Estimating Mixed-Effects Differential Equation Models

L. Wang [1], J. Cao [1*], J. O. Ramsay [2],

D. M. Burger [3], C. J. L. Laporte [4], and J. K. Rockstroh [5]

[1] *Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada, N6A5B7*

[2] *Department of Psychology, McGill University, Montreal, Canada, H3A1B1*

[3] *Department of Clinical Pharmacy, Radboud University, Nijmegen, Netherlands, 9102*

[4] *Ottawa Hospital Research Institute, University of Ottawa, Ottawa, Canada, K1Y4E9*

[5] *Department of Internal Medicine, University of Bonn, Bonn, Germany, 53115*

[*] Corresponding author. Email: cao@stats.uwo.ca

Abstract

Ordinary differential equations (ODEs) are popular tools for modeling complicated dynamic systems in many areas. When multiple replicates of measurements are available for the dynamic process, it is of great interest to estimate mixed-effects in the ODE model for the process. We propose a semiparametric method to estimate mixed-effects ODE models. Rather than using the ODE numeric solution directly, which requires providing initial conditions, this method estimates a spline function to approximate the dynamic process using smoothing splines. A roughness penalty term is defined using the ODEs, which measures the fidelity of the spline function to the ODEs. The smoothing parameter, which controls the trade-off between fitting the data and maintaining fidelity to the ODEs, can be specified by users or selected objectively by generalized cross validation. The spline coefficients, the ODE random effects, and the ODE fixed effects are estimated in three nested levels of optimization. Two simulation studies show that the proposed method obtains good estimates for mixed-effects ODE models. The semiparametric method is demonstrated with an application of a pharmacokinetic model in a study of HIV combination therapy.

KEY WORDS: Dynamic Models, Nonlinear Mixed-effects Models, smoothing splines

1 Introduction

Ordinary differential equations (ODEs) are widely used to model complex dynamic systems in many areas of science and technology. Although ODE models can often be proposed based on expert knowledge of the dynamic process of interest, the values of the ODE parameters are rarely known. Estimating these parameters from observational (noisy) data is an important but challenging statistical problem, because most ODEs have no analytic solutions, and solving ODEs numerically is computationally intensive. Some methods have been proposed for estimating parameters of such dynamic processes when the ODE parameters are considered as fixed effects, which are parameters associated with an entire population. For instance, Chen and Wu (2008) propose a two-step method and estimate the derivative using local polynomial regression. Ramsay et al. (2007) and Cao et al. (2008) introduce a parameter cascading method and approximate the ODE solutions with smoothing splines.

All the above methods assume independence of the observations, and do not take into account the presence of correlation between observations within the same group. In this paper, we focus on estimating mixed-effects ODE models from data that are grouped according to one or more classification factors. For instance, such grouped data can be multiple replicates of measurements or observations of a dynamic process. By using common random effects for observations in the same level of a classification factor, the mixed-effects ODE models flexibly represent the correlation structure induced by the grouping of the data.

Our study is motivated by Wasmuth et al. (2004), who investigated the pharmacokinetics of antiretroviral drugs in order to understand the widely used protease inhibitor combinations of indinavir (IDV) and ritonavir (RTV) for treating HIV-positive patients. Their study was designed to compare two different combinations of IDV and RTV. Each combination was taken by 16 healthy volunteers twice daily for 2 weeks before the serum concentrations of IDV and RTV were measured at 13 unequally-spaced time points in 12 hours. Some measurements are missing, which is a common problem for clinical data.

The pharmacokinetics of the IDV/RTV drug concentration, $C_i(t)$, $i = 1, \cdots, M$, can be modeled with a nonhomogeneous ordinary differential equation (ODE),

$$\frac{dC_i(t)}{dt} = -Ke_iC_i(t) + \frac{D_iKe_iKa_i}{Cl_i}\exp(-Ka_it)\,, \tag{1}$$

where D_i denotes the cumulative amount of unabsorbed drug at $t = 0$ within each subject, Cl_i denotes the rate of the total body drug clearance, and Ka_i and Ke_i denote the drug absorption and elimination rates, respectively. In our preliminary analysis, Ke_i shows strong homogeneity in all subjects, but the other two parameters, Ka_i and Cl_i, have large variations. Therefore, these two parameters, Ka_i and Cl_i, may be modeled as having a fixed effect component and a random effect component. It is then our interest to estimate the fixed and random effects of the ODE parameters from the observational data under a mixed-effects structure.

In general, estimating mixed-effects ODE models is challenging. When ODEs have analytic solutions, the problem is essentially a nonlinear mixed-effects model (Davidian and Giltinan 1995; Pinheiro and Bates 2000). This situation can be handled by many available packages such as the `nlme()` function in R (R Development Core Team 2005). Unfortunately, most ODEs have no analytic solutions and can only be solved using numerical methods. Several methods and software have been developed to estimate mixed-effects ODE models by solving ODEs numerically. The commercial software package, NONMEM, has been widely used to estimate nonlinear mixed effect models in the pharmacometric community for many years since the pioneering work by Sheiner et al. (1972) and Beal and Sheiner (1980). Wang (2007) provides a detailed introduction to the estimation methods used in the NONMEM software. Another popular software, MONOLIX, is well designed and free to use (MONOLIX 2010). Powerful EM-type algorithms such as stochastic approximation EM (SAEM) have been developed to estimate nonlinear mixed effects models (Kuhn and Lavielle 2005), and have been recently implemented in software programs such as NONMEM, MONOLIX and Matlab 2010a. These software programs require users to provide initial conditions for the ODE model in order to obtain numerical ODE solutions, and the estimates are sensitive to the initial conditions. Unfortunately, it is often hard to know the precise values of the initial conditions in real applications because any measurements or observations contain random errors. Huang et al. (2006) and Huang and Lu (2008) treat the initial conditions as extra parameters, and estimate mixed-effects for a dynamic model of the HIV infection process with a Bayesian MCMC approach by numerically solving the ODEs for each sampled parameter value. At the same time, they also have to estimate additional parameters for the variance-covariance structure of the random effects of initial conditions. Therefore, the dimension of the parameter space is often doubled after taking the initial conditions into account. The computation is very intensive due to the expanded parameter space and the high computational load of the numerical ODE solver.

We propose a semiparametric approach to estimate mixed-effects ODE models with or without analytic solutions. Our method estimates a spline function that approximates the ODE solution using smoothing splines, where the penalty is defined by the parametric ODE. The spline function is estimated by minimizing its lack-of-fit to the data while maintaining fidelity to the ODEs. The fidelity of the spline function to the ODEs describes the discrepancy between the spline function and the actual ODE solution. A smoothing parameter controls the trade-off between fitting the data and maintaining fidelity to the ODEs. The smoothing parameter can be estimated objectively by generalized cross validation, or selected by users based on the scales of measurement errors and their confidence about the proposed dynamic models. The idea of using spline approximation comes from Ramsay et al. (2007), which shows that using spline approximation has some appealing advantages compared to using ODE numerical solution. For example, using spline approximation avoids solving the ODE analytically or numerically, and hence the initial conditions of the ODE model are not required and the computation is very efficient. Moreover, Ramsay et al. (2007) point out that the optimization algorithm for searching ODE parameter values converges more quickly when using the spline approximation rather than directly using the ODE numerical solution, because the ODE solution is too sensitive to the ODE parameters which results in the corresponding optimization surface having many local optima.

There are both methodology and computing challenges in extending the method of Ramsay et al. (2007) to estimating mixed-effects ODE models. For example, the mixed-effects ODE models have both fixed and random effects of ODE parameters, while Ramsay et al. (2007) does not consider random effects in their method. How to incorporate random effects in the classic nonlinear mixed-effects models is already known as a difficult problem in the statistical literature (Pinheiro and Bates 2000). The challenges for estimating mixed-effects ODEs are even larger because most ODEs have no analytic solutions. In order to estimate mixed-effects ODE models, we propose to estimate spline coefficients, random effects of the ODE parameters, and fixed effects of the ODE parameters in three nested levels of optimization. Ramsay et al. (2007) only estimate spline coefficients and fixed effects of ODE parameters in two nested levels of optimization. Hence, the optimization procedure in our method is more complicated. In fact, we have to consider the functional relationship between spline coefficients, random and fixed effects of ODE parameters in order to increase computational efficiency, therefore, the computational complication increases exponentially.

The rest of the paper is organized as follows. Section 2 outlines the proposed semiparametric method for estimating mixed-effects ODE models. Section 3 demonstrates the approach with an application to HIV combination therapy. Section 4 includes two simulation studies to illustrate finite sample performance of the proposed method and to compare it with other available methods. The conclusions are given in Section 5.

2 Estimating Mixed-Effects ODE Models

In the following, we assume the dynamic model has only one component for easy notation. The extension to multiple components is straightforward. Suppose the dynamic process $x_i(t)$, $i = 1, \cdots, M$, for the i-th subject is modeled with an ODE as follows:

$$\frac{dx_i(t)}{dt} = h(x_i(t)|\boldsymbol{\beta}, \mathbf{b}_i) + u_i(t|\boldsymbol{\beta}, \mathbf{b}_i) \,, \tag{2}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of fixed effects, and $\mathbf{b}_i$ is a q-dimensional vector of random effects with the distribution $\texttt{Normal}(\mathbf{0}, \boldsymbol{\Psi})$, and $u_i(t|\boldsymbol{\beta}, \mathbf{b}_i)$ is called the *forcing function* in engineering. The dynamic process $x_i(t)$ is measured or observed at some discrete points $\mathbf{t}_i = (t_{i1}, \cdots, t_{in_i})^T$. Let $\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i})^T$ denote the vector of observations or measurements, and assume $\mathbf{y}_i = x(\mathbf{t}_i) + \boldsymbol{\epsilon}_i$, where the random error $\boldsymbol{\epsilon}_i \sim \texttt{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$. Our objective is to estimate the fixed effects $\boldsymbol{\beta}$, the variance-covariance matrix $\boldsymbol{\Psi}$, and the error variance σ^2 from the noisy data $\mathbf{y}_i$.

We estimate the above parameters in the mixed-effects ODE model in three nested levels of optimization. In the inner level of optimization, the dynamic process, $x_i(t)$, is estimated in the framework of smoothing splines by maximizing the penalized likelihood function, conditional on fixed and random effects of ODE parameters. The penalty term is defined with the ODE model, which quantifies how well the estimated spline function satisfies the ODE model. The spline coefficients, $\mathbf{c}_i$, are nuisance parameters, because they are not of primary interest. The estimated spline coefficients, $\widehat{\mathbf{c}}_i$, are treated as functions of fixed and random effects of ODE parameters, which is denoted as $\mathbf{c}_i(\boldsymbol{\beta}, \mathbf{b}_i)$. In the middle level of optimization, the random effects of ODE parameters, $\mathbf{b}_i$, are estimated by minimizing the negative joint log-likelihood function of data and the random effects, conditional on the fixed effect of ODE parameters, $\boldsymbol{\beta}$, and the variance-covariance matrix, $\boldsymbol{\Psi}$. The estimated random effects, $\widehat{\mathbf{b}}_i$, are treated as function of $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$, which is denoted as $\widehat{\mathbf{b}}_i(\boldsymbol{\beta}, \boldsymbol{\Psi})$. In the outer level of optimization, $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ are estimated by maximizing the profile log-likelihood function. The inner level of optimization is nested in the middle level of optimization, and the middle level of optimization is nested in the outer level of optimization. This algorithm is summarized at the end of Subsection 2.3. The next three subsections will introduce the optimization details in each of three levels of optimization.

2.1 Inner Level of Optimization: Spline Approximation

We approximate the dynamic process $x_i(t)$ with a linear combination of basis functions

$$z_i(t) = \sum_{k=1}^{K} c_{ik}\phi_{ik}(t) = \boldsymbol{\phi}_i(t)^T \mathbf{c}_i. \tag{3}$$

where $\boldsymbol{\phi}_i(t) = (\phi_{i1}(t), \cdots, \phi_{iK}(t))^T$ is a vector of basis functions, and $\mathbf{c}_i = (c_{i1}, \cdots, c_{iK})^T$ is the corresponding vector of basis coefficients. We choose cubic B-splines as the basis functions, since cubic B-splines are non-zero only in short subintervals, which is called the *compact support* property and is of the greatest importance for efficient computation.

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

To prevent the spline function $z_i(t)$ from over-fitting the data, a roughness penalty term is defined using the parametric ODE model:

$$PEN(z_i(t)) = \int [Lz_i(t)]^2 dt \,,$$

where the differential operator $Lz_i(t) = dz_i(t)/dt - h(z_i(t)|\boldsymbol{\beta}, \mathbf{b}_i) - u_i(t|\boldsymbol{\beta}, \mathbf{b}_i)$ measures the fidelity of the spline function $z_i(t)$ to the ODE model. The basis coefficient vector $\mathbf{c}_i$ is estimated by minimizing the penalized sum of squared errors J, conditional on the fixed effect $\boldsymbol{\beta}$ and random effects $\mathbf{b}_i$:

$$J(\mathbf{c}_i|\boldsymbol{\beta}, \mathbf{b}_i) = \sum_{j=1}^{n_i} (y_{ij} - z_i(t_j))^2 + \lambda \int [Lz_i(t)]^2 dt \,, \tag{4}$$

where λ is the smoothing parameter that controls the trade-off between fit to the data and fidelity to the ODE model. Since the spline function, $z_i(t)$, is penalized for departure from the ODE model, $z_i(t)$ must be smooth; otherwise, the derivative, $dz_i(t)/dt$, would be too large and far away from $h(z_i(t)|\boldsymbol{\beta}, \mathbf{b}_i)$. In other words, the second term in (4) also serves as the roughness penalty on the spline function, $z_i(t)$.

When $h(z_i(t)|\boldsymbol{\beta}, \mathbf{b}_i)$ is a linear function of $z_i(t)$, the estimate for $\mathbf{c}_i$ can be expressed explicitly:

$$\widehat{\mathbf{c}}_i(\boldsymbol{\beta}, \mathbf{b}_i) = [\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i + \mathbf{R}(\boldsymbol{\beta}, \mathbf{b}_i)]^{-1} [\boldsymbol{\Phi}_i^T \mathbf{y}_i + \mathbf{s}_i(\boldsymbol{\beta}, \mathbf{b}_i)] \,,$$

where $\mathbf{\Phi}_i$ is the order $n_i \times K$ matrix with the jk-th element $\phi_k(t_j)$, $\mathbf{R}(\boldsymbol{\beta}, \mathbf{b}_i) = \lambda \int [L\boldsymbol{\phi}_i(t)][L\boldsymbol{\phi}_i(t)]^T dt$, and $\mathbf{s}_i(\boldsymbol{\beta}, \mathbf{b}_i) = \lambda \int [u_i(t|\boldsymbol{\beta}, \mathbf{b}_i) L\boldsymbol{\phi}_i(t)] dt$. Because $\boldsymbol{\phi}_i(t)$ is a length-K column vector of basis functions, $\mathbf{R}(\boldsymbol{\beta}, \mathbf{b}_i)$ is an order $K \times K$ matrix, and $\mathbf{s}_i(\boldsymbol{\beta}, \mathbf{b}_i)$ is a K-dimensional vector. When $h(z_i(t)|\boldsymbol{\beta}, \mathbf{b}_i)$ is a nonlinear function of $z_i(t)$, the estimate for $\mathbf{c}_i$ has no closed-form expression, but can be obtained by the Newton-Raphson method. Letting $\widehat{\mathbf{c}}_i^{(s)}$ be the value of $\widehat{\mathbf{c}}_i$ at the s-th step, we update $\widehat{\mathbf{c}}_i$ by

$$\widehat{\mathbf{c}}_i^{(s+1)} = \widehat{\mathbf{c}}_i^{(s)} - \left(\frac{\partial^2 J}{\partial \mathbf{c}_i \partial \mathbf{c}_i^T} \bigg|_{\widehat{\mathbf{c}}_i^{(s)}} \right)^{-1} \left(\frac{\partial J}{\partial \mathbf{c}_i} \bigg|_{\widehat{\mathbf{c}}_i^{(s)}} \right).$$

Our method can also be extended to estimate a group of multiple ODEs,

$$\frac{dx_{si}(t)}{dt} = h_s(\mathbf{x}_i(t)|\boldsymbol{\beta}, \mathbf{b}_i) + u_{si}(t|\boldsymbol{\beta}, \mathbf{b}_i), \quad s = 1, \ldots, S,$$

where $\mathbf{x}_i(t) = (x_{1i}(t), \cdots, x_{Si}(t))^T$ is a vector of all S components. Each component, $x_{si}(t)$, is approximated by the B-spline function $z_{si}(t) = \sum_{k=1}^{K} c_{sik} \phi_{sik}(t) = \boldsymbol{\phi}_{si}(t)^T \mathbf{c}_{si}$. In practice, some components are often not observed and have no data available. Suppose $x_{ri}(t)$, $r \in \mathcal{I}$, is observed as y_{rij} at the point t_{rij}, where $j = 1, \ldots, n_{ri}$, and $\mathcal{I}$ is a subset of $\{1, 2, \ldots, S\}$, then the fitting criterion (4) can be generalized to be

$$\begin{aligned} J(\mathbf{c}_i|\boldsymbol{\beta}, \mathbf{b}_i) &= \sum_{r \in \mathcal{I}} \sum_{j=1}^{n_{ri}} [y_{rij} - z_{ri}(t_{rij})]^2 \\ &+ \sum_{s=1}^{S} \lambda_s \int \left[\frac{dz_{si}(t)}{dt} - h_s(\mathbf{z}_i(t)|\boldsymbol{\beta}, \mathbf{b}_i) - u_{si}(t|\boldsymbol{\beta}, \mathbf{b}_i) \right]^2 dt \,, \end{aligned} \tag{5}$$

where $\mathbf{c}_i = (\mathbf{c}_{1i}^T, \cdots, \mathbf{c}_{Si}^T)^T$, $\mathbf{z}_i(t) = (z_{1i}(t), \cdots, z_{Si}(t))^T$. When some components are not observable, their spline coefficients are not included in the first term in (5), but they can still be estimated by minimizing $J(\mathbf{c}_i|\boldsymbol{\beta}, \mathbf{b}_i)$ defined in (5) because they appear in the ODE-defined penalty term. Cao et al. (2008) showed one example with missing ODE components, although they did not consider the mixed effects in their ODE model.

2.2 Middle Level of Optimization

The $q \times q$ variance-covariance matrix $\mathbf{\Psi}$ has to be symmetric and positive-definite, so the Cholesky factorization is used to represent $\mathbf{\Psi}$ in the form of a relative precision factor, Δ, which is a $q \times q$ upper triangle matrix that satisfies

$$\sigma^2 \mathbf{\Psi}^{-1} = \Delta^T \Delta \,.$$

The Cholesky factorization is not uniquely defined for a given positive-definite matrix, but can be made unique by requiring all of the diagonal elements of Δ to be positive. So here the diagonal elements in Δ are parameterized in terms of their logarithms. The Cholesky decomposition helps us to transform a constrained optimization problem to a non-constrained optimization problem.

The random effects are estimated by minimizing the negative log joint-likelihood function of data and the random effects, $g(\mathbf{b}_i|\boldsymbol{\beta}, \Delta)$, conditional on $\boldsymbol{\beta}$ and Δ:

$$g(\mathbf{b}_i|\boldsymbol{\beta}, \Delta) = \| \mathbf{y}_i - \mathbf{\Phi}_i \widehat{\mathbf{c}}_i(\boldsymbol{\beta}, \mathbf{b}_i) \|^2 + \| \Delta \mathbf{b}_i \|^2,$$

where the last term, $\| \Delta \mathbf{b}_i \|^2$, comes from the density function of the random effects, $\mathbf{b}_i$, which is assumed to follow $\texttt{Normal}(0, \mathbf{\Psi})$, where $\mathbf{\Psi}^{-1} = \Delta^T \Delta / \sigma^2$. No analytic expression is available for the estimate $\widehat{\mathbf{b}}_i$, so we estimate $\mathbf{b}_i$ using the Newton-Raphson method. Letting $\mathbf{b}_i^{(s)}$ be the value of $\mathbf{b}_i$ at the s-th step, we update $\mathbf{b}_i$ by

$$\mathbf{b}_i^{(s+1)} = \mathbf{b}_i^{(s)} - \left(\frac{\partial^2 g}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T} \bigg|_{\mathbf{b}_i^{(s)}} \right)^{-1} \left(\frac{\partial g}{\partial \mathbf{b}_i} \bigg|_{\mathbf{b}_i^{(s)}} \right) . \qquad (6)$$

2.3 Outer Level of Optimization

The fixed effects, $\boldsymbol{\beta}$, Δ, and σ^2, are estimated by maximizing the likelihood function:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2, \Delta) &= \prod_{i=1}^{M} \int p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2, \Delta) p(\mathbf{b}_i | \boldsymbol{\beta}, \sigma^2, \Delta) d\mathbf{b}_i \\
&= \prod_{i=1}^{M} \int (2\pi\sigma^2)^{-(n_i+q)/2} |\Delta| \exp[-g(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \mathbf{b}_i)/(2\sigma^2)] d\mathbf{b}_i
\end{aligned}
$$

We consider a second-order Taylor expansion of g around $\widehat{\mathbf{b}}_i$

$$
g(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \mathbf{b}_i) \approx g(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i) + \frac{1}{2}[\mathbf{b}_i - \widehat{\mathbf{b}}_i]^T g''(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i)[\mathbf{b}_i - \widehat{\mathbf{b}}_i] \,.
$$

Note that $g'(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i) = 0$, so the linear term in the expansion disappears. Let $\mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i) = \boldsymbol{\Phi}_i \widehat{\mathbf{c}}_i(\boldsymbol{\beta}, \mathbf{b}_i)$, then we have the second derivative

$$
g''(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i) = -2 \frac{\partial^2 \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T}\bigg|_{\widehat{\mathbf{b}}_i} [\mathbf{y}_i - \mathbf{f}_i(\boldsymbol{\beta}, \widehat{\mathbf{b}}_i)] + 2 \frac{\partial \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial \mathbf{b}_i}\bigg|_{\widehat{\mathbf{b}}_i} \frac{\partial \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial \mathbf{b}_i^T}\bigg|_{\widehat{\mathbf{b}}_i} + 2\Delta^T \Delta \,,
$$

which involves the second derivatives of $\mathbf{f}_i$, but the contribution of the first term is usually negligible compared to that of the second term (Bates and Watts 1980). So the second derivative $g''(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i)$ is approximated with

$$
G(\boldsymbol{\beta}, \Delta, \mathbf{y}_i) = 2 \frac{\partial \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial \mathbf{b}_i}\bigg|_{\widehat{\mathbf{b}}_i} \frac{\partial \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial \mathbf{b}_i^T}\bigg|_{\widehat{\mathbf{b}}_i} + 2\Delta^T \Delta \,,
$$

where

$$
\begin{aligned}
\frac{\partial \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial \mathbf{b}_i} &= -\boldsymbol{\Phi}_i [\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i + \mathbf{R}]^{-1} \frac{\partial \mathbf{R}}{\partial \mathbf{b}_i} [\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i + \mathbf{R}]^{-1} \boldsymbol{\Phi}_i^T \mathbf{y}_i \,, \\
\frac{\partial \mathbf{R}}{\partial \mathbf{b}_i} &= \lambda \Big[\int [\frac{\partial L \boldsymbol{\phi}_i(t)}{\partial \mathbf{b}_i}][L \boldsymbol{\phi}_i(t)]^T dt + \int [L \boldsymbol{\phi}_i(t)][\frac{\partial L \boldsymbol{\phi}_i(t)}{\partial \mathbf{b}_i}]^T dt \Big] \,.
\end{aligned}
$$

Let $N = \sum_{i=1}^{M} n_i$, then the approximation to the logarithm of the likelihood function $L(\boldsymbol{\beta}, \sigma^2, \Delta)$ is obtained as

$$\ell(\boldsymbol{\beta}, \sigma^2, \Delta) = -\frac{N}{2}\log(2\pi\sigma^2) + M\log|\Delta| - \frac{1}{2}\Big[\sum_{i=1}^{M}\log|G(\boldsymbol{\beta}, \Delta, \mathbf{y}_i)| + \sigma^{-2}\sum_{i=1}^{M} g(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i)\Big] .$$

The variance σ^2 is obtained by maximizing $\ell(\boldsymbol{\beta}, \sigma^2, \Delta)$ for given $\boldsymbol{\beta}$ and Δ

$$\hat{\sigma}^2(\boldsymbol{\beta}, \Delta) = \sum_{i=1}^{M} g(\boldsymbol{\beta}, \Delta, \mathbf{y}_i, \widehat{\mathbf{b}}_i)/N .$$

We plug the estimate $\hat{\sigma}^2(\boldsymbol{\beta}, \Delta)$ into the log-likelihood function and get the profile log-likelihood function

$$\ell(\boldsymbol{\beta}, \Delta) = -\frac{N}{2}[1 + \log(2\pi) + \log(\hat{\sigma}^2)] + M\log|\Delta| - \frac{1}{2}\sum_{i=1}^{M}\log|G(\boldsymbol{\beta}, \Delta, \mathbf{y}_i)| . \qquad (7)$$

The fixed effects, $\boldsymbol{\beta}$ and Δ, are then obtained by maximizing the profile log-likelihood function $\ell(\boldsymbol{\beta}, \Delta)$ using the Newton-Raphson method. The Newton-Raphson method is more stable and able to converge to the optimum more quickly if the gradient is given analytically. Notice that $\widehat{\mathbf{b}}_i$ is a function of $\boldsymbol{\beta}$ and Δ, so the chain rule is applied to obtain the analytic gradient

$$\frac{d\ell}{d\boldsymbol{\beta}} = \frac{\partial\ell}{\partial\boldsymbol{\beta}} + \frac{\partial\ell}{\partial\widehat{\mathbf{b}}_i}\frac{\partial\widehat{\mathbf{b}}_i}{\partial\boldsymbol{\beta}} .$$

But $\widehat{\mathbf{b}}_i$ cannot be expressed as an explicit function of $\boldsymbol{\beta}$. We apply the implicit function theorem to obtain the analytic derivative of $\widehat{\mathbf{b}}_i$ with respect to $\boldsymbol{\beta}$

$$\frac{\partial\widehat{\mathbf{b}}_i}{\partial\boldsymbol{\beta}} = -\left[\frac{\partial^2 g(\boldsymbol{\beta}, \widehat{\mathbf{b}}_i, \Delta)}{\partial\widehat{\mathbf{b}}_i\partial\widehat{\mathbf{b}}_i^T}\right]^{-1}\left[\frac{\partial^2 g(\boldsymbol{\beta}, \widehat{\mathbf{b}}_i, \Delta)}{\partial\widehat{\mathbf{b}}_i\partial\boldsymbol{\beta}^T}\right] .$$

The analytic gradient $d\ell/d\Delta$ can be obtained similarly.

The algorithm of the proposed method for estimating mixed-effects ODE models can be summarized as follows:

---

1. Set the starting values for $\boldsymbol{\beta}$ and Δ, which are denoted as $\boldsymbol{\beta}^{(0)}$ and $\Delta^{(0)}$, respectively.

2. For $\boldsymbol{\beta} = \boldsymbol{\beta}^{(r)}$, and $\Delta = \Delta^{(r)}$,

   (a) Set the starting value for $\mathbf{b}_i$, which is denoted as $\mathbf{b}_i^{(0)}$.

   (b) For $\mathbf{b}_i = \mathbf{b}_i^{(s)}$, calculate $\widehat{\mathbf{c}}_i(\boldsymbol{\beta}, \mathbf{b}_i)$, $g(\mathbf{b}_i|\boldsymbol{\beta}, \Delta)$, $\frac{\partial g}{\partial \mathbf{b}_i}$, $\frac{\partial^2 g}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T}$.

   (c) Obtain $\mathbf{b}_i^{(s+1)}$ with the Newton-Raphson method, as defined in (6).

   (d) Set $s = s + 1$, and repeat (b), (c) until the optimization procedure converges. Set the converging value as the conditional estimate, $\widehat{\mathbf{b}}_i(\boldsymbol{\beta}, \Delta)$.

   (e) Calculate $\ell(\boldsymbol{\beta}, \Delta)$ as defined in (7), its gradient and Hessian matrix.

3. Obtain $\boldsymbol{\beta}^{(r+1)}$, and $\Delta^{(r+1)}$ with the Newton-Raphson method.

4. Set $r = r + 1$, and repeat 2, 3 until the optimization procedure converges. Set the converging values as the parameter estimates, which are denoted as $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Delta}$, respectively.

5. For $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, and $\Delta = \widehat{\Delta}$, go to Step 2 to obtain $\widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\Delta})$, and $\widehat{\mathbf{c}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\Delta}))$.

2.4 Smoothing Parameter Selection

The smoothing parameter λ in (4) controls the trade-off between fitting the data and maintaining fidelity to the ODE model. The smoothing splines method is appealing for its flexibility which allows users to try different values of the smoothing parameter λ

based on their confidence in their data or ODE model. Nevertheless, we propose an objective criterion, the generalized cross validation (GCV), to choose the optimal value for the smoothing parameters. The GCV is defined as

$$\texttt{GCV}(\lambda) = N \frac{\sum_{i=1}^{M} \parallel \mathbf{y}_i - \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i) \parallel^2}{(N - \texttt{df})^2},$$

where the degrees of freedom is defined as $\texttt{df} = \sum_{i=1}^{M} \text{trace}[\boldsymbol{\Phi}_i[\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i + \mathbf{R}(\boldsymbol{\beta}, \mathbf{b}_i)]^{-1} \boldsymbol{\Phi}_i^T]$, which can be understood as the effective number of parameters used in the smoothing splines. From the definition of GCV, we can view GCV as a measurement of the goodness-of-fit of the ODE model, with a penalty on the effective number of parameters. In this paper, the fitted model means the estimated spline function $\hat{z}_i(t)$ under a selected value of λ.

2.5 Estimating Initial Conditions of ODE Variables

The initial conditions of ODE variables are defined as the values of ODE variables at the first time point. The ODE solutions depend on the initial conditions of ODE variables, and a small change in initial conditions may result in quite different ODE solutions. The classic methods for estimating ODE parameters require ODE solutions in their estimation procedure, so they need users to either provide the initial conditions or add these initial conditions as some extra parameters. Either of these two requirements greatly increases the difficulty in the estimation process, as explained in Section 1.

Although our method does not require solving an ODE numerically in our estimation process, a useful byproduct of our method is that the initial conditions can be estimated by evaluating the fitted spline function for the ODE variable at the first point:

$$\hat{z}_i(t_0) = \boldsymbol{\phi}(t_0)\widehat{\mathbf{c}}_i \tag{8}$$

Our experience indicates that the ODE solution using our estimated initial conditions often fits the data better than using the first observation as the initial conditions. This is particularly useful when we would like to check the validation of ODEs by comparing ODE solutions to data.

3 Application to HIV Combination Therapy

In the pharmacokinetic study of HIV combination therapy, Wasmuth et al. (2004) test two different combinations of IDV and RTV: 400/100 mg IDV/RTV combination, and 600/100 mg IDV/RTV combination, which are called Treatment I and Treatment II, respectively. This study follows a crossover design with subjects randomized to either treatment. The serum concentration of IDV and RTV are measured at 0, 0.5, 1.0, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0 and 12.0 hours for 16 healthy volunteers after they take the dosage twice daily for two weeks. Figure 1 displays the concentrations of IDV in 16 subjects under Treatment I.

We apply the semiparametric method to estimate the mixed-effects in the nonhomogeneous ODE (1) from the separate IDV data and RTV data. Cubic B-splines with 25 equally-spaced knots in [0,12] are used to approximate the ODE solution. Since the parameters Ke_i , Ka_i , and Cl_i have to be positive, they are reparameterized in the logarithm scales as lKe_i , lKa_i , and lCl_i, respectively. The logarithm of the drug elimination rate, lKe_i, shows strong homogeneity in all subjects in our preliminary analysis, so we assume lKe_i does not have a random effect, and assume the fixed effect $lKe_i = lKe + \alpha T_i$, where T_i is a dummy variable with the value 0 for Treatment I and 1 for Treatment II. The mixed effect structure for the other two parameters is chosen as $(lKa_i, lCl_i)^T = \boldsymbol{\beta} + \boldsymbol{\gamma} * T_i + \mathbf{b}_i$, with the fixed effects $\boldsymbol{\beta} = (lKa, lCl)^T$, the treatment effects $\boldsymbol{\gamma} = (\gamma_1 , \gamma_2)^T$, and the random effects $\mathbf{b}_i = (b_i^{lKa}, b_i^{lCl})^T \sim \texttt{Normal}(0, \mathrm{diag}(\sigma^2_{lKa}, \sigma^2_{lCl}))$. In this mixed effects ODE

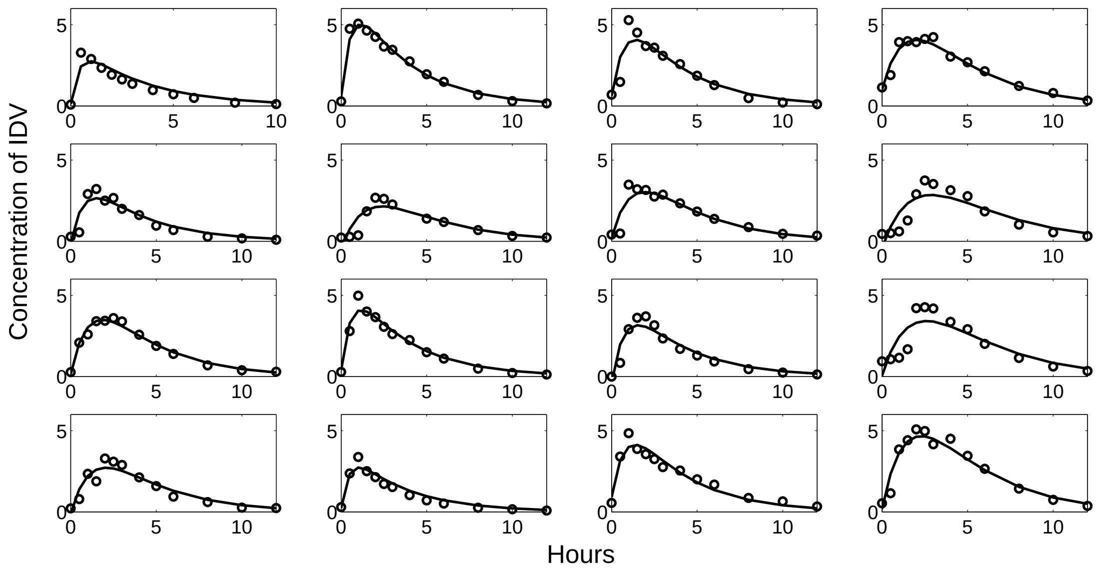

Figure 1: The serum concentrations of the indinavir (IDV) in 16 subjects are measured at 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0 and 12.0 hours after these subjects take doses of 400/100 mg IDV/RTV twice daily for 2 weeks. The solid lines are the numerical solutions of the nonhomogeneous ODE (1) with the parameters estimated using the semiparametric method and the initial conditions estimated using (8).

Table 1: The point estimates (PE) and standard errors (SE) for the mixed-effects in the nonhomogeneous ODE (1) using the semiparametric (SEMI) method for the IDV and RTV drug absorption dynamic process under Treatment I and II. The parameter Ke denotes the population drug elimination rate. The parameters Ka and Cl, denote the fix effect of the drug absorption and body drug clearance, respectively. The random effect covariance structure is assumed as $\text{diag}(\sigma^2_{Ka}, \sigma^2_{Cl})$. The parameter, σ, denotes the standard deviation of measurement errors. The standard errors for parameter estimates are obtained with the parametric bootstrap. II/I stands for the parameter ratios of Treatment II over Treatment I.

Drug		Ke I	Ke II/I	Ka I	Ka II/I	Cl I	Cl II/I	σ_{Ka}	σ_{Cl}	σ
IDV	PE	0.30	1.04	1.01	0.68	21.91	0.83	0.57	0.27	0.42
	SE	0.02	0.10	0.17	0.16	1.48	0.09	0.11	0.04	0.02
RTV	PE	0.20	0.87	0.69	0.55	42.72	1.11	0.68	0.41	0.24
	SE	0.02	0.13	0.16	0.23	4.50	0.18	0.14	0.07	0.01

model, it is of primary interest to estimate the fixed effects, $\boldsymbol{\beta}$, the treatment effects, $\boldsymbol{\gamma}$, and variances of the random effects, σ^2_{lKa} and σ^2_{lCl}.

Table 1 displays the mixed-effects estimates with the semiparametric method. Treatment I uses 33% less of the IDV drug than Treatment II, and consequently the population IDV absorption rate (Ka) and the population IDV body clearance (Cl) under Treatment II are 68% and 83% of those under Treatment I, respectively; however, the population IDV elimination rate (Ke) is very close under the two treatments. The smoothing parameters are selected as 10^4 for the IDV data and RTV data by generalized cross validation (`GCV`).

The two treatments use the same amount of the RTV drug, but the population RTV elimination rate (Ke) and absorption rate (Ka) under Treatment II are 87% and 55% of those under Treatment I, respectively; the population RTV body clearance (Cl) under Treatment II is 11% higher than that under Treatment I. This shows that the IDV drug has large effects on the absorption and elimination of the RTV drug.

Table 2: The mean and standard deviation (SD) of two pharmacokinetic parameters, the highest drug concentration (C_i^{max}) and the area under the concentration-time curve (AUC).

Drug	IDV				RTV			
Treatment	I		II		I		II	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
C_i^{max}	3.43	0.81	5.53	1.40	1.44	0.62	1.50	0.70
AUC	23.16	5.89	39.74	11.48	11.78	4.97	13.16	6.08

In this application, the subjects are dosed twice daily for two weeks before the 12-hour measurement period. We use the reported amount of the dose as the value of D_i, implicitly assuming that there is no residual unabsorbed drug from the previous doses. This is not an unreasonable assumption, because the half life of IDV is approximately $\log(0.5)/0.3 \approx 2$ hours (0.30 is the point estimate of Ke given in Table 1). Assuming that the inter-dose interval is 12 hours, the previous doses are approximately 6 half-lives distant and there is very little unabsorbed drug remaining. This assumption is also reflected in Figure 1, where all of the individual concentration profiles start at or very near 0 at $t = 0$.

The development of IDV-related toxicity depends on the highest drug concentration, expressed as $C_i^{max} = \max_{t\in[0,12]} z_i(t)$, where $z_i(t)$, defined in (3), is the spline approximation to the ODE solution. Table 2 shows that the averages of C_i^{max} are 38% and 4% smaller under Treatment I than Treatment II for the IDV and RTV drug, respectively. Higher peak plasma levels of IDV may lead to more side effects. IDV plasma concentrations above 8 mg/L are generally associated with severe side effects. No subjects under Treatment I exceed this threshold, but two subjects under Treatment II have IDV concentrations above the threshold.

The drug exposure is measured by pharmacokinetic parameters. For example, the area under the concentration-time curve is expressed as $\texttt{AUC}_i = \int_0^{12} z_i(t)dt$. Table 2 shows that the exposures to both IDV and RTV are reduced after decreasing the IDV dose from 600 mg to 400 mg. The averages of $\texttt{AUC}_i$ are 42% and 10% smaller under Treatment I than Treatment II for the IDV and RTV drug, respectively.

4 Simulations

4.1 A Mixed-Effects Nonhomogeneous ODE Model

The input/output dynamic system is widely studied in biology, engineering, and climate science. It is often modeled with the following nonhomogeneous ODE:

$$\frac{dx_i(t)}{dt} = -\theta_{1i}x_i(t) + \theta_{2i}u(t)\,, i = 1, \cdots, M\,. \tag{9}$$

For example, in the gene regulatory system, the input function, $u(t)$, is the activity of a transcription factor, and the output function, $x_i(t)$, is the expression of the i-th gene which the transcription factor regulates.

The ODE (9) can be solved analytically as follows

$$x_i(t) = e^{-\theta_{1i}t}[x_i(t_0) - \frac{\theta_{2i}}{\theta_{1i}}\int_{t_0}^{t} e^{\theta_{1i}\cdot s}u(s)ds]\,. \tag{10}$$

Define $(\theta_{1i}, \theta_{2i})^T = \boldsymbol{\beta} + \mathbf{b}_i$ with the fixed effects $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ and the random effects $\mathbf{b}_i = (b_{1i}, b_{2i})^T \sim \texttt{Normal}(0, \text{diag}(\tau_1^2, \tau_2^2))$. The data $y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$, $j = 1, \cdots, n$, where the noise ϵ_{ij} is assumed to be independently and identically distributed in $\texttt{Normal}(0, \sigma^2)$. When $u(t)$ is some special parametric function, for example, $u(t)$ is a linear function of t, then the integral in (10) has an analytic formula. However, when $u(t)$ is a general parametric function or a spline function, the integral in (10) does not have an analytic formula. In this case, it is very hard to use routine NLME packages to estimate the mixed

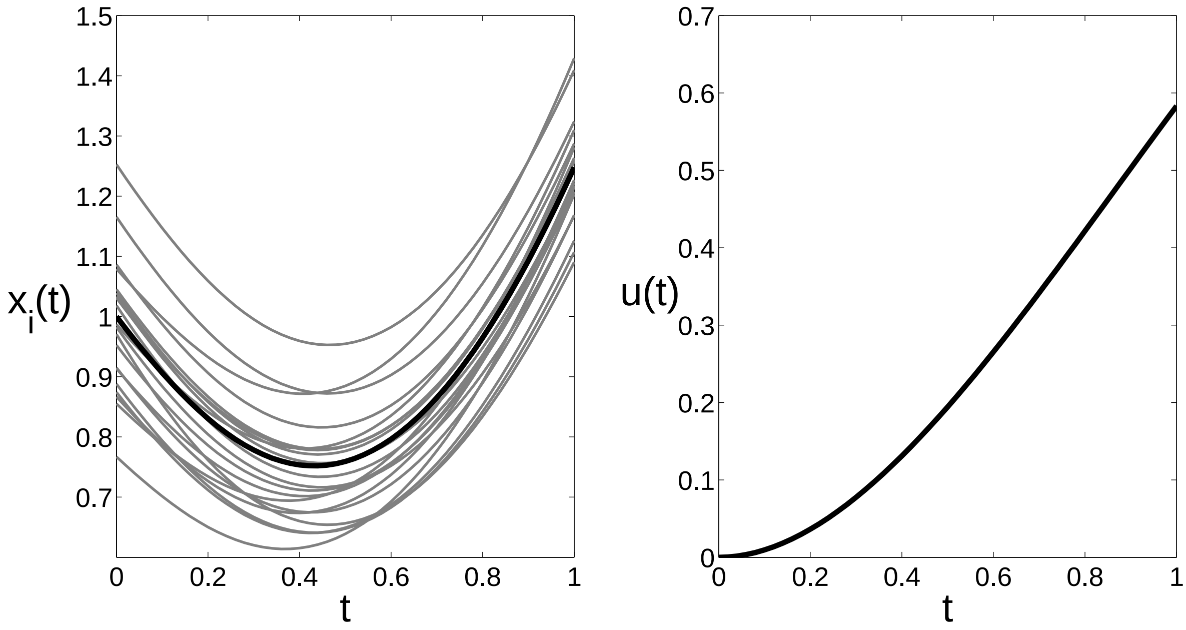

Figure 2: The numeric solutions of the ODE $dx_i(t)/dt = -\theta_{1i}x_i(t)+\theta_{2i}u(t)$, $i = 1, \cdots, 20$, which are used to generate one simulated data set. The ODE parameters θ_{1i} and θ_{2i} contains both fixed and random effects. The thick black line is the ODE solution when the parameters θ_{1i} and θ_{2i} are equal to the fixed effects. The forcing function $u(t) = \sin(t)\log(1+t)$.

effects in the model. Since our method does not need to solve ODEs, the mixed effects in the ODE model (9) can be estimated using our method, even though the integral in (10) does not have an analytic formula.

A simulation study was implemented based on the nonlinear mixed-effects model (9) with the forcing function $u(t) = \sin(t)\log(1+t)$. Notice that the integral in (10) does not have an analytic formula with this forcing function, hence routine NLME packages cannot be used in this case. The ODE (9) is numerically solved at n equally-spaced points in [0,1] with the parameter value $(\theta_{1i}, \theta_{2i})^T = \boldsymbol{\beta}+\mathbf{b}_i$ and the initial value $x_i(t_0) \sim \texttt{Normal}(1, 0.01)$. Figure 2 displays one example of the numerical ODE solutions that are used to generate one simulated data set. The simulated data are generated by adding noise distributed as $\texttt{Normal}(0, \sigma^2)$ to the ODE solutions. The true parameter values are given in Table 3 .

Table 3: The biases, standard deviations (STDs), and root mean squared errors (RMSEs) of parameter estimates for the mixed effect model (9) when varying the number of subjects (N) and the number of measurements per subject (n) using the semiparametric (SEMI) method and the stochastic approximation EM (SAEM) method implemented in MONOLIX. The parameters, β_1 and β_2, denote the fixed effects, τ_1 and τ_2 denote the STDs of random effects, and σ denotes the SD of measurement errors.

	Parameters	True	BIAS*10		SD*10		RMSE*10	
			SEMI	SAEM	SEMI	SAEM	SEMI	SAEM
$N = 20$	β_1	1.00	-0.14	0.30	0.83	0.86	0.84	0.91
	β_2	5.00	-0.29	1.35	2.87	3.00	2.88	3.29
	τ_1	0.10	-0.01	0.55	0.23	0.39	0.23	0.67
$n = 10$	τ_2	0.10	-0.04	1.32	0.26	0.81	0.26	1.55
	σ	0.10	-0.07	4e-3	0.06	0.07	0.09	0.07
$N = 50$	β_1	1.00	-0.07	0.40	0.63	0.60	0.64	0.72
	β_2	5.00	-0.24	1.55	2.15	2.01	2.17	2.54
	τ_1	0.10	0.01	0.64	0.19	0.22	0.19	0.68
$n = 10$	τ_2	0.10	-0.04	1.72	0.18	0.53	0.18	1.80
	σ	0.10	-0.05	-0.05	0.04	0.04	0.06	0.04
$N = 20$	β_1	1.00	-0.10	0.15	0.65	0.93	0.66	0.95
	β_2	5.00	-0.32	0.73	2.37	3.09	2.39	3.17
	τ_1	0.10	-0.04	0.80	0.20	0.41	0.20	0.90
$n = 20$	τ_2	0.10	-0.02	1.85	0.11	1.41	0.12	2.33
	σ	0.10	-0.03	0.04	0.04	0.04	0.05	0.06

The mixed-effects in the nonhomogeneous ODE (9) are estimated with our method from the simulated data in 100 simulation replicates. Our method is also compared with the stochastic approximation EM (SAEM) method implemented in MONOLIX on the same simulated data. Table 3 summarizes the parameter estimates when varying the number of subjects (N) and the number of measurements per subject (n). The estimates for the fixed effects and the standard deviations (STDs) of random effects with the semiparametric method have smaller biases, standard deviations and root mean squared errors (RMSEs) than those developed with the SAEM method, although the SAEM method has slightly better estimates for the standard deviation of measurement errors. When the number of measurements per subject is $n = 10$, and the number of subjects is $N = 20$, the RMSEs of the estimates for the fixed effects and STDs of the random effects using the semiparametric method are decreased by 8%, 12%, 66%, and 83% compared to those obtained using the SAEM method, respectively. For the other two cases when the number of measurements per subject n increases from 10 to 20, or the number of subjects N increases from 20 to 50, the gains using the semiparametric method are even larger. It is worth pointing out that the SAEM method implemented in MONOLIX requires users to provide the initial conditions of the ODE model for all subjects. The above SAEM estimates are based on initial conditions that are set to their true values. Otherwise, the SAEM estimates would be worse than those shown in Table 3.

The effects of measurement errors on parameter estimation are assessed by varying the standard deviation of measurement errors, $\sigma = 0.10, 0.15, 0.20$. The estimation results are displayed in Table 4. The STDs and RMSEs of the estimated fixed effects increase when σ is larger. On the other hand, the biases, STDs and RMSEs of the estimated STDs of random effects and the estimated SD of measurement errors are similar under three scales of measurement errors.

Table 4: The biases, standard deviations (STDs), and root mean squared errors (RMSEs) of parameter estimates for the mixed effect model (9) when varying SD of measurement errors, $\sigma = 0.10, 0.15, 0.20$. The number of subjects is 20, and the number of measurements per subject is 20. The estimates for fixed effects are denoted as $\hat{\beta}_1$ and $\hat{\beta}_2$, the estimates for STDs of random effects are denoted as $\hat{\tau}_1$ and $\hat{\tau}_2$, and the estimate for SD of measurement errors is denoted as $\hat{\sigma}$. Their true values are set as $\beta_1 = 1.00$, $\beta_2 = 5.00$, $\tau_1 = 0.10$, and $\tau_2 = 0.10$.

	True σ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\sigma}$
BIAS∗10	0.10	-0.10	-0.32	-0.04	-0.02	-0.03
	0.15	0.10	0.34	-0.03	-0.01	-0.04
	0.20	0.02	0.07	0.03	0.01	-0.05
SD∗10	0.10	0.65	2.37	0.20	0.11	0.04
	0.15	0.90	3.16	0.21	0.15	0.06
	0.2	1.40	4.77	0.21	0.15	0.07
RMSE∗10	0.10	0.66	2.39	0.20	0.12	0.05
	0.15	0.91	3.18	0.21	0.15	0.07
	0.20	1.40	4.77	0.21	0.15	0.08

4.2 A Mixed-Effects Homogeneous ODE Model

When the ODEs have simple analytic solutions, the mixed effects in the ODE model can also be estimated with the `nlme()` function in R. A simple population growth model with analytic solution was chosen in the second simulation study to compare the semiparametric method with the `nlme()` function in R.

The first mathematical model for population growth, referred to as the classical Malthusian scheme, is based on the work by Thomas R. Malthus (1766-1834). In *The Principle of Population* essay published in 1798, Malthus assumes that the population growth rate for one generation proportionally depends on the size of the previous generation. This population growth model can be expressed in a simple homogeneous ODE

$$\frac{dx_i(t)}{dt} = -\theta_i x_i(t)\,, \tag{11}$$

where $x_i(t)$, $i = 1, \cdots, N$, is the population for the i-th species, and θ_i is their growth rate. We assume the ODE parameter $\theta_i = \eta + b_i$, where $b_i \sim \text{Normal}(0, \sigma_b^2)$. The simulated data are generated by adding noise distributed as $\texttt{Normal}(0, \sigma^2)$ to the ODE solution in which the ODE is solved with the initial value $x_{i0} \sim \texttt{Normal}(x_0, \sigma_0^2)$. The semiparametric method is applied to estimate the fixed effect η, and variance parameters σ^2 and σ_b^2 from observations.

We choose the ODE (11) because it has an analytic solution $x_i(t) = x_{i0} \exp(-\theta_i(t - t_0))$. Then the mixed-effects ODE model (11) can also be expressed in terms of the standard nonlinear mixed-effects (NLME) model

$$y_{ij} = x_{i0} \exp(-\theta_i(t_j - t_0)) \, , j = 0, \cdots, n \, .$$

The above NLME model can be estimated with the `nlme()` function in R. In this paper, we call the estimation method that transforms an ODE model to a NLME model using the ODE's analytical solutions the *parametric method*. The parametric method constrains itself in only working for ODE models with analytic solutions. In addition, the above NLME model adds one extra parameter x_{i0} after solving ODE (11). Hence, the parametric method has to estimate the fixed effect in x_{i0}, the variance of the random effect in x_{i0}, and the correlation of the random effect in x_{i0} to other random effects. So the number of parameters to estimate is often double after introducing this extra parameter x_{i0}. In contrast, our semiparametric method estimates the ODE model directly without introducing this extra parameter.

Set $n = 50$, $\eta = 3$, $x_0 = 1$, $\sigma_0^2 = 0.1$, $\sigma = 0.03$, and $\sigma_b = 0.3$; then the relative precision parameter $\Delta = \sqrt{\sigma^2/\sigma_b^2} = 0.1$. Since Δ has to be positive, we estimate it on the logarithm scale $\rho = \ln(\Delta)$. The estimate for σ_b^2 is then obtained as $\hat{\sigma}_b^2 = \hat{\sigma}^2 \exp(-2\hat{\rho})$. The semiparametric method is compared with the parametric method under different sample

Table 5: Investigation of the effect of the sample size N by varying $N = 20$ and 50. We compare the mean, bias, standard deviation (SD) and root mean squared error (RMSE) of parameter estimates for the mixed-effects model (11) using the semiparametric (SEMI) method and the parametric (PARA) method. Notice that the semiparametric method does not solve the ODE, and thus does not need to estimate the initial conditions' fixed effect x_0, and the random effect variance σ_0^2.

$N = 20$

		BIAS$*10^2$		SD$*10^2$		RMSE$*10^2$	
	True	SEMI	PARA	SEMI	PARA	SEMI	PARA
η	3.00	-0.24	-0.16	1.57	1.51	1.59	1.52
σ_b	0.03	0.11	-1.05	1.13	2.19	1.14	2.43
σ	0.03	-0.065	-0.010	0.004	0.073	0.065	0.074
x_0	1.00	-	0.29	-	2.14	-	2.16
σ_0	0.10	-	-0.25	-	1.65	-	1.67

$N = 50$

		BIAS$*10^2$		SD$*10^2$		RMSE$*10^2$	
	True	SEMI	PARA	SEMI	PARA	SEMI	PARA
η	3.00	-0.06	-0.07	1.02	1.00	1.02	1.00
σ_b	0.03	0.06	-0.61	0.35	1.83	0.35	1.93
σ	0.03	-0.02	0.004	0.002	0.045	0.022	0.045
x_0	1.00	-	-0.13	-	1.49	-	1.50
σ_0	0.10	-	-0.23	-	0.90	-	0.93

sizes $N = 20$ and $N = 50$ using 100 simulation replicates for each scenario. The parameter estimates are summarized in Table 5. The semiparametric method obtains more accurate estimates for the random effect standard deviation σ_b and the noise standard deviation σ than the parametric method, although the estimates for the the fixed effect η using the semiparametric method are slightly worse than those using the parametric method. The root mean squared errors (RMSEs) for $\hat{\sigma}_b$ using the semiparametric method are 53% and 82% smaller than those using the parametric method when the sample size is $N = 20$ and $N = 50$, respectively. The semiparametric method has 12% and 51% smaller RMSEs for $\hat{\sigma}$ than the parametric method when the sample size is $N = 20$ and $N = 50$, respectively.

5 Conclusions

ODEs are popular modelling tools for describing dynamic systems in biology, medicine, engineering, and many other areas. It is common to have multiple replicates for the measurements of a dynamic process; hence, it is of great interest to estimate mixed-effects ODE models. In comparison with fixed-effect modeling, mixed-effects modeling has the advantages of considering and quantifying the correlation among multiple replicates, and obtaining more precise parameter estimation by pooling all data together. Most ODEs have no analytic solutions, so the available packages for nonlinear mixed-effects models cannot be used directly.

A semiparametric method is proposed to estimate mixed-effects ODE models. A spline function is estimated to approximate the dynamic process using smoothing splines, which avoids the high computational load of a numerical ODE solver. Since the method does not solve the ODE analytically or numerically, estimates of the initial conditions of the ODE variables are not required. The parametric ODEs define the penalty term, which measures the fidelity of the spline function to the ODEs. A smoothing parameter controls the trade-off between fitting the data and maintaining fidelity to ODEs, which can be selected objectively by minimizing the `GCV`. It also allows users to choose the smoothing parameter values based on their expert knowledge of their data measurement accuracy and the appropriateness of their ODE models.

Three nested levels of optimization are designed to estimate the spline coefficients, ODE random effects, and ODE fixed effects in our semiparametric method. The Newton-Raphson algorithm is used in each level, and the analytical gradients are derived to increase the computational efficiency. A very interesting problem for future research is to use the difference between the spline function and the ODE solution as a diagnostic for model misspecification.

The application and simulation studies discussed in this paper demonstrate that the proposed method provides good estimates for mixed-effects ODE models. It is also robust to the starting values of parameters when fitting the model. The semiparametric method reduces standard deviations and mean squared errors of parameter estimates when compared to current methods for estimating nonlinear mixed-effects ODE models.

Acknowledgements

The authors would like to thank Prof. Gilles Celeux, the associate editor, two reviewers, and Prof. Duncan Murdoch for their very constructive suggestions and comments, which are extremely helpful for improving this paper. This research was supported by discovery grants from the Natural Science and Engineering Research Council of Canada (NSERC) to J. Cao and J. O. Ramsay.

References

Bates, D. M. and D. B. Watts (1980). Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society, Series B 42*, 1–25.

Beal, S. L. and L. B. Sheiner (1980). The NONMEM system. *The American Statistician 34*, 118.

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

Chen, J. and H. Wu (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*(481), 369–383.

Davidian, M. and D. M. Giltinan (1995). *Nonlinear Models for Repeated Measurement Data*. London: Chapman Hall.

Huang, Y., D. Liu, and H. Wu (2006). Hierachical bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics 62*, 413–423.

Huang, Y. and T. Lu (2008). Modeling long-term longitudinal HIV dynamics with application to an aids clinical study. *The Annals of Applied Statistics 2*, 1384–1408.

Kuhn, E. and M. Lavielle (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis 49*, 1020–1038.

MONOLIX (2010). *MONOLIX 3.2 User Guide*. http://software.monolix.org.

Pinheiro, J. C. and D. M. Bates (2000). *Mixed-effects models in S and S-Plus*. New York: Springer.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. ISBN 3-900051-07-0.

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

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

Sheiner, L. B., B. Rosenberg, and K. L. Melmon (1972). Modelling of individual pharmacokinetics for computer aided drug dosage. *Comput Biomed Res 5*, 441–459.

Wang, Y. (2007). Derivation of various nonmem estimation methods. *J. Pharmacokinet Pharmacodyn 34*, 575–593.

Wasmuth, J., C. J. la Porte, K. Schneider, D. M. Burger, and J. K. Rockstroh (2004). Comparison of two reduced-dose regimens of indinavir (600 mg vs 400 mg twice daily) and ritonavir (100 mg twice daily) in healthy volunteers (coredir). *International Medical Press 2*, 1359–6535.