INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2012/17

Bayesian penalized smoothing approaches in models specified using affine differential equations with unknown error distributions

JAEGER, J. and Ph. LAMBERT

Bayesian penalized smoothing approaches in models specified using affine differential equations with unknown error distributions

Jonathan Jaeger[1] Philippe Lambert[1,2]

[1]Institut de Statistique, Biostatistique et Sciences Actuarielles,
Université Catholique de Louvain, Belgium

[2]Institut des Sciences Humaines et Sociales,
Méthodes Quantitatives en Sciences Sociales, Université de Liège, Belgium

Abstract. A full Bayesian approach based on ODE-penalized B-splines and penalized Gaussian mixture is proposed to jointly estimate ODE-parameters, state function and error distribution from the observation of some state functions involved in systems of affine differential equations. Simulations inspired by pharmacokinetic studies show that the proposed method provides comparable results to the method based on the standard ODE-penalized B-spline approach (i.e. with the Gaussian error distribution assumption) and outperforms the standard ODE-penalized B-splines when the distribution is not Gaussian. This methodology is illustrated on the Theophylline dataset.

Keywords: ODE-Penalized-B-spline, Ordinary differential equations, Penalized Gaussian mixture

1 Introduction

Ordinary differential equations (ODEs) are frequently used to model physical, chemical and biological processes.

Currently, the most commonly used estimation procedures rely on nonlinear least squares (NLS). It uses minimization techniques for the estimation of the ODE parameters and a numerical solver for the approximation of the solution. These approaches can be computationally intensive and sometimes poorly suited for statistical inference. Lunn et al. (2002) propose a Bayesian framework for these NLS procedures. The possible use of prior information about the ODE parameters is a definite advantage, but the numerical approximation to the solution leads to a posterior distribution with no-closed form for the ODE parameters.

Alternative estimation methods of the state functions and of the ODE parameters were proposed in Ramsay et al. (2007). It may be viewed as a generalization of the P-spline theory (Eilers and Marx 1996) where the unknown state functions are expressed as linear combinations of a large number of B-splines with the flexibility conter-balanced by a penalty term defined using the set of differential equations. Jaeger and Lambert (2011) adapted this approach to a full Bayesian ODE-penalized B-spline setting when the ODEs are affine and the conditional distribution of the response is assumed Gaussian. The two major drawbacks of the frequentist ODE-penalized smoothing approach are overcome in the Bayesian framework: the selection of the ODE-adhesion parameter is now automatic

and uncertainty measures about parameters can simply be quantified using MCMC. In addition, the Gaussian error density assumption enables to marginalize the joint posterior distribution with respect to the spline coefficients and therefore to get rid of the inconvenient posterior correlation between the spline coefficients and the ODE parameters.
The assumption of a Gaussian distribution for the response is often inappropriate in practice. Lindsey et al. (2000) has shown in the simultaneous modelling of flosequinan concentration and its metabolite concentration that the choice of a parametric error distribution is at least as important as the choice of the pharmacokinetic (PK) compartment model. However, these parametric distributions are restrictive by their definition.
The goal of this paper is to generalize the standard Bayesian ODE-penalized B-spline approach (Jaeger and Lambert 2011) by assuming a penalized Gaussian mixtures for the error distributions assumption.
The plan of the paper is as follows. Section 2 reminds the key ingredients of the standard ODE-penalized B-spline approach. Section 3 describes how this standard Bayesian ODE-penalized B-spline model can be extended to a non Gaussian setting. Section 4 explains how to explore efficiently the joint posterior distribution using MCMC. The results of a simulation study are presented in Section 5 to asses and compare the performances of the proposed approach with those of the traditional method. After an application on pharmacokinetic data in Section 6, we conclude the paper by a discussion in Section 7.

2 Basics on ODE-penalized B-spline approach

In this section, we recall the key ingredients in the standard ODE-penalized B-spline approach. Assume that changes in state functions $x(t) \in R^d$ are governed by a set of differential equations:

$$Dx(t) = f(x, t, \theta), t \in [0; T] \tag{1}$$

where f is a known affine function with respect to x and $\theta \in R^q$ a vector of unknown parameters. It is assumed that only a subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state functions are measured at time point t_{jk}, $j \in \mathcal{J}$, $k = 1, \ldots, n_j$ with additive measurement error ϵ_{jk}. We denote by $y_{jk} = x_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$ the corresponding measurement ($j \in \mathcal{J}$). The objective is to jointly estimate the ODE parameters θ, the state functions $x(t)$ and $\tau = \mathrm{vec}(\tau_j, j \in \mathcal{J})$ from $\{(t_{jk}, y_{jk}), j \in \mathcal{J}, k = 1, \ldots, n_j\}$. Most often, one further imposes that $E(\epsilon_j) = 0$ and $V(\epsilon_j) = 1$ such that $x_j(t)$ and τ_j are respectively the conditional mean and the inverse-variance of y_j.
The solution of the system of differential equations is approximated using a B-spline basis function expansion. Denote by $B_j(t)$ the K_j-vector of B-spline basis functions of order p evaluated at time t that is used to approximate the jth component of the state function $x(t)$ and c_j the K_j-vector of B-spline coefficients. The approximation $\widetilde{x_j}(t)$ of $x_j(t)$ is expressed as a linear combination of these B-spline basis functions:

$$\begin{aligned} \widetilde{x_j}(t) &= \sum_{k=1}^{K_j} B_{jk}(t) c_{jk} \\ &= (B_j(t))^T c_j. \end{aligned}$$

Details on the specific choice of the B-spline basis and its properties are given in Jaeger and Lambert (2011). The proximity on $[0;T]$ between the approximation to the state function, $\widetilde{x_j}(t)$, and the jth component of the solution $x(t)$ to the set of differential equations in (1) can be assessed by a penalty term:

$$\begin{aligned} PEN_j(\widetilde{x}) &= \int \left\{L_{j,\theta}\left(\widetilde{x}(t)\right)\right\}^2 dt \\ &= \int \left\{D\widetilde{x_j}(t) - f_j\left(\widetilde{x},t,\theta\right)\right\}^2 dt, \end{aligned}$$

where the integration is over the interval $[0;T]$ (Varah 1982; Ramsay and Silverman 2005; Ramsay et al. 2007; Cao and Zhao 2008). Note that for the system of affine differential equations that we consider here, the jth penalty term is an homogenous polynomial of degree 2 in the B-spline coefficients. The full fidelity-to-ODE measure is then given by:

$$\begin{aligned} PEN(\widetilde{x}|\gamma) &= \sum_{j=1}^{d} \gamma_j PEN_j(\widetilde{x}) \\ &= c^T R(\theta,\gamma)\, c + 2c^T r(\theta,\gamma) + l(\theta,\gamma), \end{aligned} \tag{2}$$

where γ is the vector of ODE-adhesion parameters, $c = \left(c_1^T,\ldots,c_d^T\right)^T$ is the vector of all spline coefficients of length $K = \sum_{j=1}^{d} K_j$, $R(\theta,\gamma)$ is a $K \times K$ block-matrix constructed by placing the corresponding penalty matrices involved in each $PEN_j(\tilde{x})$, $r(\theta,\gamma)$ is a vector of length K corresponding to a penalty vector and $l(\theta,\gamma)$ is a penalty constant (see Section 2.4 in Jaeger and Lambert (2011) for the automatic construction of these quantities).

3 Model

In this section, we give the model specification, i.e. the distributional assumption for the error distribution using Bayesian penalized Gaussian mixture (Komárek and Lesaffre 2008) and all the prior distributions necessary for the Bayesian ODE-penalized B-spline approach.

3.1 Assumption on the error distribution

We recall that only a subset $\mathcal{J} \subset \{1,\ldots,d\}$ of the d state functions are measured at time point t_{jk}, $j \in \mathcal{J}$, $k = 1,\ldots,n_j$ with additive measurement error:

$$y_{jk} = x_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}.$$

In order to have a flexible distributional assumption, we model the error density using a penalized Gaussian mixture. The density $g_j(\epsilon_j)$ of the error term associated to the jth state function is modeled using a weighted sum of Gaussian densities with respective means taken on a fixed predefined grid $\mu_j = \mathrm{vec}(\mu_{jl}; l = -L_j,\ldots,L_j)$ centered around $\mu_{j0} = 0$, and a common variance σ_j^2. Therefore,

$$\epsilon_{jk}|\pi_j \sim \sum_{l=-L_j}^{L_j} \pi_{jl}\mathcal{N}\left(\mu_{jl},\sigma_j^2\right), \tag{3}$$

with unknown positive weights π_{jl} summing to one for all $j \in \mathcal{J}$. For identifiability reasons, we suppose that for all $j \in \mathcal{J}$ and $k = 1, \ldots, n_j$:

$$E\left(\epsilon_{jk}|\pi_j\right) = 0$$

and

$$V\left(\epsilon_{jk}|\pi_j\right) = 1.$$

Komárek and Lesaffre (2008) propose to work with $\mu_{j,-L_j} = -5$ and $\mu_{jL_j} = 5$ and place the other μ_{jl} at equidistant positions between -5 and 5 (with $L_j \geq 10$). In practice, we experienced robustness issues with density estimates sensitive to extreme values so we recommend to work with a wider range of value (e.g $\mu_{j,-L_j} = -10$ and $\mu_{jL_j} = 10$ with $L_j \geq 20$). For σ_j^2, the value should be less than one (see Section 3.4 for details): in practice, one can take $\sigma_j = \frac{2}{3}\delta_j$ where δ_j is the distance between two adjacent means. With this choice, at most five means are within the 0.5% and 99.5% quantile of each Gaussian component. Note also that with that specification, label switching (Stephens 2000) does not occur.
Note that the sum to one constraint, $\sum_{l=-L_j}^{L_j} \pi_{jl} = 1$ and $\pi_{jl} > 0$ ($\forall j \in \mathcal{J}$, $\forall l \in \{-L_j, \ldots, L_j\}$) is automatically checked with the generalized logit transformation:

$$\pi_{jl} = \frac{\exp\left(a_{jl}\right)}{\sum_{k=-L_j}^{L_j} \exp\left(a_{jk}\right)}, j \in \mathcal{J}, l = -L_j, \ldots, L_j$$

where constraints $a_{jL_j} = 0$ for all $j \in \mathcal{J}$ have to be used for identifiability.

3.2 Complete likelihood function

We introduce the group label variable z_{jk} for the kth observation of the jth state function (Tanner and Wong 1987; Diebolt and Robert 1994; Richardson and Green 1997). The z_{jk} are supposed independently drawn from the distribution:

$$P\left(z_{jk} = l|\pi_j\right) = \pi_{jl}$$

with $j \in \mathcal{J}$, $l \in \{-L_j, \ldots, L_j\}$ and $k = 1, \ldots, n_j$. Given the group label variable z_{jk}, one has:

$$\epsilon_{jk}|z_{jk}, \pi_j \sim \mathcal{N}\left(\mu_{jz_{jk}}, \sigma_j^2\right).$$

The complete likelihood function is therefore equal to

$$\begin{aligned} L\left(y, z|c, \pi, \tau\right) &= \prod_{j\in\mathcal{J}} \left\{ \prod_{k=1}^{n_j} \tau_j^{1/2} \phi\left(\tau_j^{1/2}\left(y_{jk} - \left(B_j\left(t_{jk}\right)\right)^T c_j\right); \mu_{jz_{jk}}, \sigma_j^2\right) \pi_{jz_{jk}} \right\} \\ &= \prod_{j\in\mathcal{J}} \left\{ \tau_j^{n_j/2} \prod_{l=-L_j}^{L_j} \left[\pi_{jl}^{r_{jl}} \prod_{k:z_{jk}=l} \phi\left(\tau_j^{1/2}\left(y_{jk} - \left(B_j\left(t_{jk}\right)\right)^T c_j\right); \mu_{jl}, \sigma_j^2\right) \right] \right\}, \end{aligned}$$

where $\phi\left(\cdot; \mu_{jl}, \sigma_j^2\right)$ is the density function of a Gaussian distribution with mean μ_{jl} and variance σ_j^2 and r_{jl} is the number of indices $k \in \{1, \ldots, n_j\}$ for which $z_{jk} = l$; $l = -L_j, \ldots, L_j$.

3.3 Prior for the transformed weight a

In order to allow flexibility in the specification of the error distribution, Komárek and Lesaffre (2008) propose to consider a large number of components in the Gaussian mixture. The overfitting of the data is then prevented by constraining the transformed mixture weights using a s-order difference penalty:

$$\begin{aligned} Q_j\left(a_j|\lambda_j\right) &= -\frac{\lambda_j}{2}\sum_{l=-L_j+s}^{L_j}\left(\Delta^s a_{jl}\right)^2 \\ &= -\frac{\lambda_j}{2}a_j^T D_{a,j}^T D_{a,j} a_j, \end{aligned}$$

for all $j \in \mathcal{J}$, where Δ^s denotes the difference operator of order s. We suggest, as in Komárek and Lesaffre (2008), to set s equal to 3. The vector of roughness penalty parameters $\lambda = \text{vec}\left(\lambda_j; j \in \mathcal{J}\right)$ controls the smoothness of each density and has to be selected. In Bayesian terms, these penalties can be translated into a prior distribution for each a_j:

$$p\left(a_j|\lambda_j\right) \propto \exp\left(-\frac{\lambda_j}{2}a_j^T D_{a,j}^T D_{a,j} a_j\right).$$

As a_{jL_j} is set to 0 for identifiability constraints, the prior distribution $p\left(a_j|\lambda_j\right)$ is translated into a prior distribution for $d_j = \text{vec}\left(a_{jk}; k = -L_j, \ldots, L_j - 1\right)$ such that

$$p\left(d_j|\lambda_j\right) \propto \lambda_j^{\frac{2L_j+1-s}{2}} \exp\left(-\frac{\lambda_j}{2}d_j^T P_{d,j} d_j\right),$$

where $P_{d,j}$ corresponds to the penalty matrix $D_{a,j}^T D_{a,j}$ without its last line and its last column. Thus, the prior distribution for d_j corresponds to a multivariate Gaussian distribution with mean 0 and covariance matrix $\left(\lambda_j P_{d,j}\right)^{-1}$.

3.4 Identification penalty

To force a zero mean and an unit variance for the error term, we introduce correction terms to the mean and to the variance of the Gaussian components of the mixture:

$$\epsilon_{jk}|d_j \sim \sum_{l=-L_j}^{L_j} \pi_{jl}\left(d_j\right)\mathcal{N}\left(\mu_{jl} + \alpha_j, \frac{\sigma_j^2}{\beta_j^2}\right)$$

where

$$\begin{aligned} \alpha_j &= -\sum_{l=-L_j}^{L_j} \pi_{jl}\mu_{jl}, \\ \beta_j^2 &= \frac{\sigma_j^2}{1 - \sum_{l=-L_j}^{L_j} \pi_{jl}\mu_{jl}^2 + \left(\sum_{l=-L_j}^{L_j} \pi_{jl}\mu_{jl}\right)^2}. \end{aligned}$$

It appears that the conditional posterior distributions for the transformed weights are not affected by these changes. In order to force the mean correction term α_j to be closed to zero and the variance correction term β_j to be closed to one, we adapt the suggestion in Lambert (2011) by adding a frequentist identification penalty in the model specification:

$$pen_{id,j} = -\kappa \left(\alpha_j^2 + \left(\beta_j^2 - 1 \right)^2 \right). \tag{4}$$

It forces the mean to be 0 and the variance to be equal to 1 when κ tends to infinity.

3.5 Prior distribution for the spline coefficients

The frequentist ODE-penalty term $PEN\left(\widetilde{x}|\gamma\right)$ (given in Eq. (2)) is translated, in a Bayesian framework, into a prior distribution for the spline coefficients:

$$p\left(c|\theta,\gamma\right) \propto \exp\left(-\frac{1}{2} PEN\left(\widetilde{x}|\gamma\right) - \frac{1}{2}\left(c-\mu_c\right)^T \Sigma_c^{-1}\left(c-\mu_c\right) \right).$$

The vector μ_c and the matrix Σ_c^{-1} are used to translate the information available about the initial condition of the state function x in the model. The prior density corresponds to a multivariate normal distribution with mean vector $V_1^{-1}v_1$ and covariance matrix V_1^{-1} where $v_1 = -r\left(\theta,\gamma\right) + \Sigma_c^{-1}\mu_c$ and $V_1 = R\left(\theta,\gamma\right) + \Sigma_c^{-1}$. Therefore, the normalizing constant for the conditional prior density of the spline coefficients c is:

$$\left(\det\left(V_1\right)\right)^{\frac{1}{2}} \exp\left(\frac{1}{2} l\left(\theta,\gamma\right)\right) \exp\left(-\frac{1}{2} v_1^T V_1^{-1} v_1\right).$$

Jaeger and Lambert (2011) have shown the absolute necessity to include that normalizing constant in the derivarion of the joint posterior of the ODE parameters.

3.6 Prior distributions for τ, λ, γ and θ

For the conditional precision τ_j $(j \in \mathcal{J})$ of the vector of response y_j, each roughness penalty parameter λ_j $(j \in \mathcal{J})$, as well as each ODE-adhesion parameter γ_j $(j \in \{1,\ldots,d\})$, gamma prior distribution are convenient choices:

$$\tau_j \sim \mathcal{G}\left(a_{\tau_j}, b_{\tau_j}\right), \lambda_j \sim \mathcal{G}\left(a_{\lambda_j}, b_{\lambda_j}\right), \gamma_j \sim \mathcal{G}\left(a_{\gamma_j}, b_{\gamma_j}\right),$$

where $\mathcal{G}\left(a,b\right)$ denotes the gamma distribution with mean a/b and variance a/b^2. As recommanded in Lang and Brezger (2004) for standard P-spline model, we have two possibilities: either set a equal to 1 and b equal to a small quantity or set $a = b$ equal to a small quantity. We will opt for the first specification as the corresponding density is finite at 0 (see Jullion and Lambert (2007) for alternatives). This choice for the prior distributions of the ODE-adhesion parameters translates our prior confidence in the specification of the system of differential equations as descriptor of the dynamics of the state function. Indeed, such a prior for γ_j is rather flat although it puts slightly more weight on value of $\log_{10}\left(\gamma_j\right)$ around $-\log_{10}\left(b_{\gamma_j}\right)$. For the roughness penalty parameters λ, the suggested prior distributions indicate our slight prior preference for the Gaussian distribution

as the error distribution in the absence of counter-indications in the data.
For the vector θ of differential equation parameters, the chosen prior will depends on the context. Let us denote this prior distribution by $p(\theta)$.

4 Posterior distribution and parameter estimation

In this section, we explain how to obtain and to generate samples from the joint posterior distribution.

4.1 Joint posterior distribution and conditional posterior distributions

Using the complete likelihood function and all the prior distributions, one obtains the log joint posterior density:

$$
\begin{aligned}
& \log\left(p\left(c,\gamma,\tau,d,\lambda,\theta|y,z\right)\right) \\
= & \sum_{j\in\mathcal{J}}\left\{\frac{n_j}{2}\log\left(\tau_j\right)+\sum_{l=-L_j}^{L_j} r_{jl}\log\left(\pi_{jl}\right)-\frac{\beta_j^2}{2\sigma_j^2}\left(\tau_j\sum_{k=1}^{n_j} y_{jk}^2-2\tau_j^{1/2}\sum_{k=l}^{n_j}\left(\mu_{jz_{jk}}+\alpha_j\right)y_{jk}\right)\right\} \\
& -\frac{1}{2}c^T\,\mathrm{diag}\left(H_j;j=1,\ldots,d\right)c+c^T\,\mathrm{vec}\left(h_j;j=1,\ldots,d\right) \\
& +\sum_{j\in\mathcal{J}}\left\{\frac{2L_j+1-s}{2}\log\left(\lambda_j\right)-\frac{\lambda_j}{2}d_j^T P_{d,j}d_j\right\} \\
& -\sum_{j\in\mathcal{J}}\kappa\left(\alpha_j^2+\left(\beta_j^2-1\right)^2\right) \\
& +\sum_{j\in\mathcal{J}}\left\{\left(a_{\lambda_j}-1\right)\log\left(\lambda_j\right)-b_{\lambda_j}\lambda_j\right\} \\
& +\sum_{j\in\mathcal{J}}\left\{\left(a_{\tau_j}-1\right)\log\left(\tau_j\right)-b_{\tau_j}\tau_j\right\} \\
& -\frac{1}{2}\left\{c^T V_1 c-2c^T v_1+v_1^T V_1^{-1}v_1-\log\left(\det\left(V_1\right)\right)\right\} \\
& +\sum_{j=1}^{d}\left\{\left(a_{\gamma_j}-1\right)\log\left(\gamma_j\right)-b_{\gamma_j}\gamma_j\right\} \\
& +\log\left(p\left(\theta\right)\right)
\end{aligned}
$$

where $\mathrm{diag}\left(H_j, j=1,\ldots,d\right)$ corresponds to a block-diagonal matrix where H_j that is equal to

$$\tau_j\frac{\beta_j^2}{\sigma_j^2}B_j^T B_j$$

if the jth state function is observed and the null $K_j \times K_j$-matrix otherwise. The vector $\text{vec}\,(h_j, j = 1, \ldots, d)$ corresponds to the concatenation of vector h_j that is equal to

$$\tau_j \frac{\beta_j^2}{\sigma_j^2} B_j^T y_j - \tau_j^{1/2} \frac{\beta_j^2}{\sigma_j^2} \sum_{k=1}^{n_j} \left(\mu_{j z_{jk}} + \alpha_j\right) B_j\left(t_{jk}\right)$$

if the jth state function is observed and the null K_j-vector otherwise.

4.2 Conditional posterior distributions

The conditional posterior distribution for each roughness penalty parameter λ_j can be shown to be:

$$\lambda_j | d_j, y, z \sim \mathcal{G}\left(\frac{2L_j + 1 - s}{2} + a_{\lambda_j}, \frac{d_j^T P_{d,j} d_j}{2} + b_{\lambda_j}\right).$$

The conditional posterior distribution for the spline coefficients is a multivariate Gaussian distribution with mean $V_2^{-1} v_2$ and variance-covariance matrix V_2^{-1} where:

$$v_2 = v_1 + \text{vec}\,(h_j; j = 1, \ldots, d)$$

and

$$V_2 = V_1 + \text{diag}\,(H_j; j = 1, \ldots, d).$$

The other conditional posteriors are unfortunately not of a familiar type. Therefore, the Metropolis-within-Gibbs algorithm can be used to sample from the joint posterior. As in Jaeger and Lambert (2011), the joint posterior distribution can be marginalized with respect to the spline coefficients in order to avoid being forced to deal with the strong posterior conditional correlation between the spline coefficients and the differential equation parameters. This is only feasable because we are working with mixture of Gaussian distributions. The joint posterior distribution can also be marginalized with respect to λ. Finally, one can show that the conditional posterior probability for the group label variable is proportional to:

$$\begin{aligned} P\left(z_{jk} = l | y_{jk}, c_j, \tau_j, \pi_j\right) &\propto \phi\left(\tau_j^{1/2}\left(y_{jk} - \left(B_j\left(t_{jk}\right)\right)^T c_j\right); \mu_{jl}, \sigma_j^2\right) P\left(z_{jk} = l | \pi_j\right) \\ &\propto \phi\left(\tau_j^{1/2}\left(y_{jk} - \left(B_j\left(t_{jk}\right)\right)^T c_j\right); \mu_{jl}, \sigma_j^2\right) \pi_{jl}. \end{aligned}$$

Therefore, the conditional posterior distribution for the group label variable z_{jk} is a multinomial distribution with cells probabilities equal to:

$$P\left(z_{jk} = l | y_{jk}, c_j, \tau_j, \pi_j\right) = \frac{\phi\left(\tau_j^{1/2}\left(y_{jk} - \left(B_j\left(t_{jk}\right)\right)^T c_j\right); \mu_{jl}, \sigma_j^2\right) \pi_{jl}}{\sum_{m=-L_j}^{L_j} \phi\left(\tau_j^{1/2}\left(y_{jk} - \left(B_j\left(t_{jk}\right)\right)^T c_j\right); \mu_{jm}, \sigma_j^2\right) \pi_{jl}}.$$

4.3 MCMC algorithm

We use Markov Chain Monte Carlo (MCMC) techniques to generate samples from posterior distributions.

MCMC algorithm

We use a Metropolis-within-Gibbs algorithm. Each component of the transformed weights d_j, $j \in \mathcal{J}$ is updated using univariate Metropolis steps with a Gaussian proposal distribution. Gibbs steps are used to update the allocation variables z, the penalty parameters λ and the spline coefficients c. Last of all, univariate Metropolis steps are used for all the component of the precision parameters τ, the ODE-adhesion parameters γ and the ODE parameters θ. The standard deviation of each Gaussian proposal distribution is tuned during the burnin phase to ensure a 40% acceptance rate (Haario et al. 2001; Lambert and Eilers 2009).

Starting value

The starting value for the ODE parameters, the spline coefficients and the precision parameters are obtained using a standard approach (with a Gaussian distribution for the error term distribution). All components of the transformed weights d_j, $j \in \mathcal{J}$ can either be set to zero (unefficient starting values) or set to the transformed mixture weights corresponding to a standard Gaussian density.

Re-parametrization of θ and d_j using rotation and translation

To accelerate the inference procedure, we also used the strategy proposed by Lambert (2007).
First run the Metropolis-within-Gibbs algorithm for a few thousand iterations. Then, reparametrize the problem by applying a translation and a rotation to the parameter vector θ and for all vector d_j, $j \in \mathcal{J}$.
More precisely, denote by S_θ (resp. S_{d_j}) the empirical variance-covariance matrix of the parameter θ (resp. d_j) evaluated using a first set of iterations of the generated chains and $\bar{\theta}$ (resp. $\bar{d_j}$) the corresponding empirical mean vector. We reparametrize the posterior distribution using β_θ (resp. β_{d_j}),

$$\begin{aligned} \theta &= L_\theta \beta_\theta + \bar{\theta} \\ d_j &= L_{d_j} \beta_{d_j} + \bar{d_j} \end{aligned}$$

where L_θ (resp. L_{d_j}) is the lower triangular matrix in the Choleski decomposition, i.e., the matrix L_θ (resp. L_{d_j}) such that $S_\theta = L_\theta L_\theta^T$ (resp. $S_{d_j} = L_{d_j} L_{d_j}^T$). The Metropolis-Hastings algorithm described above is then used to sample from the so reparameterized joint posterior distribution. A sample of posterior distribution for θ and all d_j, $j \in \mathcal{J}$ are obtained using the previous formulas.

5 Simulation study

A simulation study was carried out to validate and to compare the performances of the proposed Bayesian ODE-penalized B-spline approach (assuming a penalized Gaussian mixture as error distribution) with the existing one (assuming a Gaussian distribution as error distribution) in the estimation of the ODE parameters and of the error precision.

5.1 Scenario

Let us consider the system of two differential equations:

$$\begin{cases} \frac{x_1(t)}{dt} &= -k_a x_1(t), \\ \frac{x_2(t)}{dt} &= \frac{k_a}{V} x_1(t) - k_e x_2(t), \\ x_1(0) &= D, \\ x_2(0) &= 0, \end{cases}$$

For this model, we are interested by the estimation of (k_a, k_e, V, τ) from a set of observations $\{(t_k, y_k) : k = 1, \ldots, n\}$ where $y_k = x_2(t_k) + \tau^{-1/2}\epsilon_k$. The sampling times correspond to n equistant time points between 0 and 2 with four sample sizes: $n = 50, 100, 200, 500$. The error terms were generated from four distributions: a standard Gaussian, a Student t_6, a standardized extreme value and a Gaussian mixture $0.4\mathcal{N}\left(-1.5, 0.9^2\right) + 0.6\mathcal{N}\left(1, 0.6^2\right)$. Each distribution is standardized to have zero mean and unit variance. For this simulation study, only the state function $x_2(t)$ is observed. For each configuration, 100 datasets were generated with (k_a, k_e, V, τ) equal to $(8, 2, 1, 5)$. For the analysis, we impose k_a to be larger than k_e. Note that this system of differential equations can possibly be used for the modelization of the concentration of drug in blood after an oral dose taking (Rowland and Tozer 1995) where D corresponds to the oral dose of drug, $x_1(t)$ is to the amount of drug in the stomach and $x_2(t)$ corresponds to the concentration of drug in the blood at time t.

5.2 Results

Tables 1, 2, 3 and 4 summarize the results of the simulation study. It gives estimation based on MCMC for the relative bias (in percent, based on the estimated posterior mean), the coverage probabilities at level 80% and 95% (based on 80% and 95%-HPD intervals), the root mean squared error and the mean posterior standard deviation for parameter k_a, k_e, V and τ for the Bayesian ODE-penalized B-spline approach with penalized Gaussian mixtures as error distributions (PGM) and Bayesian ODE-penalized B-spline approach with Gaussian error distribution (G). Figure 1 shows, for each simulation, the estimated density function for each of the 100 generated datasets (in light grey solid curve) together with the density function used for the data generation (in black solid curve).

When the generated error is Gaussian, the relative bias, estimated coverage probabilities, root mean squared error and mean posterior standard deviation for k_a, k_e, V and τ for G and PGM approaches are similar, meaning that the unnecessary flexibility in the error distribution with PGM is not an handicap (see the first column of Figure 1 for the estimated error density).

When the generated error term has a Student distribution, the relative bias is small for k_a, k_e and V. All the estimated coverage probabilities are in agreement with their nominal values. The RMSE for the ODE parameters is smaller in the PGM approach than in the G approach when the sample size is larger than 100. For the precision of measurement τ, the relative bias is positive and quite large when $n = 50$, but decreases with sample size. The estimated coverage probabilities for the precision parameter are not in agreement with their nominal value when $n \leq 100$ (except

CP80 for the PGM approach). For n larger than 200, only the coverage probabilities for the PGM approach are in agreement with their nominal values. The RMSE for τ is similar in the G and PGM approaches for each sample size. The mean posterior standard deviation of τ is larger in the PGM approach: this could partly explain why the estimated coverage probabilities for that parameter are larger than those estimated in the G approach. Note that the estimated error densities are getting closer to the target error density when the sample size increases.

When the generated error term has an extreme value distribution, the relative bias is small for all the ODE parameters. Their coverage probabilities are almost all in agreement with their nominal value in the PGM approach (except for sample size $n = 50$). The RMSE and the MPSD are smaller in the PGM approach for the ODE parameters and decrease as expected with sample size. The bias of the precision of measurement τ is slightly positive, especially for the PGM approach but decrease with the sample size. Estimated coverage probabilities at level 80% and 95% are larger in the PGM approach than in the G approach. But they are not in agreement with their nominal value, except for the PGM approach when the sample size is equal to 200 (CP95) and when the sample size is equal to 500 (CP80 and CP95). The estimated error densities are again getting closer to the target error density when increasing the sample size (see the first column of Figure 1).

When the sample size is at least 100, the relative bias for the precision of measurement τ is small but positive when the true error distribution is a mixture of two Gaussian distributions. The RMSE for that precision parameter is almost the same in the PGM and G approaches and decrease with sample size. The MPSD is smaller in the PGM approach for the precision: this could explain why the estimated coverage probabilities for τ in the PGM approach are almost in agreement with their nominal values whereas the estimated coverage probabilities for τ in the G approach are almost all superior to the corresponding nominal value for small sample sizes ($n \leq 100$). For the ODE parameters, the relative bias is small with the two approaches. The coverage probabilities are all in agreement with their nominal value in the PGM approach.

6 Application

Theophylline is a well-known anti-asthmatic agent, administrated orally. Twelve subjects are given oral dose at time 0 with blood samples taken at 11 time points over the following 25 hours (Davidian and Giltinan 1995). The blood samples are assayed for the theophylline concentration. Individual profiles are presented on Fig. 2. A common model for the kinetics of theophylline after oral administration is the two compartment model with first order absorption and elimination presented in Section 5. The modelling strategy described in Section 3 was used. For the ith subject, we approximate the amount of drug in the stomach by $x_{1i}(t) = (B_1(t))^T c_{1i}$ and the concentration of the drug in the central compartment by $x_{2i}(t) = (B_2(t))^T c_{2i}$. We parameterize the constant of absorption k_a, the constant of elimination k_e and the volume V of the central compartment in the log-scale with Gaussian random effects at the subject level. The prior information available for the study of this dataset concerns the PK parameters and the initial condition of the state functions. For the PK parameters, one has to force the constant of absorption to be larger than the constant of elimination. We have also a total confidence in the initial condition of the state functions: the

		Gaussian		Student		Extreme Value		Gaussian mixture	
		G	PGM	G	PGM	G	PGM	G	PGM
k_a	RB	0.2	0.2	0.9	0.9	0.5	0.5	0.6	0.7
	CP80	78	78	79	81	**72**	74	74	75
	CP95	95	94	91	92	94	93	91	92
	RMSE	4.76e-1	4.76e-1	5.18e-1	5.16e-1	5.35e-1	5.29e-1	5.11e-1	5.13e-1
	MPSD	4.90e-1	5.04e-1	4.89e-1	5.02e-1	4.91e-1	4.88e-1	4.80e-1	4.89e-1
k_e	RB	0.7	0.7	0.4	0.4	0.4	0.5	0.5	0.4
	CP80	78	78	81	83	**72**	**71**	**72**	73
	CP95	92	93	95	96	92	91	92	91
	RMSE	3.45e-2	3.45e-2	3.42e-2	3.39e-2	3.52e-2	3.10e-2	3.84e-2	3.62e-2
	MPSD	3.40e-2	3.44e-2	3.39e-2	3.39e-2	3.37e-2	3.15e-2	3.37e-2	3.30e-2
V	RB	-0.3	-0.3	0.1	0.1	-0.2	-0.3	-0.2	-0.2
	CP80	77	77	80	81	76	76	**72**	73
	CP95	92	92	92	92	**90**	**90**	92	93
	RMSE	2.97e-2	2.97e-2	3.14e-2	3.12e-2	3.11e-2	3.11e-2	3.25e-2	3.27e-2
	MPSD	3.00e-2	3.07e-2	2.97e-2	3.03e-2	2.94e-2	2.95e-2	2.94e-2	2.98e-2
τ	RB	4.6	4.5	11.6	9.4	12.2	12.1	6.8	6.5
	CP80	76	75	**67**	**66**	**62**	**64**	**90**	**89**
	CP95	92	93	**86**	**88**	**80**	**88**	99	98
	RMSE	1.23	1.24	1.73	1.66	1.71	1.61	1.01	9.39e-1
	MPSD	1.11	1.13	1.18	1.25	1.18	1.23	1.12	1.11

Table 1: Simulation - $n = 50$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values.

		Gaussian		Student		Extreme Value		Gaussian mixture	
		G	PGM	G	PGM	G	PGM	G	PGM
k_a	RB	0.9	0.9	-0.5	-0.5	0.2	0.2	-0.9	-0.7
	CP80	80	78	80	77	73	75	73	74
	CP95	95	95	94	94	97	96	**90**	93
	RMSE	3.45e-1	3.45e-1	3.42e-1	3.39e-1	3.52e-1	3.10e-1	3.84e-1	3.62e-1
	MPSD	3.40e-1	3.44e-1	3.39e-1	3.39e-1	3.37e-1	3.15e-1	3.37e-1	3.30e-1
k_e	RB	-0.3	-0.3	0.7	0.7	0.2	0.2	0.6	0.5
	CP80	77	77	83	83	74	75	77	80
	CP95	93	93	91	91	96	96	97	97
	RMSE	5.72e-2	5.72e-2	5.79e-2	5.73e-2	5.36e-2	4.91e-2	5.92e-2	5.39e-2
	MPSD	5.27e-2	5.30e-2	5.44e-2	5.42e-2	5.27e-2	4.95e-2	5.41e-2	5.33e-2
V	RB	0.3	0.3	-0.4	-0.4	-0.0	0.0	-0.5	-0.4
	CP80	77	77	77	78	77	79	77	82
	CP95	92	93	92	93	94	96	96	96
	RMSE	2.15e-2	2.15e-2	2.18e-2	2.17e-2	2.11e-2	1.90e-2	2.27e-2	2.07e-2
	MPSD	2.06e-2	2.08e-2	2.10e-2	2.09e-2	2.06e-2	1.92e-2	2.09e-2	2.05e-2
τ	RB	3.7	2.4	4.3	4.3	8.1	13.6	3.0	2.7
	CP80	80	78	**69**	73	**60**	**65**	**88**	87
	CP95	97	97	**86**	**90**	**80**	**88**	99	98
	RMSE	7.59e-1	7.41e-1	9.75e-1	9.36e-1	1.19	1.26	6.17e-1	6.05e-1
	MPSD	7.45e-1	7.45e-1	7.53e-1	7.91e-1	7.84e-1	9.22e-1	7.38e-1	7.35e-1

Table 2: Simulation - $n = 100$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values.

		Gaussian		Student		Extreme Value		Gaussian mixture	
		G	PGM	G	PGM	G	PGM	G	PGM
k_a	RB	0.4	0.4	0.2	0.3	-0.1	-0.3	0.1	0.2
	CP80	78	79	77	78	74	76	74	83
	CP95	94	95	96	96	93	94	93	95
	RMSE	2.45e-1	2.45e-1	2.40e-1	2.35e-1	2.38e-1	2.08e-1	2.18e-1	1.98e-1
	MPSD	2.41e-1	2.41e-1	2.36e-1	2.34e-1	2.36e-1	2.03e-1	2.37e-1	2.06e-1
k_e	RB	-0.2	-0.2	0.1	0.1	0.1	0.3	-0.0	-0.2
	CP80	80	79	82	83	76	83	84	82
	CP95	96	95	96	96	95	95	98	95
	RMSE	3.70e-2	3.70e-2	3.35e-2	3.29e-2	3.52e-2	3.20e-2	3.21e-2	3.00e-2
	MPSD	3.75e-2	3.74e-2	3.70e-2	3.66e-2	3.69e-2	3.26e-2	3.70e-2	3.31e-2
V	RB	0.0	0.0	-0.0	-0.0	-0.1	-0.2	0.1	0.2
	CP80	79	78	78	80	80	80	86	82
	CP95	96	96	94	95	95	96	99	94
	RMSE	1.42e-2	1.41e-2	1.39e-2	1.37e-2	1.41e-2	1.21e-2	1.20e-2	1.14e-2
	MPSD	1.47e-2	1.46e-2	1.45e-2	1.43e-2	1.44e-2	1.24e-2	1.45e-2	1.26e-2
τ	RB	2.1	2.3	4.3	4.9	5.0	10.2	2.2	3.5
	CP80	80	80	**66**	76	**66**	**72**	91	85
	CP95	94	94	**87**	92	**83**	92	99	95
	RMSE	5.14e-1	5.11e-1	6.93e-1	6.96e-1	8.51e-1	9.68e-1	4.26e-1	4.43e-1
	MPSD	5.04e-1	5.06e-1	5.27e-1	5.61e-1	5.33e-1	6.95e-1	5.14e-1	4.63e-1

Table 3: Simulation - $n = 200$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values.

		Gaussian		Student		Extreme Value		Gaussian mixture	
		G	PGM	G	PGM	G	PGM	G	PGM
k_a	RB	-0.0	-0.0	0.1	0.1	0.2	0.1	0.3	0.3
	CP80	83	83	82	78	**72**	78	75	79
	CP95	97	97	94	94	93	96	94	94
	RMSE	1.34e-1	1.34e-1	1.63e-1	1.53e-1	1.66e-1	1.42e-1	1.63e-1	1.20e-1
	MPSD	1.50e-1	1.50e-1	1.50e-1	1.43e-1	1.50e-1	1.26e-1	1.49e-1	1.16e-1
k_e	RB	-0.0	-0.0	0.0	-0.0	-0.1	-0.1	-0.2	-0.2
	CP80	81	82	78	78	77	79	79	79
	CP95	95	96	95	95	95	96	93	94
	RMSE	2.29e-2	2.27e-2	2.45e-2	2.33e-2	2.47e-2	2.11e-2	2.41e-2	1.97e-2
	MPSD	2.35e-2	2.34e-2	2.35e-2	2.25e-2	2.33e-2	2.03e-2	2.31e-2	1.90e-2
V	RB	0.0	0.0	-0.0	0.1	0.1	0.1	0.1	0.0
	CP80	82	83	76	79	80	77	87	83
	CP95	97	97	95	94	94	95	99	97
	RMSE	7.91e-3	7.84e-3	9.39e-3	8.85e-3	9.76e-3	7.88e-3	8.06e-3	6.42e-3
	MPSD	9.20e-3	9.18e-3	9.20e-3	8.75e-3	9.13e-3	7.67e-3	9.07e-3	6.99e-3
τ	RB	0.6	0.3	2.9	2.6	2.0	4.9	2.2	2.7
	CP80	82	81	**68**	79	**67**	76	87	83
	CP95	98	98	**89**	93	**88**	94	97	97
	RMSE	2.79e-1	2.75e-1	4.17e-1	4.02e-1	4.60e-1	5.05e-1	3.14e-1	3.22e-1
	MPSD	3.19e-1	3.12e-1	3.20e-1	3.98e-1	3.24e-1	4.37e-1	3.24e-1	2.65e-1

Table 4: Simulation - $n = 500$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values.

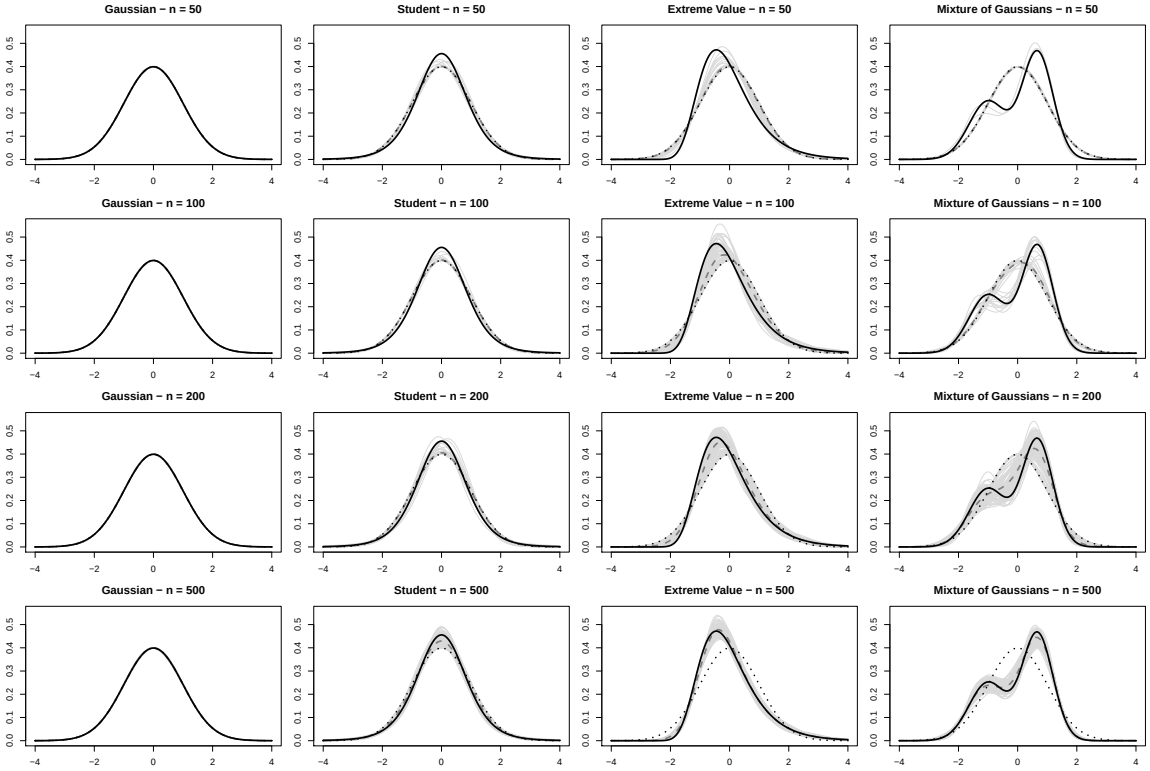

Figure 1: Estimated error density functions for sample size (50, 100, 200, 500) (one light grey curve per dataset) with pointwise average of these estimated densities in dark grey dashed curve. The black solide curve is the true density function and the black dotted curve is the standardized Gaussian density.

initial amount of drug in the stomach is equal to the oral dose and the initial concentration of the drug in the blood is null.

MCMC chains of length 10000 (including a 1000 burnin) were generated using Metropolis-within-Gibbs algorithm described in Section 4. We use this sample to reparametrize the problem by applying a translation and a rotation to the parameter vector $\theta = (k_a, k_e, V)^T$ and for all the transformed mixture weights d_j. A final chain of length 25000 (including an extra burnin of 5000) is generated for the inference. Convergence for each variable was confirmed by the Geweke's convergence diagnostic. Figure 3 shows (in solid black line) the pointwise posterior median of the error density function with the 80% and 95% pointwise credibility intervals (in gray). It suggests that the error density is symmetric and highly picked. Note that the standard Gaussian distribution (in dotted line) is not within the pointwise credibility intervals, clearly suggesting that the fitted error density is not Gaussian.

Figure 4 shows in (a) the estimated amount of drug in the stomach with pointwise 95% credibility sets, in (b) the estimated concentration of theophylline in blood with pointwise 95% credibility

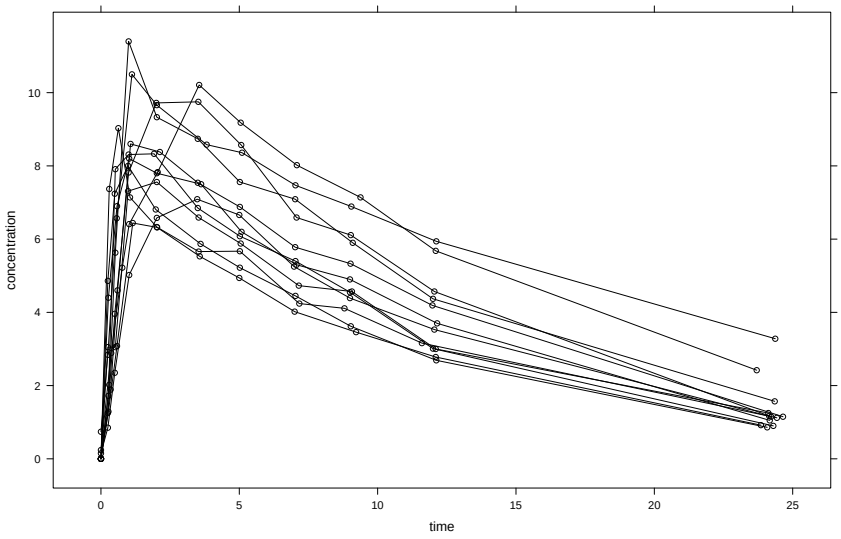

Figure 2: Pharmacokinetic profiles for the Theophylline dataset.

sets and in (c) the pointwise credibility interval for the posterior predictive distribution of the concentration of theophylline in blood for subject 2. The first two curves are obtained by estimating at each MCMC generation of the spline coefficients, the amount of drug in the stomach and the concentration of drug in the central compartment over a detailed grid of times. Then the 2.5%, 50% and 97.5% quantiles of the MCMC chain of the estimated curves are taken at each time point. The last curve corresponds to a visual predictive check for the concentration. It corresponds to a 95% credibility set of the predictive distribution of the concentration of the theophylline drug. This curve is obtained by estimating at each MCMC generation the 2.5%, 50% and 97.5% quantiles of the predictive concentration over the same detailed grid of time.

7 Discussion

A full Bayesian approach based on penalized smoothing methods was presented for the estimation of ODE parameters, the approximation of the state function involved in an affine ODE model and the estimation of error densities. The estimation of the ODE parameters and of the state functions was based on a Bayesian ODE-penalized B-spline approach with error densities estimated in a flexible way using Bayesian penalized Gaussian mixture (PGM).

The PGM approach avoids restrictive parametric assumptions on the error distribution. This Gaussian mixture assumption for the error distribution is convenient as, conditionally on the mixture component, one is back to the standard B-ODE-P-spline setting assumed in Jaeger and Lambert (2011). Label switching couldn't also occurs with this PGM approach. Simulations suggest that the B-ODE-P-spline approach combined with PGM is as efficient as the standard B-ODE-P-spline

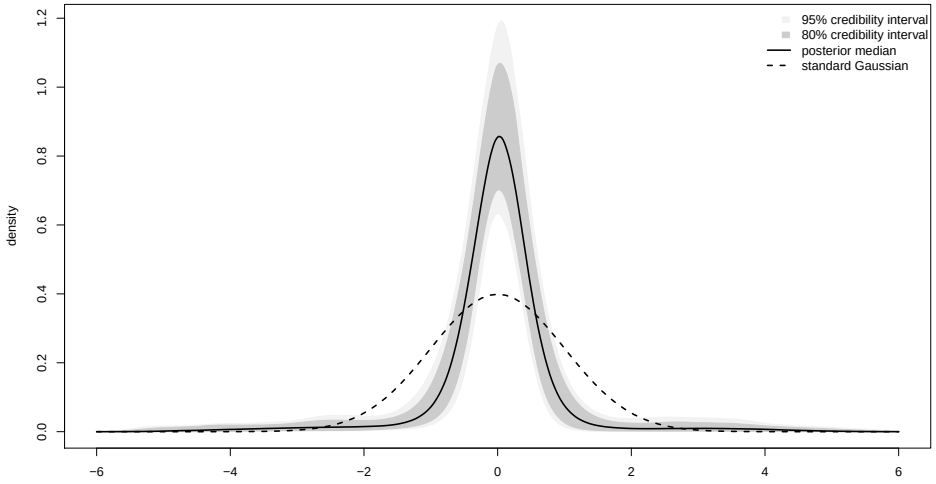

Figure 3: Fitted error density (solid curve is the pointwise posterior median and grey regions delimit pointwise 80% and 95% credible intervals), dotted curve is the Gaussian density.

approach in terms of relative bias, coverage probabilities, root mean squared errors and mean posterior standard deviation when the distribution of the error terms are Gaussian and outperforms the standard approach for small and medium sample sizes when the error distribution is homogeneous non-Gaussian.

As we work in a Bayesian framework, it is possible to inject prior informations on the ODE parameters and on the precision of measurement, as well as to include uncertainty with respect to the specification of the initial condition of the state functions. Furthemore, uncertainty measures are readily available through the credibility sets computed from the MCMC chains.

Some extensions are desirable. The Bayesian ODE-penalized B-spline approach combined with penalized Gaussian mixtures assumes that the error distribution is homogeneous. The case where the variance of the error distribution is nonhomogeneous, namely when the variance of the error depends on time and/or on the state function that is measured, has to be considered. In the case where the error term depends on the state function, it will be complicated to get rid of the posterior correlation between the generated spline coefficients and the differential equation parameters because the marginalization of the joint posterior distribution with respect to the spline coefficients will be no more easily done.

Finally, the proposed Bayesian approach was deliberately focused on the affine ODE case. Jaeger and Lambert (2011) highlight the major role of the normalization constant in the prior distribution of the spline coefficients. This is an essential information to deal with more advanced problems such as systems of nonlinear differential equations where by simplicity one might be tempted to ignore such a constant as it has no explicit analytical form. Analytical or numerical approximations will be required and should be assessed in these challenging settings.

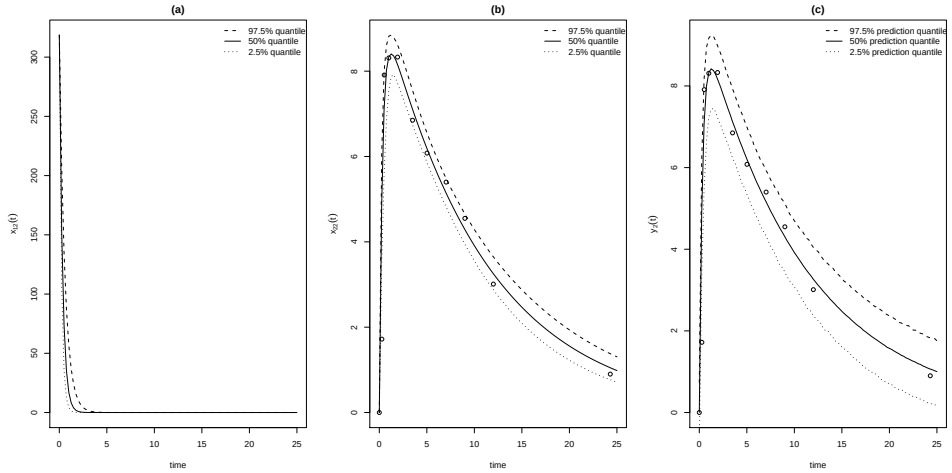

Figure 4: (a): Pointwise 95%-credibility interval for the amout of drug in stomach for subject 2. (b): Pointwise 95%-credibility interval for the mean concentration of drug in blood for subject 2. (c): Pointwise 95%-credibility interval for the prediction of drug concentration in blood for subject 2.

References

Cao, J. and Zhao, H. (2008). "Estimating dynamic models for gene regulation networks." *Bioinformatics*, 24: 1619–1624.

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

Diebolt, J. and Robert, C. (1994). "Estimation of finite mixture distributions through Bayesian sampling." *Journal of the Royal Statistical Society: Series B*, 56: 363–375.

Eilers, P. and Marx, B. (1996). "Flexible smoothing with B-splines and penalties." *Statistical Science*, 11: 89–121.

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli*, 7: 223–242.

Jaeger, J. and Lambert, P. (2011). "Bayesian Generalized Profiling Estimation in Hierarchical Linear Dynamic Systems." IAP Technical report 11001, Institut de Statistique, Biostatistique et Sciences Actuarielles.

Jullion, A. and Lambert, P. (2007). "Robust specification of the roughness penalty prior distribution in spatially adaptative Bayesian P-splines models." *Journal of Computational Statistics & Data Analysis*, 51: 2542–2558.

Komárek, A. and Lesaffre, E. (2008). "Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions." *Journal of the American Statistical Association*, 103: 523–533.

Lambert, P. (2007). "Archimedean copula estimation using Bayesian splines smoothing techniques." *Computational Statistics & Data Analysis*, 51(12): 6307–6320.

Lambert, P. and Eilers, P. (2009). "Bayesian density estimation from grouped continuous data." *Computational Statistics & Data Analysis*, 53: 1388–1399.

Lang, S. and Brezger, A. (2004). "Bayesian P-splines." *Journal of Computational and Graphical Statistics*, 13: 183–212.

Lindsey, J. K., Wang, J., Byrom, B. J. W. D., and Hind, I. D. (2000). "Simultaneous modelling of flosequinan and its metabolite." *European Journal of Clinical Pharmacology*, 55: 827–836.

Ramsay, J., 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*, 69: 741–796.

Ramsay, J. and Silverman, B. W. (2005). *Fonctional Data Analysis*. Springer Series in Statistics.

Richardson, S. and Green, P. (1997). "On Bayesian analysis of mixture with an unknown number of components." *Journal of the Royal Statistical Society: Series B*, 59: 731–792.

Rowland, M. and Tozer, T. N. (1995). *Clinical Pharmacokinetics: Concepts and Applications*. Lippincott Williams & Wilkins.

Stephens, M. (2000). "Dealing with label-switching in mixture models." *Journal of the Royal Statistical Society, Series B*, 62: 795–809.

Tanner, M. and Wong, W. (1987). "The calculation of the posterior distributions through data augmentation." *Journal of the American Statistical Association*, 82: 528–540.

Varah, J. (1982). "A spline least squares method for numerical parameter estimation in differential equations." *SIAM Journal of Computational and Graphical Statistics*, 3: 28–46.

Acknowledgments

Financial support from the IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) and financial support from the contract "Projet d'Actions de Recherche Concertées" nr. 11/16/039 of the "Communauté franaise de Belgique", granted by the "Académie universitaire Louvain", are gratefully acknowledged.