted as $X_{ijk}(t; \theta_{ijk})$, then we could estimate the parameters by the method of nonlinear least squares which minimizes

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{n=1}^{N_{ijk}} \Big\{ l(Y_{ijkn}) - l(X_{ijk}(t_{ijkn}; \theta_{ijk})) \Big\}^2. \tag{9}$$

Unfortunately, the IDE usually does not have an analytical solution, therefore, the method of nonlinear least squares is not directly applicable. In addition, the initial conditions for the IDE are unknown. Treating the initial conditions as additional parameters, one could employ numerical IDE solvers when applying the nonlinear least squares. However, this approach has some difficulties: The computation is intensive since one needs to repeatedly solve the IDE at various candidate values of the parameters during iteratively minimization of the nonlinear least squares criterion. Besides, the numeric IDE solver sometimes fails at certain candidate values of the parameters and initial conditions. To avoid the difficulties mentioned above, we propose to estimate $X_{ijk}(t; \theta_{ijk})$ using the penalized spline method where the penalty term is defined by the IDE, and then apply the method of least squares.

Specifically, the dynamic process $X_{ijk}(t)$ is represented with a linear combination of basis functions

$$X_{ijk}(t) = \sum_{m=1}^{M} c_{ijkm}\phi_{ijkm}(t) = \boldsymbol{\phi}_{ijk}^T(t)\mathbf{c}_{ijk}, \tag{10}$$

where $\boldsymbol{\phi}_{ijk}(t) = (\phi_{ijk1}(t), \phi_{ijk2}(t), \cdots, \phi_{ijkM}(t))^T$ is a vector of basis functions, and $\mathbf{c}_{ijk} = (\mathbf{c}_{ijk1}, \cdots, \mathbf{c}_{ijkM})^T$ is the associated vector of the basis coefficients. The basis system has to be flexible enough such that the dynamic process and its derivative are well represented. In practice, we choose to use B-spline basis functions, because these functions are non-zero only in short subintervals, a feature called the compact support property, which is of great importance for efficient computation (de Boor, 2001). Usually, a large number of basis functions are required to adequately represent all the IDE components. A rule of thumb is to put one knot at each observation point so that the user does not have to select the number and locations of the knots.

To obtain an estimate of the coefficient vector $\mathbf{c}_{ijk}$, we minimize the following penalized sum of squared errors criterion:

$$\begin{aligned} G(\mathbf{c}_{ijk}|\boldsymbol{\beta}) = \sum_{n=1}^{N_{ijk}} \Big\{ l(Y_{ijkn}) - l(X_{ijk}(t_{ijkn})) \Big\}^2 \\ + \gamma \int_0^T \Big| X'_{ijk}(t) - g\big(X_{ijk}(t), F_{ijk}(t)|\theta_{ijk}\big) \Big|^2 dt, \end{aligned} \tag{11}$$

where $X_{ijk}(t)$ has the representation (10), and $\gamma > 0$ is a tuning parameter. In view of $F_{ijk}(t) = \int_0^t h(X_{ijk}(s))ds$, the integral term in (11) is essentially a double integral. Both inner and outer integrals usually do not have a close-from expression, so they have to be evaluated using numerical methods, which is explained in Appendix A. The criterion function (11) depends on the vector of ANOVA parameters $\boldsymbol{\beta}$ through θ_{ijk}. Usually γ is chosen to be a very large number such that the integral term in (11) is forced to be close to 0. Indeed, if the integral term is exactly zero, $X_{ijk}(t)$ is a solution of the IDE. The data driven choice of γ will be discussed later. On the other hand, the first term in (11) provides the necessary side conditions for solving the IDE when the initial conditions are unknown.

Denote the minimizer of (11) as $\widehat{\mathbf{c}}_{ijk} = \widehat{\mathbf{c}}_{ijk}(\boldsymbol{\beta})$, then we can express the corresponding estimate of the dynamic process $X_{ijk}(t)$ as an implicit function of $\boldsymbol{\beta}$:

$$\widehat{X}_{ijk}(t|\boldsymbol{\beta}) = \boldsymbol{\phi}_{ijk}^T(t)\widehat{\mathbf{c}}_{ijk}(\boldsymbol{\beta}). \tag{12}$$

Using the above estimate of $X_{ijk}(t)$, the nonlinear least squares criterion (9) becomes

$$H(\boldsymbol{\beta}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{n=1}^{N_{ijk}}\Big\{l(Y_{ijkn}) - l(\widehat{X}_{ijk}(t_{ijkn}|\boldsymbol{\beta}))\Big\}^2. \tag{13}$$

We denote the minimizer as $\widehat{\boldsymbol{\beta}}$ for later use. Different from the usual application of the method of nonlinear least squares, the criterion function here involves some implicit functions of the parameters. The Gauss-Newton algorithm can still be applied for solving the minimization problem but some care is needed to calculate the gradients of the implict functions. Details of the algorithm are given in Appendix B.

In addition to $\widehat{\boldsymbol{\beta}}$, our method yields an estimate for the initial condition of the IDE (7)

$$\widehat{X}_{ijk}(t_0|\widehat{\boldsymbol{\beta}}) = \boldsymbol{\phi}_{ijk}^T(t_0)\widehat{\mathbf{c}}_{ijk}(\widehat{\boldsymbol{\beta}}), \tag{14}$$

which is obtained by numerically evaluating the estimated spline function (12) at the initial time point. This estimated initial condition is useful when we validate the IDE model. Specifically, using the estimated initial condition $\widehat{X}_{ijk}(t_0|\widehat{\boldsymbol{\beta}})$, the IDE with parameter $\widehat{\boldsymbol{\beta}}$ can be solved numerically and then the IDE solution can be compared with data. The ability of our method for estimating the initial condition is valuable, because the data at the initial time point often have measurement errors or are sometimes even unavailable. Our experiences also show that using our estimated initial condition usually gives better fits to the data than using the observation of $X_{ijk}(t)$ at the initial time point as the initial condition.

The value of the tuning parameter γ can be chosen by minimizing the following sum of squared prediction errors (SSPE) criterion

$$\mathrm{SSPE}(\gamma)=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{n=1}^{N_{ijk}}\Big\{l(Y_{ijkn})-l(S_{ijk}(t_{ijkn}|\widehat{\boldsymbol{\beta}}(\gamma)))\Big\}^2, \tag{15}$$

where $\widehat{\boldsymbol{\beta}}(\gamma)$ is the estimate of the ANOVA parameter for a given γ, and $S_{ijk}(t|\widehat{\boldsymbol{\beta}}(\gamma))$ is the $X_{ijk}(t)$ component of the numeric solution to the IDE (7) when the parameter value is $\widehat{\boldsymbol{\beta}}(\gamma)$ and the initial condition is set as $\widehat{X}_{ijk}(t_0|\widehat{\boldsymbol{\beta}})$. Because $S_{ijk}(t|\widehat{\boldsymbol{\beta}}(\gamma))$ is obtained by only using the estimated initial condition, it can be viewed as a prediction of the dynamic process in the whole time interval. Note that SSPE is different from the sum of squared errors (SSE) defined in (13), because $\widehat{X}_{ijk}(t|\widehat{\boldsymbol{\beta}})$ in (13) is not obtained by solving the IDE with an initial condition.

The algorithm of the proposed method can be summarized as follows:

The algorithm of the penalized spline method

```
1. Choose an initial value for the tuning parameter $\gamma$. Set an initial guess for the model parameter $\boldsymbol{\beta}$;
2. while $\gamma < \gamma_{max}$ do;
3.     For each given $\gamma$, estimate $\hat{\boldsymbol{\beta}}(\gamma)$ by minimizing
           $H(\boldsymbol{\beta})=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{n=1}^{N_{ijk}}\Big\{l(Y_{ijkn})-l(\widehat{X}_{ijk}(t_{ijkn}|\boldsymbol{\beta},\gamma))\Big\}^2,$
       where $\widehat{X}_{ijk}(t|\boldsymbol{\beta},\gamma)=\boldsymbol{\phi}_{ijk}^T(t)\widehat{\mathbf{c}}_{ijk}(\boldsymbol{\beta},\gamma)$, and $\widehat{\mathbf{c}}_{ijk}(\boldsymbol{\beta},\gamma)$ is estimated by minimizing
           $G(\mathbf{c}_{ijk}|\boldsymbol{\beta},\gamma)=\sum_{n=1}^{N_{ijk}}\Big\{l(Y_{ijkn})-l(X_{ijk}(t_{ijkn}))\Big\}^2+\gamma\int_0^T\Big|X'_{ijk}(t)-g\big(X_{ijk}(t),F_{ijk}(t)|\boldsymbol{\theta}_{ijk}\big)\Big|^2dt,$
       for which $X_{ijk}(t):=X_{ijk}(t|\boldsymbol{\beta},\gamma)=\boldsymbol{\phi}_{ijk}^T(t)\mathbf{c}_{ijk}(\boldsymbol{\beta},\gamma)$;
4.     update $\gamma$ by a value larger than the current one, for instance, update $\gamma$ by
       $a\gamma$ (for some constant $a>1$);
5. end while;
```

6. Estimate the optimal tuning parameter $\hat{\gamma}$ by minimizing

$$\mathrm{SSPE}(\gamma) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{n=1}^{N_{ijk}}\Big\{l(Y_{ijkn}) - l(S_{ijk}(t_{ijkn}|\widehat{\boldsymbol{\beta}}(\gamma)))\Big\}^2;$$

7. Obtain the final estimate for the model parameter, that is, $\hat{\boldsymbol{\beta}}(\hat{\gamma})$.

Remark 1. The proposed method is an extension of the generalized profiling method (Ramsay et al., 2007; Cao et al., 2008, 2012) to the context of IDEs. We could rewrite the IDE (7) in the form of ordinary differential equations (ODEs) as follows:

$$\begin{aligned} X'_{ijk}(t) &= g(X_{ijk}(t), F_{ijk}(t)|\theta_{ijk}),\\ F'_{ijk}(t) &= h(X_{ijk}(t)). \end{aligned}$$

and then directly apply the generalized profiling method to the ODEs. However, this approach needs introduction of an extra set of basis coefficients to represent $F_{ijk}(t)$ and thus requires significantly more computational cost of the optimization process. We thus work directly on the IDEs instead of ODEs.

Remark 2. The two-step method for estimating ODE parameters (Ramsay and Silverman, 2005; Chen and Wu, 2008; Brunel, 2008) can be extended to the current context. Specifically, in the first step we estimate the trajectories $X_{ijk}(t)$ and corresponding derivatives $X'_{ijk}(t)$ using observed data, and in the second step, we estimate the ODE parameters by minimizing the sum of squares criterion as follows:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{n=1}^{N_{ijk}}\Big\{\widehat{X}'_{ijk}(t_{ijkn}) - g(\widehat{X}_{ijk}(t_{ijkn}), \widehat{F}_{ijk}(t_{ijkn})|\theta_{ijk})\Big\}^2,$$

where $\widehat{F}_{ijk}(t) = \int_0^t h(\widehat{X}_{ijk}(s))\,ds$. We will compare the proposed penalized spline method with this two-step method in our simulation study. Any smoothing methods such as the local polynomial kernels or penalized splines with a roughness penalty can be used in the first step. In particular, the first step used in our simulation study finds $X_{ijk}(t)$ with representation (10) by minimizing the following penalized sum of squares

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{n=1}^{N_{ijk}}\Big\{l(Y_{ijkn})-l(X_{ijk}(t_{ijkn};\theta_{ijk}))\Big\}^2+\gamma_1\mathsf{Pen}(X_{ijk}), \tag{16}$$

where $\mathsf{Pen}(X_{ijk})$ is the integrated second derivative penalty, and the penalty parameter γ_1 is selected by minimizing the GCV criterion (Ramsay and Silverman, 2005).

5 Simulation

We conducted a simulation study (with four setups) to evaluate the finite sample performance of our penalized spline method. We compared our method with the two-step method for estimating the GMM-based ANOVA model (4)–(6). We also compared our method with the classical nonlinear regression method for the special case of the KPM-based ANOVA model ($s = 1$) for which the IDEs have analytical solutions.

In the first simulation setup, we generated data from the log-normal distribution with mean $S(t|\widehat{\boldsymbol{\beta}})$ and standard deviation σ, where $\widehat{\boldsymbol{\beta}}$ is the estimated parameter vector for model (4)–(6) using the cotton aphid data described in Section 3, and $S(t|\widehat{\boldsymbol{\beta}})$ is the numeric solution of the GMM (3) using $\widehat{\boldsymbol{\beta}}$ as the parameter value. In this simulation study, the time points were set to be the same as the observed data, and the standard deviation σ was set to be 0.22, which is about 10% of the standard derivation of $S(t|\widehat{\boldsymbol{\beta}})$ at the observation times. The result are summarized in Table 1. We observe that the biases of the parameter estimates using the penalized spline method are negligible, which are only around 10%, 0.1% and 20% of those using the two-step method for the birth and death parameters and the power parameter, s, respectively. Moreover, compared with the two-step method, the penalized spline method has much smaller standard deviations and root mean squared errors (RMSEs) of the parameter estimates. Specifically, the SDs and RMSEs of estimates of the birth rate parameters are decreased by at least 25%, those of the death rate parameters are decreased by 90% or more; and those of the power parameter are decreased by around 50%.

Table 1: The biases, standard deviations (SDs), and root mean squared errors (RMSEs) of parameter estimates for the GMM-based ANOVA model using the penalized spline method (PSM) and the two-step method (2-step) for the first simulation setup. The results are based on 100 simulation replicates.

	TRUTH	BIAS		SD		RMSE	
		PSM	2-step	PSM	2-step	PSM	2-step
	$(\times10^{-2})$	$(\times10^{-3})$	$(\times10^{-3})$	$(\times10^{-2})$	$(\times10^{-2})$	$(\times10^{-2})$	$(\times10^{-2})$
μ^{λ}	121.2	-0.1	160.2	0.9	1.4	0.9	1.6
α_1^{λ}	5.0	-0.6	5.9	1.3	1.9	1.3	2.0
α_2^{λ}	0.6	1.0	-14.0	1.3	1.8	1.3	1.8
ξ_1^{λ}	-1.9	-0.1	-9.2	1.3	1.8	1.3	1.8
ξ_2^{λ}	2.7	-0.8	12.8	1.2	2.0	1.2	2.0
ρ_1^{λ}	3.9	1.1	34.1	1.2	2.0	1.2	2.0
ρ_2^{λ}	-7.5	-0.9	-12.7	1.2	1.8	1.2	2.3
$(\alpha\xi)_{11}^{\lambda}$	1.5	-2.5	40.0	1.9	3.0	1.9	3.0
$(\alpha\xi)_{21}^{\lambda}$	2.5	2.9	20.9	1.9	2.3	1.9	2.3
$(\alpha\xi)_{12}^{\lambda}$	2.6	-1.3	-14.1	1.7	2.8	1.7	2.8
$(\alpha\xi)_{21}^{\lambda}$	-3.4	-1.0	20.0	1.8	2.7	1.8	2.7
	$(\times10^{-4})$	$(\times10^{-5})$	$(\times10^{-2})$	$(\times10^{-4})$	$(\times10^{-4})$	$(\times10^{-4})$	$(\times10^{-4})$
μ^{δ}	14.9	-0.2	9.3	1.9	20.3	1.9	20.3
α_1^{δ}	-0.1	0.3	0.3	1.0	31.4	1.0	30.4
α_2^{δ}	-1.2	-0.1	-0.2	0.8	26.0	0.8	25.9
ξ_1^{δ}	5.7	-0.5	1.4	1.3	33.1	1.3	33.5
ξ_2^{δ}	-1.8	1.6	-0.2	0.9	27.1	0.9	27.1
ρ_1^{δ}	2.4	-1.0	1.0	1.1	29.0	1.1	28.8
ρ_3^{δ}	-5.7	0.7	-2.3	0.9	27.0	0.9	27.3
$(\alpha\xi)_{11}^{\delta}$	0.8	2.4	0.5	1.6	7.4	1.6	7.2
$(\alpha\xi)_{21}^{\delta}$	-3.1	0.7	-0.5	1.6	44.4	1.6	43.5
$(\alpha\xi)_{12}^{\delta}$	-1.6	-1.0	-0.4	1.1	33.5	1.1	33.0
$(\alpha\xi)_{22}^{\delta}$	3.8	-0.6	0.9	1.5	40.0	1.5	40.0
		$(\times10^{-3})$	$(\times10^{-3})$	$(\times10^{-2})$	$(\times10^{-2})$	$(\times10^{-2})$	$(\times10^{-2})$
s	2.3	6.6	31.4	3.8	8.1	3.9	8.2

To investigate the effects of the number of observations and the level of noise, we designed two additional simulation setups by slightly varying the first setup. In the second setup, the observational time points were set to be 14 equally spaced points in the time interval $[0, 6.6]$, this number of time points doubles that for the observed data, and in the third setup, the time points were set to be the same as the observed data, but σ was set to be 0.32, which is 15% of the standard derivation of $S(t|\widehat{\boldsymbol{\beta}})$ at the observation times. The results, presented in the supplementary materials, show that the penalized spline method significantly outperforms the two-step method.

In the last simulation setup, we considered a model for which the IDEs have an analytic solution and the classical nonlinear regression method is applicable and can be compared with the proposed penalized spline method. Specifically, the setup is the same as first setup except that in this setup we fit the KPM-based ANOVA model to the cotton aphid data. Results summarized in Table S3 of the supplementary materials shows that the penalized spline method works as well as the nonlinear regression method in this simple case.

Our implementation was done in MATLAB (R2011b version) on a personal Mac OS X 10.5.8 machine. The computational time when applying the penalized spline method was 118 minutes in Setup 1 for 200 simulation replicates, 136 minutes in Setup 2 for 200 simulation replicates, 375 minutes in Setup 3 for 600 simulation replicates, and 117 minutes in Setup 4 for 200 simulation replicates.

6 Analysis of Aphid Data

We applied the proposed penalized spline method to fit the GMM-based ANOVA model (4)-(6) to the cotton aphid data described in Section 3. We also considered the KPM-based model which is a special case of the GMM-based model with $s = 1$. A set of quartic B-spline basis functions with knots at each observation time point was used to represent the dynamic processes. The SSPE defined in (15) suggested the penalty parameter γ to be 10^9 for the GMM-based model, and 10^3 for the KPM-based model; see Figure S1 of the supplementary file. According to the theory derived in Ramsay et al. (2007), γ should be taken as large as possible when the data agree well with the dynamic model. Selection of a small γ by SSPE for the KPM-based model indicates that the KPM-based ANOVA model does not fit the data well.

Figure 2 presented earlier in Section 3 displays the fitted curves using the GMM-based

ANOVA model and the KPM-based ANOVA model. The fitted curves using the GMM-based ANOVA model capture the skewness, the peak time, and most of the peak values much better than those using the KPM-based ANOVA model. Moreover, the sum of squared prediction errors (SSPE) is 4.6×10^3 when using the KPM-based ANOVA model, whereas SSPE is 2.4×10^3 or 47.8% smaller when using the GMM-based ANOVA model. This result on SSPE also indicates that the GMM-based ANOVA model fits the data better than the KPM-based ANOVA model.

To make a more formal comparison of the KPM-based and GMM-based ANOVA models, we tested the null hypothesis of $s = 1$ against the alternative hypothesis $s > 1$. We used the parametric bootstrap to generate the null distribution. Specifically, we fitted the KPM-based ANOVA model, generated the bootstrap samples from the fitted model, and then fitted the GMM-based ANOVA model to the bootstrap samples—the estimates of s from the bootstrap samples form the null distribution. The 99% upper quantile of the bootstrapped null distribution is 1.4, which is much smaller than the estimate of $\hat{s} = 2.3$ for the GMM-based ANOVA model using the observed data. Thus, we had enough evidence to reject the null hypothesis of the KPM-based ANOVA model and conclude that the GMM-based ANOVA model is more plausible for this data set.

Table 2 displays parameter estimates for the GMM-based and KPM-based ANOVA models. We observe that the two models give different estimate of grand means and treatment effects on the birth and death rate. In particular, the estimates of death rate-related parameters for the GMM-based model are about only 1% in magnitude of those for the KPM-based model. This is mainly due to a larger power parameter $\hat{s} = 2.3$ for the GMM-based ANOVA model. The parameters can be interpreted as the usual ANOVA. For example, according to the fitted GMM-based ANOVA model, at low irrigation level, when the nitrogen treatment changes from the zero level to the variable level, the aphid birth rate increases from $(121.2 - 1.9) \times 10^{-2} = 1.193$ to $(121.2 + 2.7) \times 10^{-2} = 1.239$ (by around 3.9%) and the death rate decreases from $(14.9 + 5.7) \times 10^{-4} = 2.06 \times 10^{-3}$ to $(14.9 - 1.8) \times 10^{-4} = 1.31 \times 10^{-3}$ (by around 36.4%), therefore, the growth rate of aphid population is increased.

Table 2: Parameter estimates and corresponding confidence intervals for the GMM-based ANOVA model and the KPM-based ANOVA model. Here, μ^λ and μ^δ are the grand means of the birth rate and the death rate, respectively; α_i^λ and α_i^δ, ξ_j^λ and ξ_j^δ, ρ_k^λ and ρ_k^δ are main effects of water, nitrogen, and blocks for the birth and death rates, respectively; $(\alpha\xi)_{ij}^\lambda$ and $(\alpha\xi)_{ij}^\delta$ are interaction effects between water and nitrogen treatments for the birth and death rates, respectively, where $i, j, k = 1, 2$. Confidence intervals are shaded in grey if they exclude zero.

	GMM ($\times 10^{-2}$)	KPM ($\times 10^{-2}$)		GMM ($\times 10^{-4}$)	KPM ($\times 10^{-2}$)
μ^λ	121.2	156.0	μ^δ	14.9	8.6
	(119.7, 122.8)	(124.5, 137.6)		(11.4, 18.7)	(7.0, 9.2)
α_1^λ	5.0	24.2	α_1^δ	−0.1	1.9
	(2.7, 7.9)	(−1.2, 14.5)		(−2.1, 1.9)	(−0.8, 2.2)
α_2^λ	0.6	−11.6	α_2^δ	−1.2	6.9
	(−1.8, 2.8)	(−11.6, 7.6)		(−2.7, 0.5)	(−1.9, 1.1)
ξ_1^λ	−1.9	−4.7	ξ_1^δ	5.7	8.0
	(−4.3, 0.2)	(−10.7, 8.2)		(3.6, 8.3)	(−1.2, 2.3)
ξ_2^λ	2.7	−2.7	ξ_2^δ	−1.8	1.3
	(0.8, 5.3)	(−7.2, 11.1)		(−3.1, −0.2)	(−1.6, 1.6)
ρ_1^λ	3.9	−2.3	ρ_1^δ	2.4	1.5
	(1.9, 6.9)	(−8.7, 9.8)		(0.1, 4.6)	(−0.9, 2.5)
ρ_2^λ	−7.5	−44.6	ρ_2^δ	−5.7	−5.1
	(−9.9, −5.3)	(−29.2, −11.8)		(−7.7, −3.9)	(−5.5, −3.1)
$(\alpha\xi)_{11}^\lambda$	1.5	6.9	$(\alpha\xi)_{11}^\delta$	1.2	2.1
	(−1.6, 5.2)	(−12.2, 14.6)		(−2.3, 4.5)	(−0.4, 4.0)
$(\alpha\xi)_{21}^\lambda$	2.5	8.0	$(\alpha\xi)_{21}^\delta$	−3.1	−0.3
	(−0.7, 6.5)	(−14.4, 18.4)		(−6.4, −0.5)	(−3.7, 1.0)
$(\alpha\xi)_{12}^\lambda$	2.6	1.3	$(\alpha\xi)_{12}^\delta$	−1.6	−0.6
	(−1.0, 6.5)	(−12.9, 14.6)		(−3.8, 1.0)	(−2.5, 1.2)
$(\alpha\xi)_{22}^\lambda$	−3.4	−11.3	$(\alpha\xi)_{22}^\delta$	3.8	−0.4
	(−7.1, −0.2)	(−15.5, 8.4)		(1.2, 6.8)	(−1.2, 3.0)

Table 2 also displays the 95% confidence intervals for parameters in the GMM-based ANOVA model. The confidence intervals were obtained using the parametric bootstrap for which bootstrap samples are drawn from the fitted model and used to form the bootstrap distribution of the parameter estimates. We used a robust estimate of the scale parameter σ based on median absolute deviation; see supplementary materials for more details. Notice that the confidence intervals for the parameters $\mu^\lambda, \alpha_1^\lambda, \xi_2^\lambda, \rho_1^\lambda$, $\rho_2^\lambda, (\alpha\xi)_{22}^\lambda$, μ^δ, $\xi_1^\delta, \xi_2^\delta, \rho_1^\delta, \rho_2^\delta$, and $(\alpha\xi)_{22}^{\delta}$ exclude zero, which implies that these parameters are significant in the GMM-based ANOVA model. At the significance level of 5%, these results can be interpreted as follows: The birth rate of aphid population appears to be significantly higher at the low irrigation level; the death rate of aphid population appears to be significantly higher under the no-nitrogen fertilization treatment; for both the birth and death rates, there are significant interaction effects between the medium irrigation leval and the nitrogen levels. Although the KPM-based ANOVA model does not fit the data, confidence intervals for its parameters are also listed in Table 2 for completeness.

7 Conclusion

This paper develops a penalized spline method for estimating parameters of an ANOVA model of IDEs. The method does not require that the IDEs have analytic solutions, and also avoids using IDE numeric solvers. When the IDEs do not have analytic solutions, the classical nonlinear regression method can not be directly applied. The two-step method known in the literature first uses smoothing techniques to estimate the trajectories of the dynamic processes in the system and then applies the nonlinear regression. An important difference of the proposed penalized spline method is that it uses the IDEs as a regularization penalty when estimating the trajectories, and thus the estimated trajectories are implicit functions of the parameters. We developed a Gauss-Newton algorithm to minimize the nonlinear least squares criterion involving implicit functions. The results of our simulation studies indicate that the proposed method outperformed the two-step method in cases that the IDEs do not have analytic solutions, and performed comparably to the classical nonlinear regression in the special case that the IDEs have analytic solutions. In an application of the proposed method to an observed data set, we compared two mechanistic models in the literature, and our results suggest that the generalized mechanistic mode is better than the KPM in describing the dynamics of cotton aphids. ANOVA also enables us to gain insight into detection of treatment effects on the population dynamics of the cotton aphids.

Appendices

Appendix A: numerical integration methods

The integral term in (11) is essentially a double integral since $F_{ijk}(t) = \int_0^t h(X_{ijk}(s))ds$. Both the inner and outer integrals usually do not have a closed-form expression, thus, they have to be calculated using numerical integration methods. In our implementation of the method, the outer integral is evaluated using the composite Simpson's rule. The composite Simpson's rule states that, for any given integrand $f(t)$,

$$\int_{t_0}^{t_Q} f(t)dt \approx \frac{\delta}{3}\left[f(t_0) + 4\sum_{q=1}^{Q/2} f(t_{2q-1}) + 2\sum_{q=1}^{Q/2-1} f(t_{2q}) + f(t_Q)\right],$$

where the quadrature points $t_q = t_0 + q\delta$ for $q = 0, 1, \cdots, Q$, with $\delta = (t_Q - t_0)/Q$, and Q is an even integer. Note that evaluation of basis functions at the quadrature points only needs to be done once during the iterative optimization process. On the other hand, the composite trapezoidal rule is applied to calculate the inner integral, $F_{ijk}(t) = \int_0^t h(X_{ijk}(s))ds$. For any given integrand, $g(s)$, the composite trapezoidal rule is

$$\int_{s_0}^{s_P} g(s)ds \approx \frac{s_P - s_0}{P}\left[\frac{g(s_0) + g(s_P)}{2} + \sum_{p=1}^{P-1} g\Big(s_0 + p\frac{s_P - s_0}{P}\Big)\right],$$

where P is a positive integer, and s_p, $p = 0, \ldots, P$, are quadrature points in $[s_0, s_P]$. To avoid reduce the computational cost of extra function evaluation, we use the same set of quadrature points for calculating the inner and outer integrals. We use two different numerical integration methods for handling two integrals in order to reach a good balance between computational cost and accuracy. If the Simpson's rule is used for both the inner and outer integrals, we need more quadrature points in the inner integral and have to increase the computing load by evaluating the basis functions at these extra quadrature points; if the trapezoidal rule is used for both the inner and outer integrals, we will lose some numerical accuracy in the outer integral with the same number of quadrature points, because the Simpson's rule is more accurate than the trapezoidal rule for numerical integration with the same number of quadrature points.

Appendix B: the Gauss-Newton algorithm for minimizing $H(\boldsymbol{\beta})$

Start with an initial guess $\boldsymbol{\beta}^{(0)}$ for the ANOVA parameter $\boldsymbol{\beta}$. The Gauss-Newton algorithm proceeds by the iterations

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

where $\mathbf{r}$ is a column vector by stacking $r_{ijkn} = l(Y_{ijkn}) - l(\widehat{X}_{ijk}(t_{ijkn}|\boldsymbol{\theta}_{ijk}))$ together, and

$$\frac{dr_{ijkn}}{d\boldsymbol{\beta}} = -l'(\widehat{X}_{ijk}(t_{ijkn}|\boldsymbol{\beta}))\Big(\frac{d\widehat{\mathbf{c}}_{ijk}(\boldsymbol{\beta})}{d\boldsymbol{\beta}}\Big)^T \boldsymbol{\phi}_{ijk}(t_{ijkn}).$$

Since the function $\widehat{\mathbf{c}}_{ijk}(\boldsymbol{\beta})$ is usually an implicit function, it is hard to calculate the derivative $d\widehat{\mathbf{c}}_{ijk}/d\boldsymbol{\beta}$. Fortunately, we can obtain an analytic expression for this derivative using the implicit function theorem as follows. Notice that the estimate $\widehat{\mathbf{c}}_{ijk}$ satisfies the identity

$$\frac{\partial G}{\partial \mathbf{c}_{ijk}}\Big|_{\mathbf{c}_{ijk}=\widehat{\mathbf{c}}_{ijk}} \equiv 0\,.$$

Differentiating both sides of the above identity with respect to $\boldsymbol{\beta}$, we obtain

$$\frac{d}{d\boldsymbol{\beta}}\frac{\partial G}{\partial \mathbf{c}_{ijk}^T}\Big|_{\mathbf{c}_{ijk}=\widehat{\mathbf{c}}_{ijk}} = \left\{ \frac{\partial^2 G}{\partial \mathbf{c}_{ijk}^T \partial\boldsymbol{\beta}}\Big|_{\mathbf{c}_{ijk}=\widehat{\mathbf{c}}_{ijk}} \right\} + \left\{ \frac{\partial^2 G}{\partial \mathbf{c}_{ijk}^T \partial\mathbf{c}_{ijk}}\Big|_{\mathbf{c}_{ijk}=\widehat{\mathbf{c}}_{ijk}} \right\}\left\{ \frac{d\widehat{\mathbf{c}}_{ijk}^T}{d\boldsymbol{\beta}} \right\} \equiv 0,$$

which yields

$$\frac{d\widehat{\mathbf{c}}_{ijk}^T}{d\boldsymbol{\beta}} = \left\{ \frac{\partial^2 G}{\partial \mathbf{c}_{ijk}^T \partial\mathbf{c}_{ijk}}\Big|_{\mathbf{c}_{ijk}=\widehat{\mathbf{c}}_{ijk}} \right\}^{-1} \left\{ \frac{\partial^2 G}{\partial \mathbf{c}_{ijk}^T \partial\boldsymbol{\beta}}\Big|_{\mathbf{c}_{ijk}=\widehat{\mathbf{c}}_{ijk}} \right\},$$

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

Acknowledgements

The authors would like to thank Prof. Montserrat Fuentes, the associate editor, and two reviewers for their very constructive suggestions and comments, which were extremely helpful for improving this paper.

References

N. D. Barlow and A. F. G. Dixon. *Simulation of Lime Aphid Population Dynamics*. Center for Agricultural Publishing and Documentation, Wageningen, Netherlands, 1980.

R. L. Blackman and V. F. Eastop. *Aphids on the World's Crops: An Identification and Information Guide*. John Wiley and Sons Ltd., Chichester, second edition, 2000.

N. J-B. Brunel. Parameter estimation of ODE's via nonparametric estimators. *Electronic Journal of Statistics*, 2:1242–1267, 2008.

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

J. Cao, J. Z. Huang, and H. Wu. Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. *Journal of Computational and Graphical Statistics*, 21:42–56, 2012.

J. Chen and H. Wu. Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics. *Journal of the American Statistical Association*, 103(481):369–383, 2008.

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

A. F. G. Dixon. *Aphid Ecology: An Optimization Approach*. Chapman and Hall, London, second edition, 1998.

B. Jones and C. J. Nachtsheim. Split-plot designs: what, why, and how. *Journal of Quality Technology*, 41:340–361, 2009.

P. Kindlmann. A model of aphid population with age structure. In V. Capasso, E. Grosso, and S. L. Paveri-Fontana, editors, *Mathematics in Biology and Medicine, Proceedings, Bari, 1983*, pages 72–77. Springer, Berlin, 1985.

P. Kindlmann, R. Arditi, and A. F. G. Dixon. A simple aphid population model. In J. C. Simon, C. A. Dedryver, C. Rispe, and M. Hulle, editors, *Aphids in a New Millennium*, pages 325–330. INRA, Paris, 2004.

F. Leclant and J. P. Deguine. Cotton aphids. In G. A. Matthews and J. P. Tunstall, editors, *Insect Pests of Cotton*, pages 285–323. CAB International, Wallingford, 1994.

A. Mashanova, A. C. Gange, and V. A. A. Jansen. Density-dependent dispersal may explain the mid-season crash in some aphid populations. *Population Ecology*, 50:285–292, 2008.

J. H. Matis, T. R. Kiffe, T. I. Matis, and D. E. Stevenson. Application of population growth models based on cumulative size to pecan aphids. *Journal of the Agricultural, Biological, and Environmental Statistics*, 11:425–449, 2006.

J. H. Matis, T. R. Kiffe, T. I. Matis, J. A. Jackman, and H. Singh. Population size models based on cumulative size, with application to aphids. *Ecological Modelling*, 205:81–92, 2007a.

J. H. Matis, T. R. Kiffe, T. I. Matis, and D. E. Stevenson. Stochastic modeling of aphid population growth with nonlinear powerlaw dynamics. *Mathematical Biosciences*, 208: 469–494, 2007b.

J. H. Matis, L. Zhou, T. R. Kiffe, and T. I. Matis. Fitting cumulative size mechanistic models to insect population data: A nonlinear mixed effects model analysis. *J. Ind. Soc. Agril. Slulisl.*, 61:147–155, 2007c.

T.I. Matis, M. N. Parajulee, J. H. Matis, and R. B. Shrestha. A mechanistic model based analysis of cotton aphid population data. *Agricultural and Forest Entomology*, 10:355–362, 2008.

R. L. Metcalf and R. A. Metcalf. *Destructive and Useful Insects.* McGraw-Hill, New Yorker, fifth edition, 1995.

C. S. Prajneshu. A nonlinear statistical model for aphid population growth. *J. Indian Soc. Agric. Stat.*, 51:73–80, 1998.

R. Rabbinge, G. W. Ankersmit, and G. A. Pak. Epidemiology and simulation of population development of sitobion avenae in winter wheat. *Netherlands Journal of Plant Pathology*, 85:197–220, 1979.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer, New York, second edition, 2005.

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

G. M. Tatchell. An estimate of the potential economic losses to some crops due to aphids in britain. *Crop Protection*, 8:25–29, 1989.

M. J. Way. The nature and causes of annual fluctuations in numbers of aphis fabae scop. on field beans (vicia faba). *Annals of Applied Biology*, 59:175–188, 1967.

M. R. Williams. *Cotton Insect Loss Estimates*, pages 980–995. National Cotton Council, Memphis, TN, 2008.