Estimation of nonlinear differential equation model for glucose-insulin dynamics in type I diabetic patients using generalized smoothing*

By Inna Chervoneva[1], Boris Freydin[1], Brian Hipszer[1], Tatiyana V. Apanasovich[2], and Jeffrey I Joseph[1]

[1] *Thomas Jefferson University,* [2] *George Washington University*

Abstract

In this work, we develop an ordinary differential equations (ODE) model of physiological regulation of glycemia in type 1 diabetes mellitus (T1DM) patients in response to meals and intravenous insulin infusion. Unlike for majority of existing mathematical models of glucose-insulin dynamics, parameters in our model are estimable from a relatively small number of noisy observations of plasma glucose and insulin concentrations. For estimation, we adopt the generalized smoothing estimation of nonlinear dynamic systems of Ramsay et al. (2007). In this framework, the ODE solution is approximated with a penalized spline, where the ODE model is incorporated in the penalty. We propose to optimize the generalized smoothing by using penalty weights that minimize the covariance penalties criterion (Efron, 2004). The covariance penalties criterion provides an estimate of the prediction error for

* This work was supported by the grant DK081088 from the NIH/NIDDK.

Key words and phrases: Generalized profiling; Covariance penalties; Parameter cascading; Penalized smoothing; Prediction error; Profiled penalty estimation.

nonlinear estimation rules resulting from nonlinear and/or non-homogeneous ODE models, such as our model of glucose-insulin dynamics. We also propose to select the optimal number and location of knots for B-spline bases used to represent the ODE solution. The results of the small simulation study demonstrate advantages of optimized generalized smoothing in terms of smaller estimation errors for ODE parameters and smaller prediction errors for solutions of differential equations. Using the proposed approach to analyze the glucose and insulin concentration data in T1DM patients we obtained good approximation of global glucose-insulin dynamics and physiologically meaningful parameter estimates.

1 Introduction

Diabetes mellitus is a chronic metabolic disease associated with abnormalities in glucose metabolism that affect the uptake of glucose by tissues and causes abnormal glucose excursions in the blood. Type 1 diabetes mellitus (T1DM) is characterized by total insulin deficiency. In insulin-deficient persons, the blood glucose levels have been roughly controlled using insulin alone. Insulin can be injected (multiple daily injection) or infused (continuous intravenous or subcutaneous insulin infusion). Glucose levels of hospitalized diabetic patients are usually managed with intravenous insulin infusions guided by glucose readings either from a lab-based blood sample assay or point of care glucose meter. However, infrequent testing of blood glucose levels does not provide sufficient trend information to achieve the desired goals of near-

normal glucose control. Better glycemic control in T1DM in-hospital patients may be achieved utilizing the continuous intravenous (IV) insulin infusion or insulin pump delivering rapid-acting insulin continuously through a subcutaneous tissue catheter, but automated regulation of the insulin delivery requires constantly updated information about the blood glucose levels and appropriate algorithms. Recently developed continuous glucose monitoring sensors continuously measure the concentration of glucose in the blood or interstitial fluid and display an averaged glucose value every one to five minutes. However, there is no commonly accepted algorithm available for controllers to allow automated regulation of the insulin delivery based on the sensor feedback. Such an algorithm requires a relatively simple parsimonious model for glucose-insulin dynamics validated using real patient data. We propose a parsimonious model, which builds upon previously considered models of glucose-insulin dynamics and may be estimated using a moderate number of plasma glucose and insulin measures and further used for designing algorithms for an automated insulin delivery system.

The motivating glucose-insulin dynamics data for four T1DM subjects were collected as a part of the VIA Blood Glucose Monitor Study conducted at the Jefferson Artificial Pancreas Center of Thomas Jefferson University, Philadelphia, PA. The VIA Blood Glucose Monitor (currently GlucoScout, International Biomedical, Austin, TX) is an FDA approved device that can be connected to an existing catheter inserted in a peripheral vein. Samples of blood are automatically transported from the blood-

stream into the flow-through sensor. The concentration of glucose is measured using an enzyme-based electrochemical method. This self-calibrating device can perform measurements as frequently as every five minutes. Blood glucose samples were collected during an 8.5-hour protocol with breakfast at 30 minutes, lunch at 240 minutes, and exercise at 450 minutes for 30 minutes. Respectively, the first 7.5 hours (450 minutes) were used to fit the proposed model of physiological regulation of glycemia in response to meals and insulin infusion. The measurements collected included blood glucose every 5 minutes with VIA Blood Glucose Analyzer and plasma insulin levels every 10 minutes. The insulin infusion protocol provided for a 120-minute square-wave bolus starting simultaneously with ingestion of the meal. The initial dose was based upon a preprandial blood glucose measurement (capillary blood sample tested on a commercial glucometer). Sixty minutes after the beginning the meal, the square-wave insulin bolus was adjusted based on a second capillary blood measurement. The basal infusion rate (0.5 U/hr) of insulin was added to the bolus dose over the 120-minute meal period. After 120 minutes, the bolus dose was completed, and the infusion rate returned to the original basal level. Glucose and insulin concentrations and IV insulin infusion rates as a function of study time as well as the times of the meals for one T1DM subject are shown in Figure 1.

Many mathematical models of glucose-insulin dynamics in humans and animals have been proposed to date (e.g. Sherwin, Kramer, Tobinet, et al, 1974; Bergman,

Ider, Bowden, et al, 1979; Fischer, Salzsieder, Jutzi, et al, 1984; Sorensen, 1985; Cobelli, Bier, and Ferrannini, 1990; Lehmann and Deutsch, 1992; Trajanoski, Wach, Kotanko, et al, 1993; Shimoda, Nishida, Sakakida, et al, 1997; Worthington, 1997; Lehman, 1997; De Gaetano and Arino, 2000; Wilinska, Chassin, Schaller, et al, 2005). All existing models describe glucose-insulin dynamics in terms of a system of non-linear ordinary differential equations (ODE). It is not feasible to convert most of these models directly into estimable statistical models because of the large number of latent (mostly not measurable) variables describing the time-dependent levels of glucose, insulin, free fatty acids and glucagon in internal physiological compartments, such as heart, liver etc. The validation of these models consist of finding a subset of parameters that produce predicted glucose profiles closest to the observed data in a least squares sense using an iterative curve-fitting algorithm and exhaustive grid search. Other parameters are fixed at known biologically plausible levels.

In this work, a relatively simple model of glucose-insulin dynamics is developed for physiological regulation of glycemia in response to meals and IV insulin infusion in T1DM subjects. The model is validated using the real data collected from four T1DM subjects studied in the clinical research unit. To our knowledge, models of glucose dynamics in response to meals and insulin infusion in T1DM subjects have not been reported in the literature before. It is understood that the proposed model is a simplified approximation of the true glucose-insulin dynamics. It is expected that

exact numeric solution of the estimated ODE model would capture global dynamics of the physiological process, but not necessarily provide good fit to observed data in all time subintervals. Therefore, we adapt the generalized smoothing (generalized profiling) approach of Ramsay, Hooker, Campbell, and Cao (2007), which is designed to handle data that may be viewed as an approximate rather than precise solution of the postulated ODE model. Section 3 provides a short formal description of the generalized profiling approach. Some alternative smoothing-based methods of ODE model estimation (e.g. Varah 1982; Brunel, 2008; Chen and Wu, 2008) proceed with first finding a nonparametric smoother of the data and then considering this smoother to be a solution of the ODE model in order to estimate ODE parameters. Nonlinear least squares (e.g. Li, Osborne and Pravan, 2005; Xue, Miao and Wu, 2010) and Bayesian methods (e.g. Huang, Liu and Wu, 2006; Donnet and Samson, 2007) are other approaches for estimating parameters of the ODE models.

Implementation of the generalized smoothing approach requires selecting the basis for representing the ODE solution and the penalty weight(s). Here, we propose to optimize the generalized smoothing solutions of non-linear and non-homogeneous ODEs by selecting the penalty weight(s) so that the estimated prediction error is minimized. As an unbiased estimate of the prediction error, we adapt the covariance penalty (Efron, 2004), which is suitable for non-linear prediction rules. In the context of generalized profiling, the penalty is used to express the difference between

the derivative of the ODE solution approximation and right side of the ODE model and does not serve the usual purpose of regularization. Therefore, for regularization and performance optimization, it is proposed to use B-spline bases with optimized number and location of the knots using the knot selection methodology and software recently developed by Spiriti, Eubank, Smith, and Young (2013).

In section 2, we develop a parsimonious model of glucose-insulin dynamics in T1DM patients. Generalized profile estimation methodology is summarized in section 3. Section 4 describes the proposed optimization for generalized profile estimation of nonlinear and/or non-homogeneous ODE models. The optimized generalized profiling is applied to estimate the new model of glucose-insulin dynamics using real data collected from T1DM patients in section 5. Simulation study is presented in section 6. Section 7 concludes with discussion.

2 Model of glucose-insulin dynamics in T1DM patients

A comprehensive glucose-insulin dynamics model in T1DM patients may be viewed as comprised of three components: a glucose metabolism model, a meal model, and an insulin kinetics model. The model of glucose metabolism describes the regulation of glucose uptake and production in the body – primarily controlled by the concentrations of glucose and insulin. The meal model describes the rate of glucose absorption from the intestine. The model of insulin kinetics is required to describe the rate at

which insulin diffuses into, and is eliminated from, the systemic circulation.

Most of glucose-insulin dynamics models developed to date focus on non-diabetic or type 2 diabetic patients, for whom the plasma glucose concentration affects the insulin production. In T1DM patients, the pancreas does not release insulin, and the model has to incorporate an exogenous insulin input. Meanwhile, it is plausible to assume that plasma insulin concentration affects the glucose concentration similarly in diabetic and non-diabetic patients. Recently proposed models for glucose and insulin dynamics after an IV glucose tolerance test (IVGTT) (De Gaetano and Arino 2000; Mukhopadhyay et al 2004) assume the following model for the derivative $\dot{G}(t)$ of plasma glucose concentration $G(t)$ $[mg/dl]$ as a function of the plasma insulin concentration $I(t)$ $[mU/L]$:

$$\dot{G}(t) = b_0 - b_1 G(t) - b_2 G(t) I(t), \tag{1}$$

where b_0 $[(mg/dl)min^{-1}]$ is the constant increase in plasma glucose concentration due to constant baseline liver glucose release, b_1 $[min^{-1}]$ is the spontaneous glucose 1st order disappearance rate, and b_2 $[min^{-1}\,(L/mU)]$ is the insulin-dependent glucose disappearance rate. This model has to be extended to incorporate the terms describing the external glucose absorption from the meals. In this work, we adapt one of the most parsimonious meal absorption models, which assumes that the rate of glucose appearance in the blood follows the model $\mu(t - t_M)_+ \exp\left(\nu(t - t_M)_+\right)$, where t_M is the starting time of the meal, ν is the parameter related to the rate of

glucose absorption from the gut and μ is the product of the distribution volume and the constant related to the total amount of carbohydrates consumed (Monks, 1990). With IV insulin infusion, the kinetics of plasma insulin concentration $I(t)$ $[mU/L]$ is described by the one-compartment model (e.g. Steil et al, 2003; Hipszer 2005):

$$\dot{I}(t) = -c_1 I(t) + c_2 r(t),$$

where c_1 $[min^{-1}]$ is the insulin 1st order disappearance rate, c_2 $[L^{-1}]$ is the reciprocal of the volume of the insulin distribution space, and $r(t)$ $[mU/min]$ is the IV insulin infusion rate. For non-diabetic and type 2 diabetic patients, $\dot{I}(t)$ also depends on the glucose concentration $G(t)$, but for T1DM patients, the pancreas does not release insulin, and insulin concentration $I(t)$ is independent of $G(t)$. Thus, we consider the following nonhomogeneous nonlinear ODE model for an individual T1DM subject on IV insulin delivery and consuming L meals at times $t_{M_1}, ..., t_{M_L}$:

$$\dot{G}(t) = b_0 - b_1 G(t) - b_2 G(t) I(t) + \sum_{i=1}^{L} \mu_i (t - t_{M_i})_+ \exp\left(\nu_i (t - t_{M_i})_+\right) \tag{2}$$

$$\dot{I}(t) = -c_1 I(t) + c_2 r(t), \tag{3}$$

For nonlinear ODE systems, identifiability of all parameters is not a trivial issue. In our models, the second differential equation (3) is independent of $G(t)$, linear, and has a closed form solution. Hence parameters c_1 and c_2 are identifiable based on measured $I(t)$ and known $r(t)$. Identifiability of parameters in (2) is straightforward to show using the multiple time points method of Wu et al (2008).

3 Generalized profiling estimation of ODE models

Following Ramsay et al (2007), it is assumed that a vector of d output functions or states $\mathbf{x}(t) = [x_1(t), ..., x_d(t)]$, $t \in [0, T]$ depends on a vector of g input functions $\mathbf{u}(t) = [u_1(t), ..., u_g(t)]$. The ODE model is written for the derivative vector $\dot{\mathbf{x}}(t)$ as a known vector function $\mathbf{f}$ of t, $\mathbf{x}(t)$, $\mathbf{u}(t)$, and unknown $p \times 1$ parameter vector $\boldsymbol{\theta}$,

$$\dot{\mathbf{x}}(t) = \mathbf{f}\left(\mathbf{x}(t), \mathbf{u}(t), t, \boldsymbol{\theta}\right), \ t \in [0, T]. \tag{4}$$

It is assumed that a subset of $m \leq d$ output functions are measured with error on some grid of points $0 \leq t_{jl} \leq T, \ j \in \mathcal{I} \subset \{1, ..., d\}$, $1 \leq l \leq N_j$, (possibly different grids for different j). The measurement error model is $y_{jl} = x_j(t_{jl}) + \varepsilon_{jl}$, where $y_{jl} = y\left(t_{jl}\right)$ are observed values and $\varepsilon_{jl} \sim i.i.d.N(0, \sigma_j^2)$. Each state $x_j(t)$, $j = 1, ..., d$, is approximated by an expansion with respect to some basis $B_j = \left\{\psi_{jk}(t), 1 \leq k \leq K_j\right\}$,

$$\widetilde{x}_j(t) = \sum_{k=1}^{K_j} \alpha_{jk}\psi_{jk}(t) = \boldsymbol{\alpha}_j'\boldsymbol{\psi}_j(t) = \boldsymbol{\psi}_j'(t)\boldsymbol{\alpha}_j \tag{5}$$

where $\boldsymbol{\alpha}_j = \left[\alpha_{j1}, ..., \alpha_{jK_j}\right]'$ and $\boldsymbol{\psi}_j(t) = \left[\psi_{j1}(t), ..., \psi_{jK_j}(t)\right]'$.

The data fitting criterion is defined as a negative log-likelihood

$$\mathbf{H}(\boldsymbol{\sigma}, \boldsymbol{\alpha}) = -\sum_{j=1}^{m} \ln\left\{g\left(\mathbf{e}_j | \sigma_j, \boldsymbol{\alpha}_j\right)\right\} \tag{6}$$

of the error terms vectors $\mathbf{e}_j = \mathbf{y}_j - \boldsymbol{\Psi}_j\boldsymbol{\alpha}_j$, where $\mathbf{y}_j = \left[y_{j1}, ..., y_{jN_j}\right]$, $\boldsymbol{\Psi}_j$ is the $N_j \times K_j$ matrix of basis functions $\psi_{j1}(t), ..., \psi_{jK_j}(t)$ evaluated at N_j time points,

$\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1', ..., \boldsymbol{\alpha}_m']'$, $\boldsymbol{\sigma} = [\sigma_1, ..., \sigma_m]'$, and $g(\mathbf{e}_j|\sigma_j, \boldsymbol{\alpha}_j)$ is the density of the error terms. The estimate $\widehat{\boldsymbol{\alpha}}$ of the parameter $\boldsymbol{\alpha}$ is computed by minimizing the penalized criterion

$$\mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathbf{H}(\boldsymbol{\sigma}, \boldsymbol{\alpha}) + \sum_{j=1}^{d} \lambda_j PEN_j(\widetilde{\mathbf{x}}|\boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_d]'$ and the penalty for the jth output function is defined by

$$PEN_j(\widetilde{\mathbf{x}}|\boldsymbol{\theta}) = \int \left\{ \frac{d}{dt}\widetilde{x}_j(t) - f_j\left(\widetilde{\mathbf{x}}(t), \mathbf{u}(t), t, \boldsymbol{\theta}\right) \right\}^2 dt. \tag{8}$$

The composite penalty $\sum_{j=1}^{m} \lambda_j PEN_j(\widetilde{\mathbf{x}}|\boldsymbol{\theta})$ measures the closeness of $\widetilde{\mathbf{x}}(t)$ to the solution of (4). Parameter $\boldsymbol{\alpha}$ is considered a nuisance parameter, which depends on the structural parameter $\boldsymbol{\theta}$ and smoothing parameter $\boldsymbol{\lambda}$, with function $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ defined implicitly (under the assumptions of implicit function theorem) by

$$\frac{\partial \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}} = \mathbf{0} \tag{9}$$

in some open neighborhood of the minimum of the penalized criterion (7). The generalized profiling approach (a.k.a. generalized smoothing) estimates structural parameter vector $\boldsymbol{\theta}$ by minimizing (6) with respect to $\boldsymbol{\theta}$ using Gauss-Newton algorithm. The necessary gradient vector is

$$\frac{d\mathbf{H}(\boldsymbol{\alpha}(\boldsymbol{\theta}))}{d\boldsymbol{\theta}'} = \frac{\partial \mathbf{H}(\boldsymbol{\alpha}(\boldsymbol{\theta}))}{\partial \boldsymbol{\alpha}'} \frac{d\boldsymbol{\alpha}(\boldsymbol{\theta})}{d\boldsymbol{\theta}'}$$

where $d\boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda}) / d\boldsymbol{\theta}'$ is computed using the implicit function theorem,

$$\frac{d\boldsymbol{\alpha}(\boldsymbol{\theta})}{d\boldsymbol{\theta}'} = -\left(\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}' \partial \boldsymbol{\alpha}} \right)^{-1} \frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{d\boldsymbol{\theta}' \partial \boldsymbol{\alpha}}. \tag{10}$$

After each update of $\boldsymbol{\theta}$, parameter $\boldsymbol{\alpha}$ is updated by minimizing (7) conditionally on the current value of $\boldsymbol{\theta}$ and a priori chosen $\boldsymbol{\lambda}$. Representation of $\widetilde{x}_j(t)$ through the basis expansion (5) automatically yields the derivative needed to evaluate (8),

$$\frac{d}{dt}\widetilde{x}_j(t) = \sum_{k_0}^{K_j} \alpha_{jk} \frac{d}{dt}\psi_{jk}(t) = \boldsymbol{\alpha}_j' \dot{\boldsymbol{\psi}}_j(t) = \dot{\boldsymbol{\psi}}_j(t)' \boldsymbol{\alpha}_j.$$

Notably, the penalized criterion (7) is not minimized jointly with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ because joint minimization yielded unsatisfactory estimates (Heckman and Ramsay, 2000). When $e_{jl} \sim i.i.d.N(0, \sigma_j^2)$, minimizing (6) is equivalent to minimizing

$$\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\sigma}) = \sum_{j=1}^{m} w_j \left(\mathbf{y}_j - \boldsymbol{\Psi}_j \boldsymbol{\alpha}_j\right)' \left(\mathbf{y}_j - \boldsymbol{\Psi}_j \boldsymbol{\alpha}_j\right), \tag{11}$$

where $w_j = \sigma_j^{-2}$ (Ramsay et al (2007) mention other alternatives for w_j). We use (11) with weights $w_j = \sigma_j^{-2}$ estimated prior to the generalized profile estimation along with the starting values and drop dependence of $\mathbf{H}$ on $\boldsymbol{\sigma}$ from further notation.

In the original profiled estimation work, Ramsay et al (2007) assume that $\boldsymbol{\lambda}$ is fixed a priori. Cao and Ramsay (2007) view $\boldsymbol{\lambda}$ as a complexity parameter and propose optimization with respect to $\boldsymbol{\lambda}$ as the third outer level of optimization. Denoting a suitable outer optimization criterion by $\mathbf{F}(\boldsymbol{\lambda}, \boldsymbol{\theta}(\boldsymbol{\lambda}), \boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda}))$, they propose to optimize $\mathbf{F}(\boldsymbol{\lambda}, \boldsymbol{\theta}(\boldsymbol{\lambda}), \boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda}))$ as a function of $\boldsymbol{\lambda}$, so that every time $\boldsymbol{\lambda}$ is changed, $\boldsymbol{\theta}$ is updated by minimizing $\mathbf{H}(\boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda}))$ as described above. Respectively, for every change in $\boldsymbol{\theta}$, the estimates of the nuisance parameters $\boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ are updated by minimizing $\mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})$. Thus, the overall (outer) optimization is carried on with respect to $\boldsymbol{\lambda}$, the middle level

of optimization of $\mathbf{H}(\boldsymbol{\alpha}(\boldsymbol{\theta}, \boldsymbol{\lambda}))$ is nested within the outer level optimization, and the inner level of optimization of $\mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})$ is nested within the middle level optimization. The term "parameter cascade" was introduced to reflect this multistage optimization with respect to $\mathbf{F}$, $\mathbf{H}$, and $\mathbf{J}$.

4 Optimization of generalized profiling estimation

For generalized smoothing, the most important question is the choice of the penalty weights $\boldsymbol{\lambda}$. Cao and Ramsay (2007, 2009) adapted the generalized cross validation (GCV) criterion (Craven and Wahba, 1979) to select the weights for the penalties that may be represented as a quadratic form in the spline coefficients, which corresponds to linear differential operators defined by the ODE model. For the models with non-linear and/or non-homogeneous ODEs, this approach is not appropriate because there is no associated smoothing matrix that would linearly transform data into spline coefficients. On the other hand, a direct implementation of leave-one-out cross validation to minimize the prediction error would be extremely computationally intense for non-sparse functional data and nonlinear ODE models. The GCV criterion provides an estimate of the prediction error for linear estimation rules. In particular, the GCV criterion is an unbiased estimate of the prediction error for penalized splines (e.g. Gu, 2002). Optimization of tuning parameter $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_m]'$ for the generalized smoothing of nonlinear and/or nonhomogeneous ODEs may be based on an estimate

of the prediction error, which would be suitable for nonlinear estimation rules as well as for the linear ones. The covariance penalties approach (Efron, 2004) provides such an estimate of the prediction error. Following Efron (2004), let us denote the mean function by $\boldsymbol{\mu}$ and corresponding prediction by $\widehat{\boldsymbol{\mu}} = m(\mathbf{y})$, where $m(\mathbf{y})$ is a nonlinear estimation rule. An unbiased estimate of the prediction error is given by

$$\widehat{Err} = \sum_{j=1}^{m} w_j \left\{ \sum_{1 \le l \le N_j} \left(y_{jl} - \widehat{\mu}_{jl}\right)^2 + 2 \sum_{1 \le l \le N_j} cov(\widehat{\mu}_{jl}, y_{jl}) \right\}.$$

Efron (2004) used the parametric bootstrap to estimate $2\sum_{1 \le l \le N_j} cov(\widehat{\mu}_{jl}, y_{jl})$ (covariance penalty) in the general case. In the case of Gaussian errors and a differentiable mapping $\widehat{\boldsymbol{\mu}} = m(\mathbf{y})$, $cov(\widehat{\mu}_{jl}, y_{jl})$ may be estimated by $\sigma_j^2 \left(\partial \widehat{\mu}_{jl} / \partial y_{jl}\right)$ (Stein, 1981; Ye, 1998; Efron, 2004), so that with $w_j = \sigma_j^{-2}$, one obtains

$$\widehat{Err} = \sum_{j=1}^{m} \left\{ \sum_{1 \le l \le N_j} w_j \left(y_{jl} - \widehat{\mu}_{jl}\right)^2 + 2 \sum_{1 \le l \le N_j} \frac{\partial \widehat{\mu}_{jl}}{\partial y_{jl}} \right\}.$$

For linear prediction rules, using the covariance penalties is asymptotically equivalent to using the GCV criterion (Efron, 2004).

If $\left\{\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}\right)\right\}$ is the generalized smoothing solution computed for fixed $\boldsymbol{\lambda}$, then the predicted mean function for state $x_j(t)$ is $\widehat{\boldsymbol{\mu}}_j = \boldsymbol{\Psi}_j \widehat{\boldsymbol{\alpha}}_j$. Respectively, $\widehat{\mu}_{jl} = \boldsymbol{\psi}_j(t_{jl})' \widehat{\boldsymbol{\alpha}}_j$ and (11) implies that $\sum_{j=1}^{m} \sum_{1 \le l \le N_j} w_j \left(y_{jl} - \widehat{\mu}_{jl}\right)^2 = \mathbf{H}\left(\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}\right)\right)$. Further, $\frac{\partial}{\partial y_{jl}} \boldsymbol{\psi}_j(t_{jl})' \widehat{\boldsymbol{\alpha}}_j = \boldsymbol{\psi}_j(t_{jl})' \frac{\partial}{\partial y_{jl}} \widehat{\boldsymbol{\alpha}}_j$ and the covariance penalties criterion for the generalized profiling may be written as

$$\mathbf{F}\left(\boldsymbol{\lambda}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\right) = \mathbf{H}\left(\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}\right)\right) + 4 \sum_{j=1}^{m} \sum_{1 \le l \le N_j} \boldsymbol{\psi}_j(t_{jl})' \frac{\partial}{\partial y_{jl}} \widehat{\boldsymbol{\alpha}}_j.$$

Since equation $\partial \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})/\partial \boldsymbol{\alpha} = \mathbf{0}$ is satisfied by $\widehat{\boldsymbol{\alpha}}$, the derivatives $\partial \widehat{\boldsymbol{\alpha}}_j / \partial y_{jl}$ may be computed using the inverse function theorem,

$$\frac{\partial \widehat{\boldsymbol{\alpha}}_j}{\partial y_{jl}} = -\left(\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j}\right)^{-1} \frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial y_{jl} \partial \boldsymbol{\alpha}_j}\Bigg|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}},$$

where $\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})/\partial y_{jl} \partial \boldsymbol{\alpha}_j = -2 w_j \boldsymbol{\psi}_j(t_{jl})$ and expression for $\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})/\partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j$ is given in the Appendix. Hence, the covariance penalties criterion is given by

$$\mathbf{F}(\boldsymbol{\lambda}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = \mathbf{H}(\widehat{\boldsymbol{\alpha}}) + 4 \sum_{j=1}^{m} w_j \sum_{1 \le l \le N_j} \boldsymbol{\psi}_j(t_{jl})' \left(\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j}\right)^{-1}\Bigg|_{\substack{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}} \\ \boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}} \boldsymbol{\psi}_j(t_{jl}). \quad (12)$$

It is proposed to compute $\widehat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \mathbf{F}\left((\boldsymbol{\lambda}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\right)$ using the Gauss–Newton optimization to find a point of minimum, such that at every iteration step, the profiled solution $\left\{\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}\right)\right\}$ is updated conditionally on $\boldsymbol{\lambda}$ as described in section 3.

The Gauss–Newton optimization requires the gradient $\frac{d\mathbf{F}(\boldsymbol{\lambda}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})}{d\boldsymbol{\lambda}'}$. Assuming that differentiation under the integral is appropriate, this gradient is computed as

$$\frac{d}{d\boldsymbol{\lambda}'} \mathbf{F}\left(\boldsymbol{\lambda}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\right) = \frac{\partial \mathbf{F}}{\partial \boldsymbol{\lambda}'} + \frac{\partial \mathbf{F}}{\partial \boldsymbol{\alpha}'} \frac{d\boldsymbol{\alpha}}{d\boldsymbol{\lambda}'}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}$$

where $d\boldsymbol{\alpha}/d\boldsymbol{\lambda}'$ may be computed the same way as $d\boldsymbol{\alpha}/d\boldsymbol{\theta}'$ in (10),

$$\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\lambda}'} = -\left(\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}' \partial \boldsymbol{\alpha}}\right)^{-1} \frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}' \partial \boldsymbol{\alpha}},$$

$\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})/\partial \lambda_k \partial \boldsymbol{\alpha}'$ is given in the Appendix, and

$$\begin{aligned}
\frac{\partial \mathbf{F}}{\partial \lambda_k} &= \sum_{j=1}^{m} 4 w_j \sum_{1 \le l \le N_j} \frac{\partial}{\partial \lambda_k} trace\left(\left(\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j}\right)^{-1}\Bigg|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}} \boldsymbol{\psi}_j(t_{jl}) \boldsymbol{\psi}_j(t_{jl})'\right) \\
&= \sum_{j=1}^{m} 4 w_j \sum_{1 \le l \le N_j} \left[Vec\left\{\left(\frac{\partial^3 \mathbf{J}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})}{\partial \lambda_k \partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j}\right)^{-T}\Bigg|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}\right\}^T Vec\left(\boldsymbol{\psi}_j(t_{jl}) \boldsymbol{\psi}_j(t_{jl})'\right)\right]
\end{aligned}$$

$$\frac{\partial \mathbf{F}}{\partial \boldsymbol{\alpha}_k'} = -2w_k \left(\mathbf{y}_k - \boldsymbol{\Psi}_k \boldsymbol{\alpha}_k\right)' \boldsymbol{\Psi}_k + \sum_{j=1}^{m} 4w_j \sum_{1 \le l \le N_j} \frac{\partial}{\partial \boldsymbol{\alpha}_k'} \left[Vec \left\{ \left(\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j} \right)^{-T} \Bigg|_{\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}} \right\}^T Vec\left(\boldsymbol{\psi}_j(t_{jl}) \boldsymbol{\psi}_j(t_{jl})'\right) \right]$$

with $\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j}$ and $\frac{\partial^3 \mathbf{J}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \lambda_k \partial \boldsymbol{\alpha}_j' \partial \boldsymbol{\alpha}_j}$ also given in the Appendix.

For generalized profile estimation, it is standard to use B-spline bases with equally spaced knots. By analogy with penalized splines, it is expected that a sufficiently large number of basis vectors K (e.g. $n/2$, where n is the number of data points) would be appropriate to minimize the bias (Ramsay et al, 2007). However, the results of Claeskens et al (2009) imply that with a large number of equally spaced knots, penalty weight is the only effective smoothing parameter for penalized splines. Thus, if nonsmooth input functions are present in the ODE model, then penalty (8) may not serve the purpose of regularization as a roughness penalty. In particular, this is the case in our data with insulin infusion rate being a piece-wise constant step function. Therefore, it is proposed to optimize the number and location of knots for B-spline bases. With a small number of knots, penalized splines behave asymptotically as least-squares splines and the knots become operative smoothing parameters (Claeskens et al, 2009). Optimizing the number and/or location of knots has been previously considered in the context of free-knot splines (e.g. Jupp, 1978; Hu, 1993; Lindstrom, 1999). It may be performed using the standard model selection criteria such as Akaike information criterion (AIC), Bayesian information criterion

(BIC), and GCV, but it is known to be very computationally intensive and often intractable using derivative-based optimization methodology. Recently, Spiriti et al (2013) developed two stochastic search algorithms for selecting both the number and location of knots. The algorithms combined with adjusted GCV, AIC, or BIC optimization criteria are implemented in the R package `freeknotsplines` available at the Comprehensive R Archive Network (CRAN) (http://cran.r-project.org/). Spiriti et al (2013) report excellent numerical performance of the new algorithms at producing knot locations that are near optimal in the sense of average squared error loss. We propose to adapt this state-of-the-art methodology for selection of free knots for B-spline bases representing the ODE solution(s) for generalized smoothing. Knots selection is optimized using only the observed data, without taking in account the ODE model. This allows using the standard penalization of the first or second derivative for the purpose of selecting the knots for penalized spline optimally smoothing the data. Then the optimized sequence of knots is used to construct corresponding B-spline basis for representing the ODE solution for generalized profiling estimation, which can be performed using the R package `CollocInfer` or Matlab software functions for profiled estimation of differential equations (Giles Hooker, 2010) available at http://faculty.bscb.cornell.edu/~hooker/profile_webpages/. To obtain the starting values and weights for the error sum-of-squares required for implementation of the generalized profile estimation, we first smooth the data points nonparametrically and

optimally in the classic sense of cross validation and then use the standard nonlinear regression methods to estimate the structural parameters of ODEs. The resulting estimates serve as the starting values for the generalized profile estimation. The necessary weights are the reciprocals of the cross validation estimates for the standard deviation of the errors, $w_j = \widehat{\sigma}_j^{-2}$.

5 Application to glucose-insulin dynamics in T1DM subjects

For parameter estimation, model (2) was re-parameterized as following to impose physiologically meaningful signs of the original parameters:

$$\begin{aligned} \dot{G}(t) &= \theta_1 - e^{\theta_2} G(t) - e^{\theta_3} G(t) I(t) \\ &\quad + |\theta_6| (t - t_{M_1})_+ e^{-|\theta_7|(t-t_{M_1})_+} + |\theta_8| (t - t_{M_2})_+ e^{-|\theta_9|(t-t_{M_2})_+} \qquad (13) \\ \dot{I}(t) &= -e^{\theta_4} I(t) + \theta_5 r(t), \end{aligned}$$

so that $b_0 = \theta_1$, $b_1 = e^{\theta_2}$, $b_2 = e^{\theta_3}$, $c_1 = e^{\theta_4}$, $c_2 = \theta_5$, $\mu_1 = |\theta_6|$, $\nu_1 = -|\theta_7|$, $\mu_2 = |\theta_8|$, $\nu_2 = -|\theta_9|$, and 9×1 vector of all structural parameters is $\boldsymbol{\theta} = [\theta_1, ..., \theta_9]$.

System (13) may be written in the general form (4) with $\mathbf{x}(t) = [x_1(t), x_2(t)]'$ and

$$\begin{aligned} f_1(\mathbf{x}(t), \mathbf{u}(t), t, \boldsymbol{\theta}) &= \theta_1 - e^{\theta_2} x_1(t) - e^{\theta_3} x_1(t) x_2(t) \\ &\quad + |\theta_6| u_1(t) \exp\{-|\theta_7| u_1(t)\} + |\theta_8| u_2(t) \exp\{-|\theta_9| u_2(t)\} \\ f_2(\mathbf{x}(t), \mathbf{u}(t), t, \boldsymbol{\theta}) &= -e^{\theta_4} x_2(t) + \theta_5 u_3(t), \end{aligned}$$

where $x_1(t) = G(t)$, $x_2(t) = I(t)$, $u_1(t) = (t - t_{M_1})_+$, $u_2(t) = (t - t_{M_2})_+$, $u_3(t) = r(t)$, $\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t)]'$.

The number and locations of knots were optimized separately for glucose and insulin in each T1DM subject using the corrected Akaike information criterion (AICc) as implemented in R package `freeknotsplines`. The resulting numbers of optimized knots ranged between 12 and 18 for the glucose component and between 6 and 13 for the insulin component. Using these small numbers of knots resulted in good non-parametric smoothing of observed values, but ODE parameter estimates were not as physiologically plausible as using 30-40 equally spaced knots. Therefore, for each subject, the sets of optimal nodes for the glucose and insulin components were pooled into one set of 15-23 knots (dropping one of the knots from resulting pairs of knots with the distance of less than 5 min). Figure 2 shows the resulting B-spline approximations of the ODE solutions for all subjects. It is clear that the approximate ODE solutions track well the global trends in the observed data. Parameter estimates from the fitted ODE model were used to compute most important physiologically meaningful quantities, for which the ranges were previously described in diabetes research literature. This serves as additional validation step for the models fitted to the real data. The metabolic clearance rate (MCR) $[(min^{-1})mL/kg]$ is the product of the fractional clearance rate and distribution volume in mL of insulin normalized by the subject's weight. The fractional clearance rate and distribution volume are c_1 and

$1000c_2^{-1}$, respectively, where c_1 and c_2 are parameters in (3). Thus, the metabolic clearance rate is computed as $MCR = 1000c_2^{-1}c_1/weight$. The basal insulin infusion rate (r_b) required to maintain euglycemia was calculated to investigate the physiological validity of the parameters in equation (2). By setting to 0 the glucose derivative and glucose appearance from a meal (steady-state conditions), and setting the glucose concentration to the euglycemic value of $G_b = 80$ $[mg/dL]$ in (2), it is possible to compute the basal insulin value required to achieve euglycemic value of 80 mg/dl as $I_b = (b_0 - b_1 G_b)/b_2 G_b$. Then the infusion rate required to achieve I_b is computed from (3) by setting $\dot{I}(t) = 0$ ($r_b = c_2^{-1}c_1 I_b$). Table 1 presents the estimates of parameters in model (2), computed from the optimized generalized profiling estimates of structural parameters in (13), subject weights and computed MCR and r_b. The estimated values of MCR fall within the range ($7.5 - 35.2$ $(min^{-1})mL/kg$) previously reported in the literature (Thorsteinsson, 1990). The computed r_b values range from 0.22 to 0.34 U/hr (3.7 to 5.7 mU/min). These estimates are feasible for T1DM patients with low to normal basal insulin requirements (Scheiner and Boyer, 2005).

6 Simulation study

A small simulation study was conducted to evaluate performance of the proposed optimization for a nonlinear and nonhomogeneous ODE model (2). The input functions used in the model were generated separately according to the study protocol for

IV insulin and Monks (1990) model of glucose appearance in the blood. Then exact numeric solutions were computed for the postulated ODE model (2) with known parameters $\boldsymbol{\theta} = [2.08, -4.60, -7.99, -2.91, 0.08, 2.15, -0.07, 0.39, -0.03]$ (similar to parameters estimated from real data) and input functions. Finally, 100 data sets with $N = 61$ observations each (corresponds to sampling every 10 minutes during 6 hours) were generated according to the measurement error model with exact numeric ODE solution as the mean function and independent Gaussian errors with mean zero and variance 25. The optimized generalized profiling estimation was performed as described in sections 3-4 using B-spline bases with (i) equally spaced knots a priori chosen from the set $\boldsymbol{\Omega} = \{10, 20, 30, 40, 50, 60\}$, (ii) equally spaced knots with the number of basis functions K that minimizes (12) over $K \in \boldsymbol{\Omega}$; (iii) optimized number and location of knots in the range 5-60 as described in section 4. Option (ii) corresponds to optimizing the number of equally spaced knots using a sparse full-search algorithm similar to the one described in Ruppert (2002). Figure 3 shows a sample simulated data set (blue points) with the true ODE solution (blue line), generalized profile solution (red line), and locations of optimized knots (47 knots for glucose and 54 knots for insulin with locations shown by vertical green lines). As expected for optimized knot selection, more knots are placed in the intervals of rapid change of output functions as compared to the intervals with small gradients.

Figure 4 shows the error distributions of the resulting ODE parameter estimates.

The errors corresponding to ODE solutions represented by B-splines with fixed a priori chosen number $K \in \mathbf{\Omega}$ of basis functions are shown with x-coordinate equal to K. The errors corresponding to equally spaced knots and K minimizing (12) over $K \in \mathbf{\Omega}$ are shown with x-coordinate zero. The errors corresponding to optimized number and location of knots are shown with x-coordinate -10. None of the priori chosen values of K yields accuracy of estimating ODE parameters comparable to the accuracy achieved using equally spaced knots with K minimizing (12) over $K \in \mathbf{\Omega}$ or optimized unequally spaced knots. For each θ_i, some values of a priori chosen K yielded similar error distributions, however such values of K are different for different ODE parameters, which does not allow selecting a common fixed optimal K. For majority of θ_i, the optimized generalized profiling with unequally spaced knots yields the smallest estimation errors. Figure 5 shows the distribution of root mean prediction errors of approximated ODE solutions. For each simulated data set, the mean prediction error was computed on the same grid as original data points using the true values $x_j(t_{jl})$, which are known for the simulated data. Optimizing both the number and location of knots provides dramatic reduction in the root mean prediction error of ODE solutions as compared to using an a priori chosen K and substantial reduction as compared to optimizing just the number of equally spaced knots. Also optimizing just the number of equally spaced knots provides large reduction in root mean prediction error of ODE solution as compared to using an a priori chosen K.

In conclusion, optimizing the number and location of knots for B-spline approximating the ODE solution provides the most accurate estimates of the ODE parameters and solution. The improvement well justifies additional computational cost for such optimization. Meanwhile, computational costs of repeated optimized generalized smoothing for multiple numbers of equally spaced knots are substantially higher than optimizing the number and location of knots once.

7 Discussion

In this work, we proposed an approach for optimizing the generalized profiling estimation of models defined by nonlinear and/or nonhomogeneous ODE systems. Our approach includes optimization with respect to the penalty parameters and with respect to the number and location of knots for B-spline basis used to approximate the ODE solution. Additional regularization from optimizing the knots for B-spline basis is especially important in the case of discontinuous input functions. The covariance penalties criterion for outer optimization of penalty weights is equivalent to the generalized cross validation criterion in case of linear prediction rules, but it is applicable to a much wider range of nonlinear and/or nonhomogeneous ODE models that result in nonlinear prediction rules.

Applying the optimized generalized profiling to glucose and insulin concentration data in T1DM patients we obtained physiologically plausible results for the proposed

parsimonious model of glucose-insulin dynamics. Thus, unlike the majority of existing mathematical models of glucose-insulin dynamics, our model is estimable from a relatively small number of noisy observations of glucose and insulin concentrations. This makes the proposed model an attractive candidate for developing automated algorithms of IV insulin delivery for in-hospital glucose management in T1DM patients.

Acknowledgment

We are grateful to the Editor and Reviewers for constructive comments that helped to improve the quality and accessibility of this manuscript. Support for this research was provided by the NIH/NIDDK grant DK081088.

Appendix

Assuming that it is appropriate to integrate under the integral in the penalty function, omitting the limits of integration, and dropping the explicitly written dependence on t, derivatives for Gauss-Newton optimization are

$$\begin{aligned}\frac{\partial \mathbf{J}(\boldsymbol{\alpha}|\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_j} &= -2w_j\boldsymbol{\Psi}_j'\left(\mathbf{y}_j-\boldsymbol{\Psi}_j\boldsymbol{\alpha}_j\right)\\ &\quad+\sum_{k=1}^{d}2\lambda_k\int\left(\dot{\boldsymbol{\psi}}_k'\boldsymbol{\alpha}_k-f_k\left(\widetilde{\mathbf{x}},\mathbf{u},s,\boldsymbol{\theta}\right)\right)\left(\delta_{kj}\dot{\boldsymbol{\psi}}_k-\frac{\partial f_k\left(\widetilde{\mathbf{x}},\mathbf{u},s,\boldsymbol{\theta}\right)}{\partial\boldsymbol{\alpha}_j}\right)ds,\end{aligned}$$

where $\delta_{kj}=1$, if $k=j$, and $\delta_{kj}=0$, if $k\neq j$. Then,

$$\frac{\partial \mathbf{J}(\boldsymbol{\alpha}|\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\lambda_k\partial\boldsymbol{\alpha}_j}=2\int\left(\dot{\boldsymbol{\psi}}_k'\boldsymbol{\alpha}_k-f_k\left(\widetilde{\mathbf{x}},\mathbf{u},s,\boldsymbol{\theta}\right)\right)\left(\delta_{kj}\dot{\boldsymbol{\psi}}_k-\frac{\partial f_k\left(\widetilde{\mathbf{x}},\mathbf{u},s,\boldsymbol{\theta}\right)}{\partial\boldsymbol{\alpha}_j}\right)ds,$$

$$
\begin{aligned}
\frac{\partial^2 \mathbf{J}(\boldsymbol{\alpha}|\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\lambda})}{\partial \boldsymbol{\alpha}_i' \partial \boldsymbol{\alpha}_j} &= \delta_{ij} 2 w_j \boldsymbol{\Psi}_j' \boldsymbol{\Psi}_j \\
&+ \sum_{k=1}^{d} 2\lambda_k \int \left(\delta_{ki} \dot{\boldsymbol{\psi}}_k - \frac{\partial f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_i} \right) \left(\delta_{kj} \dot{\boldsymbol{\psi}}_k - \frac{\partial f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_j} \right)' ds \\
&+ \sum_{k=1}^{d} 2\lambda_k \int \left(\dot{\boldsymbol{\psi}}_k' \boldsymbol{\alpha}_k - f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta}) \right) \left(- \frac{\partial^2 f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_i' \partial \boldsymbol{\alpha}_j} \right) ds,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^3 \mathbf{J}(\boldsymbol{\alpha}_j|\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\lambda})}{\partial \lambda_k \partial \boldsymbol{\alpha}_i' \partial \boldsymbol{\alpha}_j} &= 2 \int \left(\delta_{ki} \dot{\boldsymbol{\psi}}_k - \frac{\partial f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_i} \right) \left(\delta_{kj} \dot{\boldsymbol{\psi}}_k - \frac{\partial f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_j} \right)' ds \\
&+ 2 \int \left(\dot{\boldsymbol{\psi}}_k' \boldsymbol{\alpha}_k - f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta}) \right) \left(- \frac{\partial^2 f_k(\widetilde{\mathbf{x}}, \mathbf{u}, s, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_i' \partial \boldsymbol{\alpha}_j} \right) ds.
\end{aligned}
$$

References

Bergman, R.Y., Ider, Z., Bowden, C., Cobelli, C. (1979) Quantitative estimation of insulin sensitivity. Am J Physiol. 236(6), E667-E676.

Brunel, N.J.B. (2008) Parameter estimation of ODE's via nonparametric estimators. Electron. J. Statist. Volume 2 (2008), 1242-1267.

Cao, J. and Ramsay, J. O. (2009) Generalized Profiling Estimation for Global and Adaptive Penalized Spline Smoothing. Computational Statistics and Data Analysis 53, 2550-2562.

Chen, J. and Wu, H. (2008), Efficient Local Estimation for Time-varying Coefficients in Deterministic Dynamic Models with Applications to HIV-1 Dynamics, Journal of the American Statistical Association, 103, 369-384.

Claeskens, G., Krivobokova, T., & Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. Biometrika, 96(3), 529-544.

Cobelli C., Bier D.M., Ferrannini, E. (1990) Modeling glucose metabolism in man: Theory and practice. Horm Metab Res., 24(S):1-10.

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numer.Math., 31, 377–403.

De Gaetano, A. and Arino, O. (2000) Mathematical Modelling of the Intravenous Glucose Tolerance Test. J. Math. Biol., 40, 136-168.

Donnet, S. and Samson, A. (2007) Estimation of parameters in incomplete data models defined by dynamical systems., Journal of Statistical Planning and Inference, 137(9) :2815–2831.

Efron, B.(2004).The estimation of prediction error: covariance penalties and cross-validation. J.Amer.Statist.Assoc., 99(467): 619–642.

Fang, Y, Wu, H, Zhu, L-X (2011). A two-stage estimation method for random coefficient differential equation models with application to longitudinal HIV dynamic data. Vol. 21, No. 3, 1145-1170.

Fischer, U., Salzsieder, E., Jutzi, E., Albrecht, G., Freyse, E.J. (1984) Modeling the

glucose-insulin system as a basis for the artificial beta cell. Biomed Biochim Acta., 43(5):597-605.

Heckman, N. and Ramsay, J. (2000) Penalized regression with model based penalties. Can J Stat 28:241–258.

Hipszer B., Joseph J., Kam M. (2005) Pharmacokinetics of intravenous insulin delivery in humans with type 1 diabetes. Diabetes Technol Ther., 7(1) :83-93.

Hu, Y. (1993). An algorithm for data reduction using splines with free knots. IMA journal of numerical analysis, 13(3), 365-381.

Huang, Y., Liu, D. and Wu, H. (2006), Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System, Biometrics, 62, 413-423.

Jupp, D. L. (1978). Approximation to data by splines with free knots. SIAM Journal on Numerical Analysis, 15(2), 328-343.

Lehmann, E.D. (1997). "Interactive educational simulators in diabetes care". Med Inform (Lond) 22 (1): 47–76.

Lehmann, E.D., Deutsch, T. A (1992) A physiological model of glucose-insulin interaction in T1DM mellitus. J Biomed Eng., 14(3):235-42.

Li, Z., Osborne, M.R. and Pravan, T. (2005) Parameter estimation of ordinary differential equations. IMAJ. Numer. Anal. 25, 264–285.

Lindstrom, M. J. (1999). Penalized estimation of free-knot splines. Journal of Computational and Graphical Statistics, 8(2), 333-352.

Monks J. (1990) The construction of diabetic diets using linear programming. MSc Thesis, University of Salford.

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

Ruppert, D (2002) Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics 11,735–757.

Scheiner G and Boyer BA. (2005) Characteristics of basal insulin requirements by age and gender in Type-1 diabetes patients using insulin pump therapy. Diabetes Res Clin Pract., 69(1), 14-21.

Sherwin R, Kramer K, Tobin J, Insel P, Liljenquist, Berman M, Andres R. (1974) A model of the kinetics of insulin in man. J Clin Invest., 53:1481-1492.

Shimoda S, Nishida K, Sakakida M, Konno Y, Ichinose K, Uehara M, Nowak T, Shichiri M. (1997) Closed-loop subcutaneous insulin infusion algorithm with a short-acting insulin analog for long-term clinical application of a wearable artificial endocrine pancreas. Front Med Biol Eng., 8(3), 197-211.

Sorensen J.T. (1985) A Physiological Model of Glucose Metabolism in Man and its Use to Design and Assess Improved Insulin Therapies for Diabetes. PhD Thesis, Massachusetts Institute of Technology.

Spiriti, S., Eubank, R., Smith, P. W., & Young, D. (2013). Knot selection for least-squares and penalized splines. Journal of Statistical Computation and Simulation, 83(6), 1020-1036.

Steil G, Rebrin K, Mastrototaro J, Bernaba B, Saad M. (2003) Determination of plasama glucose during rapid glucose excursions with a subcutaneous glucose sensor. Diabetes Tech. & Ther., 5(1):27-31.

Stein, C. (1981) Estimation of the mean of a multivariate normal distribution . The Annals of Statistics, 9, 1135-1151.

Trajanoski Z, Wach P, Kotanko P, Ott A, Skraba F. (1993) Pharmacokinetic model for the absorption of subcutaneously injected soluble insulin and monomeric insulin analogues. Biomed Tech (Berl)., 38(9):224-31.

Thorsteinsson B. (1990) Kinetic models for insulin disappearance from plasma in man. Dan Med Bull., 37(2):143–153.

Varah, J. M. (1982) A spline least squares method for numerical parameter estimation in differential equations., SIAM J.sci. Stat. Comput., 3(1):28–46

Wilinska ME, Chassin LJ, Schaller HC, Schaupp L, Pieber TR, Hovorka R. (2005) Insulin kinetics in type-I diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Trans Biomed Eng., 52(1), 3-12.

Worthington DR. (1997) Minimal model of food absorption in the gut. Med Informatics., 22:35-45.

Wu, H., Zhu, H., Miao, H., and Perelson, A.S. (2008), Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models, Bulletin of Mathematical Biology, 70(3), 785-799.

Xue, H., Miao, H., Wu, H. (2010), Sieve Estimation of Constant and Time-Varying Coefficients in Nonlinear Ordinary Differential Equation Models by Considering Both Numerical Error and Measurement Error, Annals of Statistics, 38, 2351-2387.

Ye, J. (1998) On measuring and correcting the effects of data mining and model selection. Journal of the American Statistical Association, 93, 120-131.

Table 1 ODE parameter estimates and corresponding derived physiological parameters, metabolic clearance rate (MCR) and basal insulin infusion rate (r_b), for all T1DM subjects.

Subj.	b_0 [(mg/dl)min^{-1}]	b_1 [min^{-1}]	b_2 [min^{-1}(L/mU)]	c_1 [min^{-1}]	c_2 [L^{-1}]	Weight [kg]	MCR [min^{-1}mL/kg]	r_b [mU/min]	r_b [U/h]
1	0.95	0.0011	0.00020	0.054	0.035	73.5	20.9	4.97	0.30
2	0.46	0.0010	0.00009	0.253	0.166	65.8	23.2	4.75	0.29
3	1.94	0.0010	0.00020	0.030	0.036	61.3	13.4	5.72	0.34
4	1.24	0.0049	0.00030	0.106	0.060	74.9	23.5	3.71	0.22

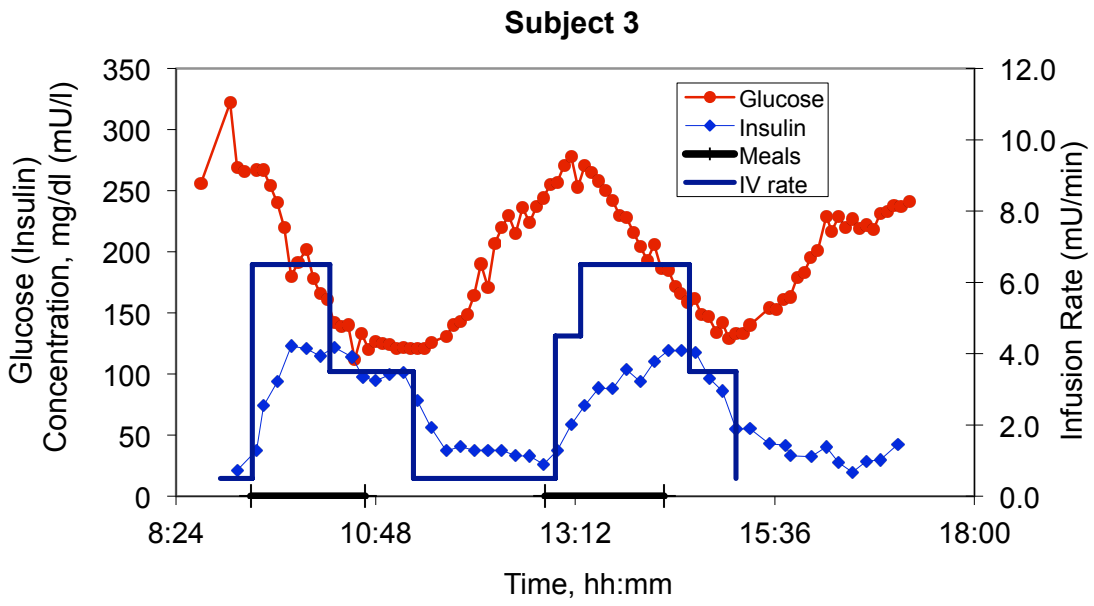

Figure 1: Glucose concentration, insulin concentration, and intravenous insulin infusion rates and meal times as a function of study time in one T1DM subject.

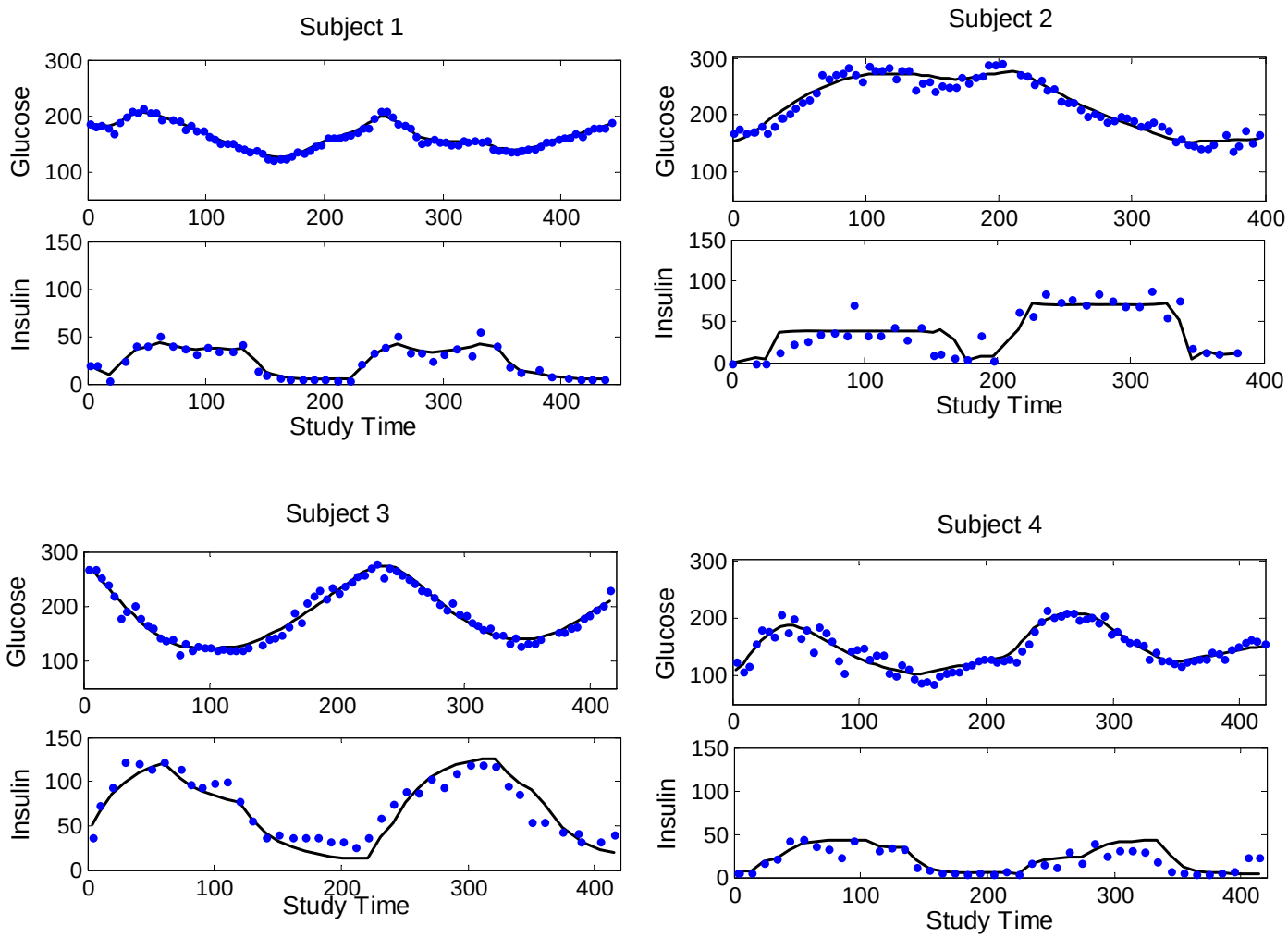

Figure 2: Fitted ODE solutions for 4 T1DM subjects

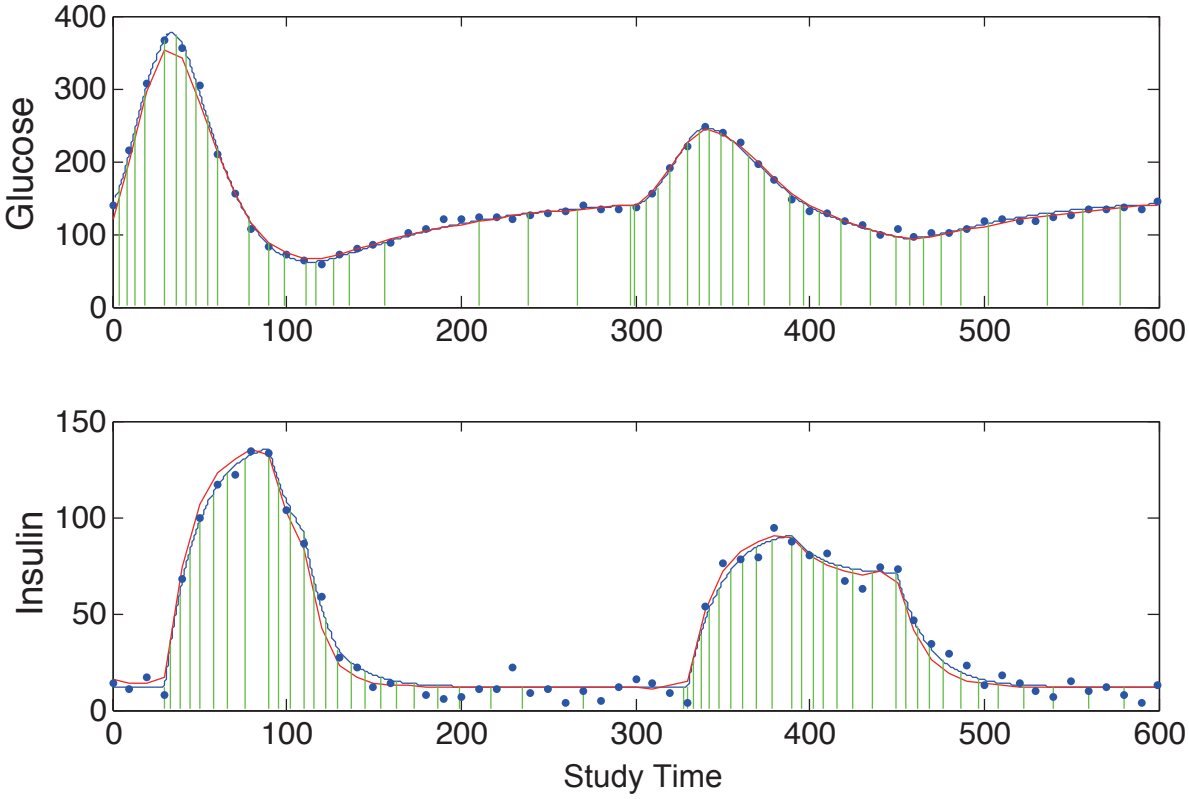

Figure 3: Sample simulated data (blue points) with the true ODE solution (blue line) and generalized profile solution (red line). Locations of optimized knots for B-spline basis shown as vertical green lines (47 knots for glucose and 54 knots for insulin).

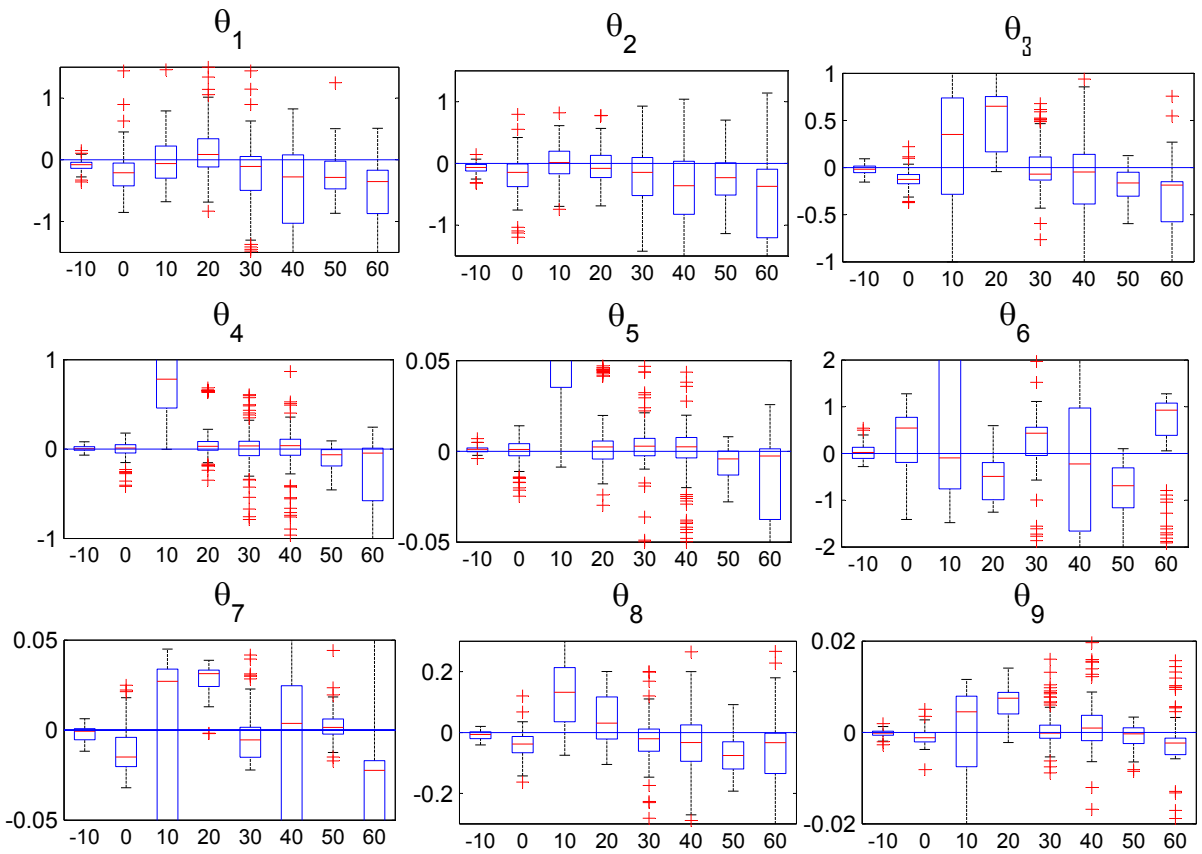

Figure 4: Error distribution of ODE parameter estimates. The errors corresponding to ODE solutions represented by B-splines with fixed a priori chosen number K of basis functions (K=10, 20, 30, 40, 50, 60) are shown with x-coordinate equal to K. The errors corresponding to ODE solutions represented by B-splines with equally spaced knots and the number of basis functions K that minimizes the outer optimization criterion over the grid {10, 20, 30, 40, 50, 60} are shown with x-coordinate zero. The errors corresponding to ODE solutions using B-splines with optimized number and location of knots are shown with x-coordinate -10.

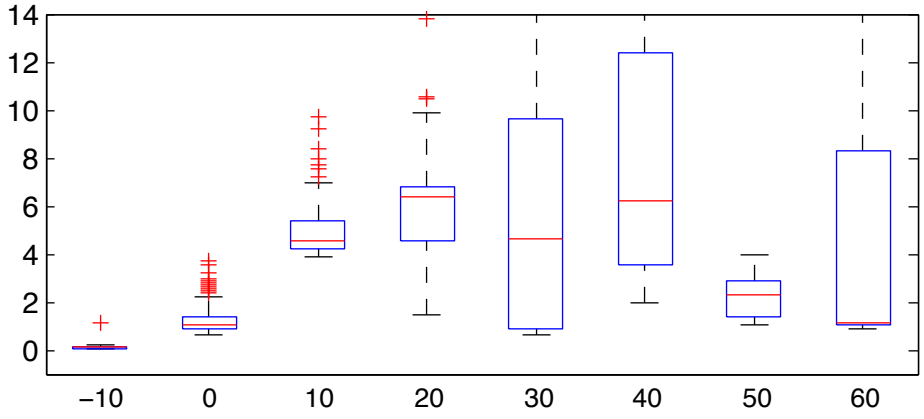

Figure 5: Distribution of root mean prediction error of ODE solutions estimated on the same grid as original data points. The prediction errors of ODE solutions represented by B-splines with fixed a priori chosen number K of basis functions (K=10, 20, 30, 40, 50, 60) are shown with x-coordinate equal to K. The prediction errors of ODE solutions represented by B-splines with equally spaced knots and the number of basis functions K that minimizes the outer optimization criterion over the grid {10, 20, 30, 40, 50, 60} are shown with x-coordinate zero. The prediction errors of ODE solutions using B-splines with optimized number and location of knots are shown with x-coordinate -10.