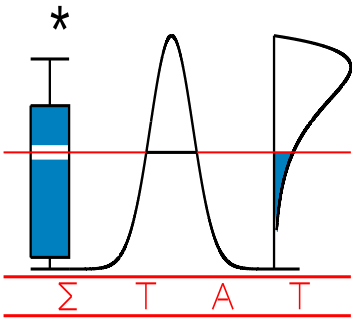
T E C H N I C A L

R E P O R T

11001

Bayesian generalized profiling estimation in hierarchical linear dynamic systems

JAEGER, J. and P. LAMBERT



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Bayesian Generalized Profiling Estimation in Hierarchical Linear Dynamic Systems

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

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

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

[*]*email:*jonathan.jaeger@uclouvain.be
[**]*email:*p.lambert@ulg.ac.be

January 27, 2010

Abstract

Ordinary differential equations (ODEs) are widely used to model physical, chemical and biological processes. Current methods for parameter estimation are computationally intensive and not suitable for inference and prediction. Frequentist approaches based on ODE-penalized smoothing techniques have recently solved part of these drawbacks.
In this paper we propose a full Bayesian approach based on ODE-penalized B-splines to jointly estimate ODE-parameters and state function from linear systems of differential equations. Simulations inspired by pharmacokinetic studies show that the proposed method provides comparable results to methods based on explicit solution of the ODEs and outperforms the frequentist ODE-penalized smoothing approach. We extend the basic model to a hierarchical one in order to study cases where several subjects are involved. This Bayesian hierarchical approach is illustrated on real data for the study of perfusion ratio after a femoral artery occlusion.

Keywords: Dynamic systems; Ordinary differential equations; Parameter estimation; Penalized B-spline.

Acknowledgement: *Financial support from the IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) is gratefully acknowledged.*

1 Introduction

Assume that changes in the states $\boldsymbol{x}(t) \in \mathbb{R}^d$ of a dynamic system are governed by a set of differential equations:

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

where f is a known function and $\boldsymbol{\theta} \in \mathbb{R}^q$ an unknown vector of parameters.
It is assumed that only a subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state function $\boldsymbol{x}$ 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 $y_{jk} = \boldsymbol{x}_j(t_{jk}) + \epsilon_{jk}$ the corresponding measurement. An ordinary differential equation (ODE) system with initial values $\boldsymbol{x}(0) = \boldsymbol{x}_0$ of the output variable has a unique solution if f is Lipschitz continuous with respect to $\boldsymbol{x}$.
In this paper, we will consider the case where f is an affine transformation with respect to $\boldsymbol{x}$ ensuring the existence and uniqueness of the solution of the dynamic system model. We propose a full Bayesian approach to jointly estimate the ODE parameters $\boldsymbol{\theta}$ and the state functions $\boldsymbol{x}(t)$ from $\{(t_{jk}, y_{jk}), j \in \mathcal{J}, k = 1, \ldots, n_j\}$.
In the linear ODEs case, an explicit solution is available for (1) and the estimation procedure reduces to a nonlinear

regression problem. But most of the time, when f is nonlinear with respect to $\boldsymbol{x}(t)$, ODEs do not have explicit solution. Several methods have been proposed for estimating the parameter $\boldsymbol{\theta}$ and the state functions $\boldsymbol{x}(t)$ of ODEs and can be classified in two parts.

The most commonly used estimation procedures rely on nonlinear least squares (NLS) procedures (Biegler et al., 1986). First the ODE system is solved numerically given the current estimate of $\boldsymbol{\theta}$ and the initial value $\boldsymbol{x}(0)$. Then the numerical solution is compared with the measured responses, suggesting an update of the estimate for $\boldsymbol{\theta}$. Iterations between computation of numerical approximation of the solution and parameter updating continue until some convergence criteria for the parameter estimate is met. These NLS procedures have many drawbacks. First it's time consuming since a numerical approximation of the state function must be computed for each updated value of the parameter estimate. Furthermore the accuracy of the parameter estimate is highly dependent on the accuracy of the numerical integration. Bayesian approaches of these NLS procedures have been proposed in Lunn et al. (2002), but the posterior distribution of the parameter $\boldsymbol{\theta}$ has no closed form when the numerical solution of the ODE has no closed form. The most troublesome disadvantage of this Bayesian approach is that the reconstruction of the solution of the system of differential equations is needed at each iteration of the MCMC algorithm.

An alternative procedure was introduced by Varah (1982) that does not require numerical integration for estimating the parameters involved in the ODEs. The proposed spline smoothing approach is a two-stage procedure. First, each state function is estimated by a standard spline smoothing method. Then parameter $\boldsymbol{\theta}$ is estimated through the minimization of a criterion that involves the differential equation with the state functions replaced by their spline approximations. Varah's approach works well in simple cases but is highly dependent on the spline fit: a poor spline fit can lead to a poor parameter estimate.

Poyton et al. (2006) and Ramsay et al. (2007) proposed a generalization of Varah's method by estimating the spline basis coefficients with an ODE-based penalty and by iterating the two-step procedure. The use of an ODE-penalized likelihood as fitting criterion is a trade off between fitting the data and solving the ODE. This method converges quickly and gives less biased and more precise estimates. Moreover, this procedure allows estimating unobserved state functions. Nevertheless, in this ODE-penalized B-spline approach, two critical points have to be highlighted. The first point is the delicate choice of the ODE-adhesion parameter that controls the relative emphasis between goodness-of-fit and solving the differential equation. The second point is the approximation that must be done for the variance of the parameter estimate. A Bayesian framework of this ODE-penalized B-spline approach was proposed by Campbell (2007). Based on the idea of parallel tempering, this approach is composed of three components. The first is the prior distribution for the ODE parameter $\boldsymbol{\theta}$ using the ODE-penalty term and prior information about this parameter. The second component is the use of a sequence of parallel MCMC chains of the ODE-parameter, based on increasing values of ODE-adhesion parameters. The last component is the estimation of the spline coefficents: knowing the ODE-parameter and the level of ODE-adhesion parameters, point estimates for the spline coefficients are plugged-in. Thus, that method is not fully Bayesian.

In this paper, we attempt to develop a fully Bayesian approach to jointly estimate parameters and state functions of linear ODE models. In Section 2, we first give a brief overview of the generalized profiling estimation procedure for ODE models. Then we aim at providing a Bayesian framework which can be viewed as a generalization of Bayesian P-splines models (Lang and Brezger, 2004). Section 3 presents the strategies used to explore the joint posterior distribution with Markov Chain Monte Carlo (MCMC). Section 4 generalizes the Bayesian ODE-penalized B-spline approach to multi-subject studies. Section 5 gives the results of some simulations comparing the performance of our approach with those of traditional methods. Application on real data is given in Section 6. Conclusions and generalizations of the proposed method are discussed in Section 7.

2 Bayesian Generalized Profiling Estimation for Linear Dynamic Systems Model

In this section, we first remind the B-spline definition and give an overview of the generalized profiling estimation procedure for ODE proposed in Ramsay et al. (2007). This methodology may be viewed as a generalization of P-spline theory (Eilers and Marx, 1996). Then we propose a full Bayesian framework for this method.

2.1 B-spline definition and generalized profiling estimation procedure

To obtain a joint estimation of the vector of parameters $\boldsymbol{\theta}$ and of the state functions $\boldsymbol{x}(t)$ governed by the system of differential equations, Poyton et al. (2006), Ramsay et al. (2007), Cao and Zhao (2008) proposed to use a B-spline basis

function expansion combined with a penalty related to the system of differential equations.
A B-spline basis of order p is defined using m inner knots $\tau_1 \leq \ldots \leq \tau_m$ and two p-multiple knots τ_0 and τ_{m+1} corresponding respectively to the lower and the upper bound of the study interval with conditions $\tau_0 < \tau_1$ and $\tau_m < \tau_{m+1}$. These $m+2$ knots are typically placed equidistantly on the time interval of the study.
Each of the $K = p+m$ basis functions consists at most of p polynomial pieces, each of degree $(p-1)$. These polynomial pieces joint at most at $(p-1)$ inner knots with continuous derivatives up to order $(p-2)$. Figure 1 presents a B-spline basis of order 4, also called cubic B-splines, with knots at each tenth between 0 and 1.

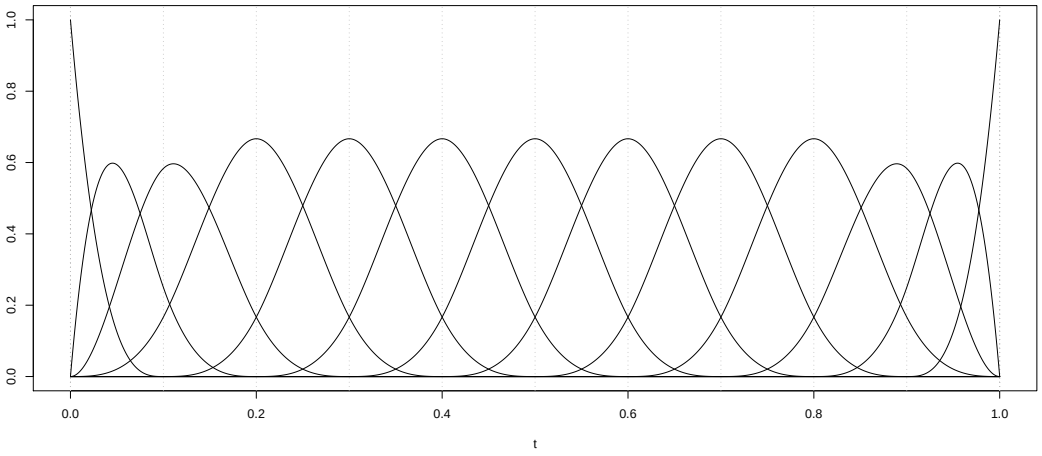


Figure 1: Cubic B-splines on $[0;1]$ for 9 equidistant inner knots.

Denote by $\boldsymbol{B}_j(t)$ the K_j-vector of B-spline basis functions of order p evaluated at time t that is used to approximate the j-th component of the state function $\boldsymbol{x}(t)$. The approximation $\widetilde{\boldsymbol{x}}_j(t)$ of $\boldsymbol{x}_j(t)$ is expressed as a linear combination of these B-spline basis functions:

$$\begin{aligned} \widetilde{\boldsymbol{x}}_j(t) &= \sum_{k=1}^{K_j} \boldsymbol{B}_{jk}(t)\, \boldsymbol{c}_{jk} \\ &= \left(\boldsymbol{B}_j(t)\right)^T \boldsymbol{c}_j \end{aligned}$$

where $K_j = p + m_j$ is the number of B-spline basis functions in vector $\boldsymbol{B}_j(t)$ that is chosen to ensure enough flexibility to capture the variation in $\boldsymbol{x}_j(t)$ (with the possibility to increase the order p or the number m_j of inner knots). That choice for the spline basis has many advantages. First it permits to directly link the initial condition of the state function (i.e. $\boldsymbol{x}_j(0)$) to the first component of the basis function expansion (i.e. $\boldsymbol{c}_{j1}$). Note also that tied knots (i.e. inner knots of order p) on $(0,T)$ could also be used to deal with discontinuities in $\boldsymbol{x}_j(t)$. More details on B-spline properties are available in de Boor (2001).

Data fitting criterion and fidelity to the ODE

Let $\boldsymbol{\epsilon}_j$ denote the vector of errors associated to the observed variable $\boldsymbol{y}_j^T = \left(y_{j1}, \ldots, y_{jn_j}\right)$, $j \in \mathcal{J}$ and let $g_j(\boldsymbol{\epsilon}_j|\boldsymbol{\sigma}_j)$ be the parametric density of these errors conditional on the vector $\boldsymbol{\sigma}_j$. Let $\boldsymbol{\sigma}$ denote the concatenation of these $\boldsymbol{\sigma}_j$, $j \in \mathcal{J}$. The data fitting criterion can be taken to be the log-likelihood:

$$H(\boldsymbol{c}, \boldsymbol{\sigma}|\boldsymbol{y}) = \sum_{j \in \mathcal{J}} \log\left(g_j(\boldsymbol{\epsilon}_j|\boldsymbol{\sigma}_j)\right)$$

where $\boldsymbol{\epsilon}_{jk} = \boldsymbol{y}_{jk} - (\boldsymbol{B}_j(t_{jk}))^T \boldsymbol{c}_j$. In practice, one usually assumes independent Gaussian errors with mean 0 and standard deviations σ_j. Then,

$$H(\boldsymbol{c}, \boldsymbol{\sigma}|\boldsymbol{y}) = \sum_{j \in \mathcal{J}} \left\{ -n_j \log(\sigma_j) - \frac{1}{2\sigma_j^2} \left\| \boldsymbol{y}_j - \boldsymbol{B}_j \boldsymbol{c}_j \right\|^2 \right\}.$$

For such objective criterion, the estimation of B-spline coefficients are highly dependent on the number of knots and their location. If there are too few knots, the approximated functions $\widetilde{\boldsymbol{x}}(t)$ will be too smooth and it will not capture all the available information on the ODE system. If the number of knots is too large, the approximated functions will capture noisy variation in the data.
To avoid this problem, Ramsay and Silverman (2005) propose to consider a large set of knots and to control the spline coefficients by adding a model-based-penalty. To do that, for each equation of the differential equation system, a differential equation operator is introduced:

$$L_{j,\boldsymbol{\theta}}(\boldsymbol{x}(t)) = D\boldsymbol{x}_j(t) - f_j(\boldsymbol{x}, t, \boldsymbol{\theta}), j = 1, \ldots, d.$$

The proximity between the approximation of the output function, $\widetilde{\boldsymbol{x}}_j(t)$, and the solution $\boldsymbol{x}_j(t)$ of the corresponding j-th equation over $[0; T]$ can be assessed by:

$$PEN_j(\widetilde{\boldsymbol{x}}) = \int \left\{ L_{j,\boldsymbol{\theta}}(\widetilde{\boldsymbol{x}}(t)) \right\}^2 dt$$

where the integration is over an interval of length $\mathcal{L}$ which contains all times of measurement. Note that for the linear differential equation system that we consider here, this j-th penalty term is an homogenous polynomial of degree 2 in the B-spline coefficients. The full fidelity-to-ODE measure is then given by:

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

where $\boldsymbol{c} = \left(\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_d^T\right)^T$ is the vector of all spline coefficients of length $K = \sum_{j=1}^{d} K_j$, $\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a $K \times K$ supermatrix that is constructed by placing the corresponding penalty matrices involved in each $PEN_j(\tilde{\boldsymbol{x}})$, $\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a vector of length K corresponding to a penalty vector and $l(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a penalty constant.
The ODE-adhesion coefficients $\gamma_j \geq 0$ permit to weight and to control the relative emphasis on goodness-of-fit and solving the system of differential equations, i.e. to express the confidence that one has in the differential equation system as description of the dynamic system. For a given vector $\boldsymbol{\gamma}$, the spline parameters estimates are implicitly defined using the current estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$, i.e. as maximizers of the fitting criterion

$$J(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = \sum_{j \in \mathcal{J}} \log\left(g_j(\boldsymbol{\epsilon}_j|\boldsymbol{\sigma}_j)\right) - \frac{1}{2} PEN(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma}). \quad (2)$$

Then, the data fitting criterion J is optimized with respect to $(\boldsymbol{\theta}, \boldsymbol{\sigma})$. As explained more precisely in Ramsay et al. (2007), the dependence of J on $(\boldsymbol{\theta}, \boldsymbol{\sigma})$ is twofold: directly by its definition, but also implicitly through the spline parameters estimates. Note that, if large values are taken for the $\boldsymbol{\gamma}_j$s, one simply forces the state of the system, $\boldsymbol{x}(t)$, to be the exact solution of the system of differential equations. At the opposite, if the $\boldsymbol{\gamma}_j$s are small, the penalty term PEN plays a small role and goodness of fit to the data becomes the leading principle in the estimation of $\boldsymbol{c}$.

2.2 Bayesian generalized profiling estimation for linear ODE

In terms of likelihood, the penalty $PEN(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma})$ appears as a term that is substracted from the log-likelihood H. The penalized function has the following form:

$$J = H - \frac{1}{2} PEN(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma}).$$

The same log-posterior is obtained in a Bayesian framework with the following model specification:

$$\begin{cases} y_{jk}|\boldsymbol{c}_j, \tau_j \sim \mathcal{N}\left((\boldsymbol{B}_j(t_{jk}))^T \boldsymbol{c}_j; \tau_j^{-1}\right) & \forall j \in \mathcal{J}, \forall k \in \{1, \ldots, n_j\} \\ \pi(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\gamma}) \propto \exp\left(-\frac{1}{2}\left\{\boldsymbol{c}^T \boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) \boldsymbol{c} + 2\boldsymbol{c}^T \boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + l(\boldsymbol{\theta}, \boldsymbol{\gamma})\right\}\right) \end{cases}$$

where τ_j is the precision of measurment, i.e. $\tau_j = \frac{1}{\sigma_j^2}$.
The penalty from the frequentist penalized likelihood approach translates, in a Bayesian framework, into a prior distribution for the spline coefficients $\boldsymbol{c}$. Note that this model specification does not express information about the initial condition of the state function. If these are available, one could choose to work with the following alternative model specification:

$$\left\{ \begin{array}{lcll} y_{jk}|\boldsymbol{c}_j, \tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B}_j\left(t_{jk}\right)\right)^T \boldsymbol{c}_j; \tau_j^{-1} \right) & \forall j \in \mathcal{J}, \forall k \in \{1, \ldots, n_j\} \\ \pi(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\gamma}) & \propto & \exp\left(-\frac{1}{2} PEN\left(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma}\right) - \frac{1}{2}\left(\boldsymbol{c}-\boldsymbol{\mu_c}\right)^T {\boldsymbol{\Sigma_c}}^{-1}\left(\boldsymbol{c}-\boldsymbol{\mu_c}\right) \right) & \end{array} \right. \tag{3}$$

The K-vector $\boldsymbol{\mu_c}$ and the $K \times K$-matrix ${\boldsymbol{\Sigma_c}}^{-1}$ are used to translate the information available about the initial condition of the state function $\boldsymbol{x}$ in the model. Typically, the first component in each $\boldsymbol{\mu_{c_j}}, j \in \{1, \ldots, d\}$ of the vector $\boldsymbol{\mu_c} = \left({\boldsymbol{\mu_{c_1}}}^T, \ldots, {\boldsymbol{\mu_{c_d}}}^T\right)^T$ are fixed to the corresponding initial value $\boldsymbol{x}_j\left(0\right)$ of the state function. The corresponding precision in ${\boldsymbol{\Sigma_c}}^{-1}$ allows to express uncertainty with respect to this initial condition. In the Section 3, the corresponding prior distribution (including the normalizing constant) for $\boldsymbol{c}$ will be given and the case where the precision tends to infinity will be considered in a simple case in Appendix 1.
Further priors are required to have a full model specification, namely for the precision parameters $\boldsymbol{\tau} = \left(\tau_j, j \in \mathcal{J}\right)^T$, for the vector $\boldsymbol{\gamma} = \left(\gamma_1, \ldots, \gamma_d\right)^T$ of ODE-adhesion parameters and for the vector of ODE-parameters $\boldsymbol{\theta}$. For each τ_j, $j \in \mathcal{J}$, the conditional precision of the vector of response $\boldsymbol{y}_j$, it is convenient to take a gamma prior distribution:

$$\tau_j \sim \mathcal{G}\left(a_{\tau_j}, b_{\tau_j}\right), \forall j \in \mathcal{J}$$

The prior of each ODE-adhesion parameter γ_j, $j = 1, \ldots, d$ can also be chosen to be a gamma distribution:

$$\gamma_j \sim \mathcal{G}\left(a_{\gamma_j}, b_{\gamma_j}\right), \forall j = 1, \ldots, d$$

where $\mathcal{G}\left(a, b\right)$ is the gamma distribution with mean a/b and variance a/b^2. As recommended in Lang and Brezger (2004) for standard P-splines 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).
For the vector $\boldsymbol{\theta}$ of differential equation parameters, the chosen prior will depend on the context. Let us denote this prior distribution by $\pi\left(\boldsymbol{\theta}\right)$.

2.3 Examples of Dynamic Systems

The two first examples given here are based on models that are used in pharmacokinetics. Pharmacokinetic (PK) studies are set up to analyze the evolution of drug concentration in plasma over time. With PK compartment models, the basic idea is that the body can be seen as a system of compartments communicating with each other. Each compartment corresponds to a tissue or a group of tissues with similar blood flow and uniform drug concentration.
The last example is based on model used to describe the evolution of perfusion ratio after a femoral artery occlusion.

One-compartment model with single IV bolus injection

This model corresponds to an instantaneous injection of a dose D of a drug into the blood. Here it is assumed that the entire body is assimilated to one central compartment of volume V and that the drug is eliminated at a rate k_e. The differential equation system with the initial condition is:

$$\left\{ \begin{array}{lcl} \frac{dC(t)}{dt} & = & -k_e C\left(t\right), \\ C\left(0\right) & = & \frac{D}{V}. \end{array} \right.$$

It is also assumed that k_e and V are positive quantities. Under this differential equation model, the concentration $C\left(t\right)$ at time t in the central compartment is easily shown to be:

$$C\left(t\right) = \frac{D}{V}\exp\left(-k_e t\right).$$

For this model, we are interested by the estimation of $\boldsymbol{\theta}^T = \left(k_e, V\right)$ from a set of observations $\{(t_k, y_k) : k = 1, \ldots, n\}$ where:

$$y_k = C\left(t_k\right) + \epsilon_k.$$

For this one-compartment model with single IV bolus injection, the concentration is approximated by:

$$\widetilde{C}(t) = V^{-1} (\boldsymbol{B}(t))^T \boldsymbol{c}$$

and the differential equation operator is defined by:

$$\begin{aligned} L_{\boldsymbol{\theta}}\left(\widetilde{C}(t)\right) &= \frac{d\widetilde{C}(t)}{dt} + k_e \widetilde{C}(t) \\ &= \frac{1}{V}\left\{\left(\boldsymbol{B}^{(1)}(t)\right)^T \boldsymbol{c} + k_e (\boldsymbol{B}(t))^T \boldsymbol{c}\right\}. \end{aligned}$$

The penaly term is then defined by:

$$\begin{aligned} PEN &= \gamma \int \left(\frac{1}{V}\left\{\left(\boldsymbol{B}^{(1)}(t)\right)^T \boldsymbol{c} + k_e (\boldsymbol{B}(t))^T \boldsymbol{c}\right\}\right)^2 dt \\ &= \frac{\gamma}{V^2} \boldsymbol{c}^T \left(\boldsymbol{P^{11}} + k_e \left(\boldsymbol{P^{10}} + \boldsymbol{P^{01}}\right) + k_e^2 \boldsymbol{P^{00}}\right) \boldsymbol{c} \\ &= \boldsymbol{c}^T \boldsymbol{R}(k_e, V, \gamma)\, \boldsymbol{c}, \end{aligned} \tag{4}$$

where $\boldsymbol{B}^{(1)}(t)$ corresponds to the vector of first derivative of each component of the vector of B-spline functions $\boldsymbol{B}(t)$ and $\boldsymbol{P^{ij}} = \int \boldsymbol{B}^{(i)}(t) \left(\boldsymbol{B}^{(j)}(t)\right)^T dt$, $i, j \in \{0, 1\}$.

Two-compartment model with oral dosing

This compartment model arises when drug at dose D is administrated orally in the previous setting. Let $Q_a(t)$ denote the quantity of drug into the stomach (first compartment) at time t. This quantity goes to the plasma with an absorption rate k_a. The drug has concentration $C_e(t)$ in this central compartment of volume V and is eliminated at rate k_e. The differential equation systems is:

$$\left\{\begin{array}{lcl} \frac{dQ_a(t)}{dt} &=& -k_a Q_a(t), \\ \frac{dC_e(t)}{dt} &=& \frac{k_a}{V} Q_a(t) - k_e C_e(t), \\ Q_a(0) &=& D, \\ C_e(0) &=& 0. \end{array}\right.$$

For this model, it is supposed that k_a, k_e and V are positive and that $k_a > k_e$ (the drug cannot be eliminated more quickly than it is absorbed). Note that in this setting, only the concentration of the drug in the central compartment is usually observed.
The analytic solution of this dynamic system is:

$$\left\{\begin{array}{lcl} Q_a(t) &=& D \exp(-k_a t), \\ C_e(t) &=& \frac{D}{V} \frac{k_a}{k_a - k_e} \left(\exp(-k_e t) - \exp(-k_a t)\right). \end{array}\right.$$

Again, one is interested in estimating $\boldsymbol{\theta}^T = (k_a, k_e, V)$ from observations $\{(t_k, y_k) : k = 1, \ldots, n\}$ where:

$$y_k = C_e(t_k) + \epsilon_k$$

For this two-compartment model with oral dosing, the quantity of drug in the stomach is approximated by $\widetilde{Q}_a(t) = (\boldsymbol{B_a}(t))^T \boldsymbol{c_a}$ and the concentration of the drug in the central compartment is approximated by $\widetilde{C}_e(t) = (\boldsymbol{B_e}(t))^T \boldsymbol{c_e}$.
For the first differential equation of the system, the corresponding differential equation operator is:

$$\begin{aligned} L_{1,\boldsymbol{\theta}}\left(\widetilde{Q}_a(t)\right) &= \frac{d\widetilde{Q}_a(t)}{dt} + k_a \widetilde{Q}_a(t) \\ &= \left(\boldsymbol{B_a}^{(1)}(t)\right)^T \boldsymbol{c_a} + k_a (\boldsymbol{B_a}(t))^T \boldsymbol{c_a} \end{aligned}$$

and the corresponding penalty term is equal to:

$$\begin{aligned} PEN_1 &= \int \left(\left(\boldsymbol{B_a}^{(1)}(t) \right)^T \boldsymbol{c_a} + k_a \left(\boldsymbol{B_a}(t) \right)^T \boldsymbol{c_a} \right)^2 dt \\ &= \boldsymbol{c_a}^T \left(\boldsymbol{P_{aa}^{11}} + k_a \left(\boldsymbol{P_{aa}^{10}} + \boldsymbol{P_{aa}^{01}} \right) + k_a^2 \boldsymbol{P_{aa}^{00}} \right) \boldsymbol{c_a} \\ &= \boldsymbol{c_a}^T \boldsymbol{R_a} \boldsymbol{c_a}. \end{aligned}$$

The second differential equation operator is defined by:

$$\begin{aligned} L_{2,\boldsymbol{\theta}} \left(\widetilde{C}_e(t) \right) &= \frac{d\widetilde{C}_e(t)}{dt} + k_e \widetilde{C}_e(t) - \frac{k_a}{V} \widetilde{Q}_a(t) \\ &= \left(\boldsymbol{B_e}^{(1)}(t) \right)^T \boldsymbol{c_e} + k_e \left(\boldsymbol{B_e}(t) \right)^T \boldsymbol{c_e} - \frac{k_a}{V} \left(\boldsymbol{B_a}(t) \right)^T \boldsymbol{c_a} \end{aligned}$$

and the second penalty term is equal to:

$$\begin{aligned} PEN_2 &= \int \left(\left(\boldsymbol{B_e}^{(1)}(t) \right)^T \boldsymbol{c_e} + k_e \left(\boldsymbol{B_e}(t) \right)^T \boldsymbol{c_e} - \frac{k_a}{V} \left(\boldsymbol{B_a}(t) \right)^T \boldsymbol{c_a} \right)^2 dt \\ &= \boldsymbol{c_e}^T \left(\boldsymbol{P_{ee}^{11}} + k_e \left(\boldsymbol{P_{ee}^{10}} + \boldsymbol{P_{ee}^{01}} \right) + k_e^2 \boldsymbol{P_{ee}^{00}} \right) \boldsymbol{c_e} + \frac{k_a^2}{V^2} \boldsymbol{c_a}^T \boldsymbol{P_{aa}^{00}} \boldsymbol{c_a} \\ &\quad -2 \boldsymbol{c_e}^T \frac{k_a}{V} \left(\boldsymbol{P_{ea}^{10}} + k_e \boldsymbol{P_{ea}^{00}} \right) \boldsymbol{c_a} \\ &= \boldsymbol{c_e}^T \boldsymbol{R_e} \boldsymbol{c_e} + \frac{k_a^2}{V^2} \boldsymbol{c_a}^T \boldsymbol{P_{aa}^{00}} \boldsymbol{c_a} - 2 \boldsymbol{c_e}^T \boldsymbol{M_{ea}} \boldsymbol{c_a}. \end{aligned}$$

where $\boldsymbol{P_{\alpha\beta}^{ij}} = \int \boldsymbol{B}_\alpha^{(i)}(t) \left(\boldsymbol{B}_\beta^{(j)}(t) \right)^T dt$, $i, j \in \{0, 1\}$ and $\alpha, \beta \in \{a, e\}$.
The global penalty is then defined by:

$$\begin{aligned} PEN &= \gamma_1 PEN_1 + \gamma_2 PEN_2 \\ &= \boldsymbol{c}^T \begin{pmatrix} \gamma_1 \boldsymbol{R_a} + \gamma_2 \frac{k_a^2}{V^2} \boldsymbol{P_{aa}^{00}} & -\gamma_2 \left(\boldsymbol{M_{ea}} \right)^T \\ -\gamma_2 \boldsymbol{M_{ea}} & \gamma_2 \boldsymbol{R_e} \end{pmatrix} \boldsymbol{c} \\ &= \boldsymbol{c}^T \boldsymbol{R}(k_a, k_e, V, \gamma_1, \gamma_2) \boldsymbol{c}. \end{aligned} \tag{5}$$

Perfusion ratio model

The differential equation that models the evolution of the perfusion ratio $x(t)$ (see Section 6 for more details) is:

$$\begin{cases} \frac{dx(t)}{dt} = -\beta x(t) + \alpha\beta, \\ x(0) = 0. \end{cases}$$

It is assumed that α and β are positive quantities. Under this differential equation model, the perfusion ratio $x(t)$ at time t is shown to be:

$$x(t) = \alpha \left(1 - \exp(-\beta t) \right).$$

For this model, we are interested by the estimation of $\boldsymbol{\theta}^T = (\alpha, \beta)$ from the set of observations $\{(t_k, y_k) : k = 1, \ldots, n\}$ where

$$y_k = x(t_k) + \epsilon_k.$$

The perfusion ratio is approximated by $\widetilde{x}(t) = \left(\boldsymbol{B}(t) \right)^T \boldsymbol{c}$ and the differential equation operator is defined by:

$$\begin{aligned} L_{\boldsymbol{\theta}} \left(\widetilde{x}(t) \right) &= \frac{d\widetilde{x}(t)}{dt} + \beta \widetilde{x}(t) - \alpha\beta \\ &= \left(\boldsymbol{B}^{(1)}(t) \right)^T \boldsymbol{c} + \beta \left(\boldsymbol{B}(t) \right)^T \boldsymbol{c} - \alpha\beta. \end{aligned}$$

The penaly term is then defined by:

$$\begin{aligned} PEN &= \gamma \int \left(\left(\boldsymbol{B}^{(1)}(t) \right)^T \boldsymbol{c} + \beta \left(\boldsymbol{B}(t) \right)^T \boldsymbol{c} - \alpha\beta \right)^2 dt \\ &= \gamma \left\{ \boldsymbol{c}^T \left(\boldsymbol{P^{11}} + \beta \left(\boldsymbol{P^{10}} + \boldsymbol{P^{01}} \right) + \beta^2 \boldsymbol{P^{00}} \right) \boldsymbol{c} - 2\boldsymbol{c}^T \alpha\beta \left(\boldsymbol{p^1} + \beta \boldsymbol{p^0} \right) + \mathcal{L}\alpha^2\beta^2 \right\} \\ &= \boldsymbol{c}^T \boldsymbol{R}(\alpha,\beta,\gamma)\, \boldsymbol{c} + 2\boldsymbol{c}^T \boldsymbol{r}(\alpha,\beta,\gamma) + l(\alpha,\beta,\gamma), \end{aligned} \tag{6}$$

where $\mathcal{L}$ is the length of the interval of integration, $\boldsymbol{P^{ij}} = \int \boldsymbol{B}^{(i)}(t) \left(\boldsymbol{B}^{(j)}(t) \right)^T dt$ and $\boldsymbol{p^i} = \int \boldsymbol{B}^{(i)}(t)\, dt$ $i,j \in \{0,1\}$

2.4 Penalty matrix in a general system of linear differential equations

We have seen in the three previous examples that the overall penalty term is always a quadratic form in the spline parameters. In this subsection, we give an automatic procedure for constructing of the penalty matrix $\boldsymbol{R}(\boldsymbol{\theta},\boldsymbol{\gamma})$, the penalty vector $\boldsymbol{r}(\boldsymbol{\theta},\boldsymbol{\gamma})$ and the penalty constant $l(\boldsymbol{\theta},\boldsymbol{\gamma})$.
As we work in the case where the function f is an affine transformation with respect to $\boldsymbol{x}$, the system of differential equations can be written as:

$$\left\{ \begin{array}{lcl} \frac{dx_1(t)}{dt} &=& \sum_{k=1}^d a_{1k} x_k(t) + b_1 \\ \vdots & & \\ \frac{dx_d(t)}{dt} &=& \sum_{k=1}^d a_{dk} x_k(t) + b_d \end{array} \right.$$

where a_{jk} is the k-th coefficient used in the j-th differential equation. If the state function $x_k(t)$ is not involved in the j-th differential equation, then $a_{jk} = 0$. Otherwise, it is a given fixed constant or a quantity to be estimated. The term b_j is constant or a quantity to be estimated for the j-th differential equation of the system.
The penalty term for the j-th differential equation is:

$$\begin{aligned} PEN_j &= \int \left\{ L_{j,\boldsymbol{\theta}}\left(\widetilde{x}(t)\right) \right\}^2 dt \\ &= \int \left(\frac{d\widetilde{x}_j(t)}{dt} - \sum_{k=1}^d a_{jk} \widetilde{x}_k(t) - b_j \right)^2 dt \\ &= \int \left(\boldsymbol{c}_j^T \boldsymbol{B}_j^{(1)}(t) - \sum_{k=1}^d a_{jk} \boldsymbol{c}_k^T \boldsymbol{B}_k(t) - b_j \right) \left(\left(\boldsymbol{B}_j^{(1)}(t) \right)^T \boldsymbol{c}_j - \sum_{k=1}^d a_{jk} \left(\boldsymbol{B}_k(t) \right)^T \boldsymbol{c}_k - b_j \right) dt. \end{aligned}$$

Denote by $\boldsymbol{P_{jk}^{ab}}$ the matrix and $\boldsymbol{p_j^a}$ the vector defined for all $j,k \in \{1,\ldots,d\}$ and $a,b \in \{0,1\}$ by:

$$\begin{aligned} \boldsymbol{P_{jk}^{ab}} &= \int \boldsymbol{B}_j^{(a)}(t) \left(\boldsymbol{B}_k^{(b)}(t) \right)^T dt, \\ \boldsymbol{p_j^a} &= \int \boldsymbol{B}_j^{(a)}(t)\, dt. \end{aligned}$$

Then the penalty term for the j-th differential equation is equal to:

$$\begin{aligned} PEN_j &= \boldsymbol{c_j}^T \boldsymbol{P_{jj}^{11}} \boldsymbol{c_j} - \boldsymbol{c_j}^T \left(a_{j1} \boldsymbol{P_{j1}^{10}}, a_{j2} \boldsymbol{P_{j2}^{10}}, \ldots, a_{jd} \boldsymbol{P_{jd}^{10}} \right) \boldsymbol{c} \\ &\quad - \boldsymbol{c}^T \begin{pmatrix} a_{j1} \boldsymbol{P_{1j}^{01}} \\ a_{j2} \boldsymbol{P_{2j}^{01}} \\ \vdots \\ a_{jd} \boldsymbol{P_{dj}^{01}} \end{pmatrix} \boldsymbol{c_j} + \boldsymbol{c}^T \begin{pmatrix} a_{j1}a_{j1} \boldsymbol{P_{11}^{00}} & a_{j1}a_{j2} \boldsymbol{P_{12}^{00}} & \ldots & a_{j1}a_{jd} \boldsymbol{P_{1d}^{00}} \\ a_{j2}a_{j1} \boldsymbol{P_{21}^{00}} & a_{j2}a_{j2} \boldsymbol{P_{22}^{00}} & \ldots & a_{j2}a_{jd} \boldsymbol{P_{2d}^{00}} \\ \vdots & \vdots & \vdots & \vdots \\ a_{jd}a_{j1} \boldsymbol{P_{d1}^{00}} & a_{jd}a_{j2} \boldsymbol{P_{d2}^{00}} & \ldots & a_{jd}a_{jd} \boldsymbol{P_{dd}^{00}} \end{pmatrix} \boldsymbol{c} \\ &\quad + 2b_j \boldsymbol{c}^T \begin{pmatrix} a_{j1} \boldsymbol{p_1^0} \\ a_{j2} \boldsymbol{p_2^0} \\ \vdots \\ a_{jd} \boldsymbol{p_d^0} \end{pmatrix} - 2b_j \left(\boldsymbol{p_j^1} \right)^T \boldsymbol{c_j} + b_j^2 \mathcal{L}. \end{aligned}$$

One can show that the overall penalty term $PEN = \sum_{j=1}^{d} \gamma_j PEN_j$ is equal to:

$$PEN = \boldsymbol{c}^T \boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma})\, \boldsymbol{c} + 2\boldsymbol{c}^T \boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + l(\boldsymbol{\theta}, \boldsymbol{\gamma}).$$

The matrix $\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a symetric block matrix of dimension $K \times K$ with (k,l)-th block of dimension $K_k \times K_l$ defined by:

$$\delta_{k=l}\gamma_k \boldsymbol{P_{kk}^{11}} - \gamma_k a_{kl} \boldsymbol{P_{kl}^{10}} - \gamma_l a_{lk} \boldsymbol{P_{lk}^{01}} + \sum_{i=1}^{d} \gamma_i a_{ik} a_{il} \boldsymbol{P_{kl}^{00}}$$

where $k, l \in \{1, \ldots, d\}$ and $\delta_{k=l} = 1$ if $k = l$ and 0 otherwise. The vector $\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a vector of length K that corresponds to the concatenation of vectors $\left(\sum_{j=1}^{d} \gamma_j b_j a_{jk} \boldsymbol{p_k^0} - \gamma_k b_k \boldsymbol{p_k^1}\right)_{k=1,\ldots,d}$. The penalty constant $l(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is equal to $\mathcal{L} \sum_{j=1}^{d} \gamma_j b_j^2$.

3 Posterior distribution and parameter estimation

In this section, we explain how to explore the joint posterior distribution using MCMC techniques.

3.1 Joint posterior distribution

In order to obtain the joint posterior distribution, it is necessary to derive the normalization constant of the prior distribution of the spline coefficients $\boldsymbol{c}$ in (3). Note that

$$\log(\pi(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\gamma})) \doteq -\frac{1}{2}\left\{\boldsymbol{c}^T \left(\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1}\right) \boldsymbol{c} - 2\boldsymbol{c}^T \left(-\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1} \boldsymbol{\mu_c}\right) + l(\boldsymbol{\theta}, \boldsymbol{\gamma})\right\}$$

is composed of two parts. The first corresponds to the log-density of a multivariate normal distribution with mean $\left(\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1}\right)^{-1} \left({\boldsymbol{\Sigma_c}}^{-1} \boldsymbol{\mu_c} - \boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma})\right)$ and variance-covariance matrix $\left(\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1}\right)^{-1}$. The second part is a constant term independent of $\boldsymbol{c}$. Therefore, the normalizing constant for the conditional prior density for the spline coefficients $\boldsymbol{c}$ is:

$$(\det(\boldsymbol{M_1}))^{\frac{1}{2}} \exp\left(\frac{1}{2} l(\boldsymbol{\theta}, \boldsymbol{\gamma})\right) \exp\left(-\frac{1}{2} {\boldsymbol{v_1}}^T {\boldsymbol{M_1}}^{-1} \boldsymbol{v_1}\right),$$

where $\boldsymbol{v_1} = -\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1} \boldsymbol{\mu_c}$ and $\boldsymbol{M_1} = \boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1}$.
We remind that $\mathcal{J}$ represents the subset of $\{1, \ldots, d\}$ of all indices of observed state functions. Using the model specification and the normalization constant in the conditional prior for $\boldsymbol{c}$, one obtains the log joint posterior density:

$$\begin{aligned}
\log(p(\boldsymbol{c}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}|\boldsymbol{y})) \;\doteq\; & \sum_{j \in \mathcal{J}} \left\{\frac{n_j}{2} \log(\tau_j) - \frac{\tau_j}{2} \left\|\boldsymbol{y}_j - \boldsymbol{B}_j \boldsymbol{c}_j\right\|^2\right\} \\
& -\frac{1}{2}\left\{\boldsymbol{c}^T \boldsymbol{M_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1} + {\boldsymbol{v_1}}^T {\boldsymbol{M_1}}^{-1} \boldsymbol{v_1} - \log(\det(\boldsymbol{M_1}))\right\} \\
& + \sum_{j \in \mathcal{J}} \left\{\left(a_{\tau_j} - 1\right) \log(\tau_j) - b_{\tau_j} \tau_j\right\} \\
& + \sum_{j=1}^{d} \left\{\left(a_{\gamma_j} - 1\right) \log(\gamma_j) - b_{\gamma_j} \gamma_j\right\} \\
& + \log(\pi(\boldsymbol{\theta})) \\
\doteq\; & \sum_{j \in \mathcal{J}} \left\{\frac{n_j}{2} \log(\tau_j) - \frac{\tau_j}{2} \boldsymbol{y}_j^T \boldsymbol{y}_j\right\} - \frac{1}{2}\left\{{\boldsymbol{v_1}}^T {\boldsymbol{M_1}}^{-1} \boldsymbol{v_1} - \log(\det(\boldsymbol{M_1}))\right\} \\
& -\frac{1}{2} \boldsymbol{c}^T \left(\operatorname{diag}(\boldsymbol{Z_j}, j = 1, \ldots, d) + \boldsymbol{M_1}\right) \boldsymbol{c} + \boldsymbol{c}^T \left(\boldsymbol{v_1} + \operatorname{vec}(\boldsymbol{z_j}, j = 1, \ldots, d)\right) \\
& + \sum_{j \in \mathcal{J}} \left\{\left(a_{\tau_j} - 1\right) \log(\tau_j) - b_{\tau_j} \tau_j\right\} \\
& + \sum_{j=1}^{d} \left\{\left(a_{\gamma_j} - 1\right) \log(\gamma_j) - b_{\gamma_j} \gamma_j\right\} \\
& + \log(\pi(\boldsymbol{\theta}))
\end{aligned} \tag{7}$$

where $\text{diag}\,(\boldsymbol{Z_j}, j = 1, \ldots, d)$ corresponds to a block-diagonal matrix with matrix $\boldsymbol{Z_j}$ that is equal to $\tau_j {\boldsymbol{B_j}}^T \boldsymbol{B_j}$ if the j-th state function is observed and the null $K_j \times K_j$-matrix otherwise. The vector $\text{vec}\,(\boldsymbol{z_j}, j = 1, \ldots, d)$ corresponds to the concatenation of vector $\boldsymbol{z_j}$ that is equal to $\tau_j {\boldsymbol{B_j}}^T \boldsymbol{y_j}$ if the j-th state function is observed and the null K_j-vector otherwise.

3.2 Exploring the joint posterior distribution using MCMC

Markov Chain Monte Carlo (MCMC) technique is a powerful method to generate samples from posterior distributions in a Bayesian framework.

3.2.1 Conditional posterior distributions

From expression (7), one can show that the conditional posterior distribution for the vector $\boldsymbol{c} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d)^T$ of spline coefficients is a multivariate normal distribution :

$$\boldsymbol{c}|\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{y} \sim \mathcal{N}\left({\boldsymbol{M_2}}^{-1}\boldsymbol{v_2}; {\boldsymbol{M_2}}^{-1}\right),$$

where $\boldsymbol{v_2} = \boldsymbol{v_1} + \text{vec}\,(\boldsymbol{z_j}, j = 1, \ldots, d)$ and $\boldsymbol{M_2} = \boldsymbol{M_1} + \text{diag}\,(\boldsymbol{Z_j}, j = 1, \ldots, d)$.
For each precision τ_j, $j \in \mathcal{J}$, the conditional posterior is a gamma distribution:

$$\tau_j|\boldsymbol{c}, \boldsymbol{y} \sim \mathcal{G}\left(\frac{n_j}{2} + a_{\tau_j}; \frac{\left\|\boldsymbol{y}_j - \boldsymbol{B}_j\boldsymbol{c}_j\right\|^2}{2} + b_{\tau_j}\right).$$

Unfortunately, the conditional posterior distribution for each ODE-adhesion parameter γ_j, $j = 1, \ldots, d$ is not of a familiar type. Therefore, a Metropolis-Hastings step was used to generate a random sample of γ_j from the conditional posterior distribution. Whatever the prior distribution for $\boldsymbol{\theta}$, the conditional posterior distribution for parameter $\boldsymbol{\theta}$ does not correspond to a common family due to $\boldsymbol{M_1}$, $\boldsymbol{v_1}$. Therefore, a Metropolis-Hastings step was also used to generate a random sample of θ from the conditional posterior distribution.

3.2.2 Marginal joint posterior distribution for $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$

In order to avoid being forced to deal with the strong posterior conditional correlation between the spline coefficients and the differential equation parameters, the joint posterior distribution can be marginalized with respect to the spline coefficients. The log of it can be shown to be:

$$\begin{aligned}
\log\left(p\left(\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}|\boldsymbol{y}\right)\right) \quad \doteq \quad & \sum_{j\in\mathcal{J}}\left\{\frac{n_j}{2}\log\left(\tau_j\right) - \frac{\tau_j}{2}\boldsymbol{y}_j^T\boldsymbol{y}_j\right\} \\
& -\frac{1}{2}\left\{{\boldsymbol{v_1}}^T{\boldsymbol{M_1}}^{-1}\boldsymbol{v_1} - \log\left(\det\left(\boldsymbol{M_1}\right)\right) - {\boldsymbol{v_2}}^T{\boldsymbol{M_2}}^{-1}\boldsymbol{v_2} + \log\left(\det\left(\boldsymbol{M_2}\right)\right)\right\} \\
& + \sum_{j\in\mathcal{J}}\left\{\left(a_{\tau_j} - 1\right)\log\left(\tau_j\right) - b_{\tau_j}\tau_j\right\} \\
& + \sum_{j=1}^{d}\left\{\left(a_{\gamma_j} - 1\right)\log\left(\gamma_j\right) - b_{\gamma_j}\gamma_j\right\} \\
& + \log\left(\pi\left(\boldsymbol{\theta}\right)\right)
\end{aligned}$$

As $\boldsymbol{M_1} = \boldsymbol{M_1}\left(\boldsymbol{\gamma}, \boldsymbol{\theta}\right)$, $\boldsymbol{v_1} = \boldsymbol{v_1}\left(\boldsymbol{\gamma}, \boldsymbol{\theta}\right)$, $\boldsymbol{M_2} = \boldsymbol{M_2}\left(\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{y}\right)$ and $\boldsymbol{v_2} = \boldsymbol{v_2}\left(\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{y}\right)$, the conditional distribution of $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ cannot be identified. Therefore, Metropolis-Hastings steps must be used to generate samples from $p\left(\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}|\boldsymbol{y}\right)$. When the MCMC-chains for parameters $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ have converged, one can generate a sample from the conditional posterior of $\boldsymbol{c}$ if desired, i.e. from the multivariate normal distribution with conditional mean ${\boldsymbol{M_2}}^{-1}\boldsymbol{v_2}$ and conditional variance-covariance ${\boldsymbol{M_2}}^{-1}$.

3.2.3 Metropolis-Hastings algorithm with adaptive proposals

In order to reduce the rejection rate and to improve the mixing in the exploration of $p\left(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\tau}|\boldsymbol{y}\right)$, we use for each component in $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ a Metropolis-Hasting step with adaptive proposals. We use the strategy developed by Cai

et al. (2008) which is based on parallel chains and proposal densities built from part of the already available sample from the posterior distribution.
For the proposal densities, two possibilities are left to the user: either use a mixture of triangular and exponential distributions or use a mixture of trapezoidal and exponential distributions. The first solution has the advantage to be simple and fast, but is limited by the acceptance rate (approximately 60%). The second one has an acceptance rate of about 80% but is slower due to the fact that the weights for the mixture depend on the posterior distribution. In practice, we use mixture of triangular and exponential distributions since it is less time consuming. More details are available in Cai et al. (2008).

3.2.4 Re-parametrization of $\boldsymbol{\theta}$ using rotation and translation

To accelerate the inference procedure, we also used the strategy proposed by Lambert (2007).
First run the Metropolis-Hastings algorithm with adaptive proposals for a few thousand iterations. Then, reparametrize the problem by applying a translation and a rotation to the parameter vector $\boldsymbol{\theta}$.
More precisely, denote by $\boldsymbol{S}$ the empirical variance-covariance matrix of the parameter $\boldsymbol{\theta}$ evaluated using the first iterations of the generated chains and $\bar{\boldsymbol{\theta}}$ the corresponding empirical mean vector. We reparametrize the posterior distribution using $\boldsymbol{\beta}$ where:

$$\boldsymbol{\theta} = \boldsymbol{L}\boldsymbol{\beta} + \bar{\boldsymbol{\theta}}$$

with $\boldsymbol{L}$ is the lower triangular matrix in the Choleski decomposition, i.e., the matrix $\boldsymbol{L}$ such that $\boldsymbol{S} = \boldsymbol{L}\boldsymbol{L}^T$. Then we use a Metropolis-Hastings algorithm with adaptive proposals to generate sample from the posterior distribution for $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$. A sample of posterior distribution for $\boldsymbol{\theta}$ is obtained using the previous formula. As before, when the MCMC-chains for these parameters have converged, we sample from the conditional posterior of $\boldsymbol{c}$, i.e. from the multivariate normal distribution with conditional mean ${\boldsymbol{M_2}}^{-1}\boldsymbol{v_2}$ and conditional variance-covariance ${\boldsymbol{M_2}}^{-1}$.

4 Hierarchical Bayesian Generalized Profiling Estimation for Linear Dynamic Systems Model

In this section, we introduce the concept of hierarchical Bayesian generalized profiling estimation for linear dynamic systems model. This hierarchical approach may be viewed as a simple generalization of the standard approach where the variability of the individual differential equation parameters is specified.

4.1 Individual system of differential equations

Assume now that states are observed comming from I subjects and that the changes in each state $\boldsymbol{x_i}(t) \in \mathbb{R}^d, i = 1, \ldots, I$ are governed by a linear system of differential equations:

$$D\boldsymbol{x_i}(t) = f(\boldsymbol{x_i}, t, \boldsymbol{\theta_i}), t \in [0; T_i],$$

where $\boldsymbol{\theta_i} \in \mathbb{R}^q$ is the unknown individual vector of ODE-parameters and f a known function, identical for each subject, that is supposed to be an affine transformation with respect to $\boldsymbol{x_i}$.
It is assumed, as in the standard case, that the same subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d states in $\boldsymbol{x_i}$ are observed at time point t_{ijk}, $i = 1, \ldots, I$, $j \in \mathcal{J}$, $k = 1, \ldots, n_{ij}$ with measurement error ϵ_{ijk}. We denote $y_{ijk} = \boldsymbol{x_i}(t_{ijk}) + \epsilon_{ijk}$ the corresponding measurement. As in the standard case, it is assumed that the measurement errors ϵ_{ijk} are independent and distributed according to a Gaussian distribution with mean 0 and with precision τ_j.

4.2 B-spline basis functions expansion and ODE-penalty

The approximation $\widetilde{\boldsymbol{x}}_{ij}(t)$ of the j-th state function for the i-th subject is expressed as a linear combination of B-spline basis functions:

$$\widetilde{\boldsymbol{x}}_{ij}(t) = (\boldsymbol{B_{ij}}(t))^T \boldsymbol{c_{ij}}$$

where $\boldsymbol{B_{ij}}(t)$ is the K_{ij}-vector of B-spline basis functions evaluated at time t and ${\boldsymbol{c_{ij}}}^T = (c_{ijk}; k = 1, \ldots, K_{ij})$ the K_{ij}-vector of B-spline coefficients.
The total ODE-penalty term is constructed in a three steps procedure. First the j-th penalty term for the i-th subject is constructed, then the total penalty term for the i-th subject is computed and finally, the overall penalty is obtained

by summing all the individual penalties.
In the first step, the proximity between the approximation of the state function $\widetilde{\boldsymbol{x}}_{ij}(t)$ and the solution $\boldsymbol{x}_{ij}(t)$ is given by:

$$PEN_{ij}(\widetilde{\boldsymbol{x}}_i) = \int \left\{L_{i,j,\boldsymbol{\theta_i}}(\widetilde{\boldsymbol{x}}_i)\right\}^2 dt,$$

where $L_{i,j,\boldsymbol{\theta_i}}(\widetilde{\boldsymbol{x}}_i) = D\widetilde{\boldsymbol{x}}_{ij}(t) - f_j(\widetilde{\boldsymbol{x}}_i, t, \boldsymbol{\theta_i})$ is the differential equation operator for the j-th state function of the i-th subject. Then the total penalty term for the i-th subject is given by:

$$\begin{aligned} PEN_i(\widetilde{\boldsymbol{x}}_i|\boldsymbol{\gamma}) &= \sum_{j=1}^{d} \gamma_j PEN_{ij}(\widetilde{\boldsymbol{x}}_i) \\ &= \boldsymbol{c_i}^T \boldsymbol{R_i}(\boldsymbol{\theta_i}, \boldsymbol{\gamma})\, \boldsymbol{c_i} + 2\boldsymbol{c_i}^T \boldsymbol{r_i}(\boldsymbol{\theta_i}, \boldsymbol{\gamma}) + l_i(\boldsymbol{\theta_i}, \boldsymbol{\gamma}). \end{aligned}$$

The vector $\boldsymbol{c_i}$ corresponds to the concatenation of all spline coefficients for the i-th subject. The matrix $\boldsymbol{R_i}(\boldsymbol{\theta_i}, \boldsymbol{\gamma})$, the vector $\boldsymbol{r_i}(\boldsymbol{\theta_i}, \boldsymbol{\gamma})$ and the constant $l_i(\boldsymbol{\theta_i}, \boldsymbol{\gamma})$ are constructed in the same way as the penalty terms presented in the standard case (see Section 2.4). As the same function f is common to all the systems of differential equations, the same ODE-adhesion parameter $\boldsymbol{\gamma}$ is used for all the individual penalties. Finally, the overall penalty term is the sum of all individual penalties:

$$\begin{aligned} PEN(\widetilde{\boldsymbol{x}}) &= \sum_{i=1}^{d} PEN_i(\widetilde{\boldsymbol{x}}_i|\boldsymbol{\gamma}) \\ &= \boldsymbol{c}^T \boldsymbol{R}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})\, \boldsymbol{c} + 2\boldsymbol{c}^T \boldsymbol{r}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma}) + l(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma}) \end{aligned}$$

where $\boldsymbol{c}$ is the vector of all spline coefficients and $\boldsymbol{R}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})$ is a block diagonal matrix with the individual penalty matrices $\boldsymbol{R_i}(\boldsymbol{\theta_i}, \boldsymbol{\gamma})$ on its diagonal, $\boldsymbol{r}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})$ concatenates the individual penalty vector $\boldsymbol{r_i}(\boldsymbol{\theta_i}, \boldsymbol{\gamma})$ and $l(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})$ is the sum of all individual penalty constants.

4.3 Bayesian hierarchical model

The Bayesian model is similar to the one used in individual approach, but with an additional specification: the inter-individual variability of the ODE parameters and the prior distributions of the population paramerters has to be specified.

$$\left\{ \begin{array}{rcl} y_{ijk}|\boldsymbol{c_{ij}}, \tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B_{ij}}(t_{ijk})\right)^T \boldsymbol{c_{ij}}; \tau_j^{-1}\right) \\ p(\boldsymbol{c}|\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma}) & \propto & \exp\left(-\frac{1}{2}\left\{PEN + \boldsymbol{c}^T {\boldsymbol{\Sigma_c}}^{-1} \boldsymbol{c} - 2\boldsymbol{c}^T {\boldsymbol{\Sigma_c}}^{-1} \boldsymbol{\mu_c}\right\}\right) \\ \gamma_j & \sim & \mathcal{G}\left(a_{\gamma_j}; b_{\gamma_j}\right) \\ \tau_j & \sim & \mathcal{G}\left(a_{\tau_j}; b_{\tau_j}\right) \\ \boldsymbol{\theta_i}|\boldsymbol{\theta}, \boldsymbol{P_\theta} & \sim & \mathcal{N}_q\left(\boldsymbol{\theta}; {\boldsymbol{P_\theta}}^{-1}\right) \\ \boldsymbol{\theta} & \sim & \pi(\boldsymbol{\theta}) \\ \boldsymbol{P_\theta} & \sim & \mathcal{W}_q\left(\boldsymbol{V}^{-1}; r\right) \end{array} \right.$$

The vector $\boldsymbol{\mu_c}$ and the matrix ${\boldsymbol{\Sigma_c}}^{-1}$ are used to express the possible uncertainty on the initial condition of the state functions. As in the standard approach, we have to note that the prior distribution for the spline coefficients is composed of two parts. The first corresponds to a multivariate Gaussian distribution with mean ${\boldsymbol{M_1}}^{-1}\boldsymbol{v_1}$ and variance-covariance matrix ${\boldsymbol{M_1}}^{-1}$ where $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c} - \boldsymbol{r}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \gamma)$ and $\boldsymbol{M_1} = \boldsymbol{R}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma}) + {\boldsymbol{\Sigma_c}}^{-1}$ and the second part corresponds to a term independent of $\boldsymbol{c}$. Therefore, the normalizing constant

$$\left(\det(\boldsymbol{M_1})\right)^{1/2} \exp\left(\frac{1}{2} l(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})\right) \exp\left(-\frac{1}{2} {\boldsymbol{v_1}}^T {\boldsymbol{M_1}}^{-1} \boldsymbol{v_1}\right),$$

has to be used to obtain the joint posterior distribution.
The prior distributions for the ODE-adhesion parameter $\boldsymbol{\gamma}$ and for the precision of measurement $\boldsymbol{\tau}$ are the same as those used in the standard approach. The inter-individual variability is expressed using a standard choice: it is supposed that each individual ODE-parameter $\boldsymbol{\theta_i}$ is distributed according to a multivariate Gaussian distribution with mean $\boldsymbol{\theta}$ and precision matrix $\boldsymbol{P_\theta}$. The prior distribution for the mean population parameter $\boldsymbol{\theta}$ is here specified by $\pi(\boldsymbol{\theta})$. Note that, if this prior distribution is a multivariate Gaussian distribution, then it corresponds to a conditional conjugate

prior distribution. The precision matrix $\boldsymbol{P_\theta}$ characterizes the variability of the individual ODE-parameters about the mean population ODE-parameter $\boldsymbol{\theta}$. The prior distribution for this precision matrix $\boldsymbol{P_\theta}$ is specified by a Wishart distribution. There, $\boldsymbol{V}$ corresponds to the prior expected value for the variance-covariance matrix of the random effects and r quantifies the confidence in that specific value, large values of r translating a large prior confidence. Note that r has to be grather than q.
The strategy used to generate a sample from the joint posterior distribution is similar to the one used in the standard approach, see Section 3.2.
First, one can show that the conditional posterior distribution for the vector $\boldsymbol{c}$ is a multivariate Gaussian distribution with mean $\boldsymbol{M_2}^{-1}\boldsymbol{v_2}$ and variance-covariance matrix $\boldsymbol{M_2}^{-1}$ where:

$$\begin{aligned}\boldsymbol{v_2} &= \boldsymbol{v_1} + \mathrm{vec}\left(\mathrm{vec}\left(\boldsymbol{z_{ij}}; j = 1, \ldots, d\right); i = 1, \ldots, I\right)\\ \boldsymbol{M_2} &= \boldsymbol{M_1} + \mathrm{diag}\left(\mathrm{diag}\left(\boldsymbol{Z_{ij}}; j = 1, \ldots, d\right); i = 1, \ldots, I\right).\end{aligned}$$

The K_{ij}-vector $\boldsymbol{z_{ij}}$ is equal to $\tau_j \boldsymbol{B_{ij}}^T \boldsymbol{y_{ij}}$ if the state function $x_{ij}(t)$ is observed and the null vector otherwise. The $K_{ij} \times K_{ij}$-matrix $\boldsymbol{Z_{ij}}$ is equal to $\tau_j \boldsymbol{B_{ij}}^T \boldsymbol{B_{ij}}$ if the state function $x_{ij}(t)$ is observed and the null matrix otherwise.
Then, the joint posterior distribution is marginalized with respect to the spline coefficients $\boldsymbol{c}$. The aim is to get rid of the inconvenient posterior conditional correlation between the generated spline coefficients and the individual differential equation parameters. One can show that the log of this marginal joint posterior distribution is equal to:

$$\begin{aligned}\log\left(p\left(\boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\theta}, \boldsymbol{P_\theta} | \boldsymbol{y}\right)\right) \doteq\ & \frac{1}{2}\sum_{i=1}^{I}\sum_{j\in\mathcal{J}}\left\{n_{ij}\log\left(\tau_j\right) - \tau_j \boldsymbol{y_{ij}}^T \boldsymbol{y_{ij}}\right\}\\ & -\frac{1}{2}\left\{\boldsymbol{v_1}^T \boldsymbol{M_1}^{-1} \boldsymbol{v_1} - \log\left(\det\left(\boldsymbol{M_1}\right)\right) - \boldsymbol{v_2}^T \boldsymbol{M_2}^{-1} \boldsymbol{v_2} + \log\left(\det\left(\boldsymbol{M_2}\right)\right)\right\}\\ & +\sum_{j=1}^{d}\left\{\left(a_{\gamma_j} - 1\right)\log\left(\gamma_j\right) - b_{\gamma_j}\gamma_j\right\}\\ & +\sum_{j\in\mathcal{J}}\left\{\left(a_{\tau_j} - 1\right)\log\left(\tau_j\right) - b_{\tau_j}\tau_j\right\}\\ & +\frac{I}{2}\log\left(\det\left(\boldsymbol{P_\theta}\right)\right) - \frac{1}{2}\sum_{i=1}^{I}\left(\boldsymbol{\theta_i} - \boldsymbol{\theta}\right)^T \boldsymbol{P_\theta}\left(\boldsymbol{\theta_i} - \boldsymbol{\theta}\right)\\ & +\log\left(\pi\left(\boldsymbol{\theta}\right)\right)\\ & +\frac{r - q - 1}{2}\log\left(\det\left(\boldsymbol{P_\theta}\right)\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}\boldsymbol{P_\theta}\right)\end{aligned}$$

In that marginal joint posterior distribution, only the conditional posterior for the precision matrix $\boldsymbol{P_\theta}$ is of a familar type:

$$\boldsymbol{P_\theta} | \boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\theta}, \boldsymbol{y} \sim \mathcal{W}_q\left(\left(V + \sum_{i=1}^{I}\left(\boldsymbol{\theta_i} - \boldsymbol{\theta}\right)\left(\boldsymbol{\theta_i} - \boldsymbol{\theta}\right)^T\right)^{-1}; r + I\right).$$

The other conditional posterior distributions cannot be identified: Metropolis-Hastings steps with adaptive proposals have been used in practice to generate samples from $p\left(\boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\theta} | \boldsymbol{P_\theta}, \boldsymbol{y}\right)$. If desired, when the MCMC-chains for parameters $\boldsymbol{\tau}$, $\boldsymbol{\gamma}$, $\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}$, $\boldsymbol{\theta}$ and $\boldsymbol{P_\theta}$ have converged, one can generate a sample from the conditional posterior distribution of $\boldsymbol{c}$.

5 Simulation study

In this part, we perform some simulations on the first two examples presented in Section 2.3. The aim is to compare the performances of different estimation strategies.

5.1 Preliminary remarks

In order to compare the proposed Bayesian general profiling estimation (BGPE) method, the frequentist general profiling estimation (FGPE) method (Ramsay et al., 2007), the standard Bayesian (BES) and frequentist (FES) methods using

the explicit solution to the system of differential equations, it is necessary to explain how the initial conditions of the state functions were specified in each case.
For the FGPE method, the first component of each vector of splines coefficients $\boldsymbol{c_j}, j = 1, \ldots, d$ is fixed to the initial condition of the corresponding state function. The other spline parameters are estimated using criterion J, see Eq. (2), modified in order to take into account that the first component is fixed. A search grid is done for each ODE-adhesion parameter $\boldsymbol{\gamma}$, starting from small values and gradually increasing. The selected $\boldsymbol{\gamma}$ corresponds to the minimizer of the BIC criterion. Note that the method proposed in Ye (1998) was used to compute the number of effective parameters at each grid point. Concerning the BGPE method, the model described in Section 2.2 is used (i.e. including $\boldsymbol{\mu_c}$ and ${\boldsymbol{\Sigma_c}}^{-1}$). The precision parameters in ${\boldsymbol{\Sigma_c}}^{-1}$ corresponding to the initial condition of the state function in $\boldsymbol{\mu_c}$ is set to infinity in order to force equality of the approximation of the state function to the initial condition at $t = 0$. For the BGPE and BES methods, the posterior mean, 80% and 95% credibility intervals and the posterior empirical variance-covariance matrix of PK parameters are estimated from the generated chains. For the FGPE method, the variance-covariance matrix of the estimated parameter is calculated using the approximation given in Ramsay et al. (2007). Confidence intervals at 80% and 95% levels are computed using Gaussian quantiles. For the FES method, the variance-covariance matrix of the parameter estimates is based on the hessian matrix and confidence intervals at 80% and 95% levels are computed using gaussian quantiles.

5.2 One-compartment model with single IV bolus

Data

Each sample corresponds to 21 measurements made on a single subject at equidistant time points between 0h and 1h. At each time point t_k, the concentration y_k is generated using the explicit solution of the one-compartment model with single IV bolus injection and some measurement error:

$$y_k = \frac{D}{V} \exp\left(-k_e t_k\right) + \epsilon_k$$

where $D = 20mg$, $k_e = 5h^{-1}$, $V = 5L$ and $\epsilon \sim \mathcal{N}\left(0; \sigma^2\right)$. We consider three different values for the standard deviation σ: $\sigma = 0.1$, 0.3 and 0.5. These three levels correspond to a low, medium and large level of noise in the data measurement. For each configuration, 1000 datasets were generated.

Models

For the FGPE and the BGPE methods, the concentration $C(t)$ of drug is approximated by $\widetilde{C}(t) = \frac{1}{V}\left(\boldsymbol{B}(t)\right)^T \boldsymbol{c}$ using a 5-order B-spline expansion with inner knots at each tenth between 0 and 1.
For the FGPE method, the first spline coefficient c_1 is fixed to the dose D. PK-parameters k_e and V, the spline coefficients $\boldsymbol{c}_{2:K}$ and precision τ of measurements are estimated by minimizing the criterion presented in Eq. (2).
For the BGPE approach, the following Bayesian model is considered:

$$\left\{\begin{array}{rcll} y_k|\boldsymbol{c}, \tau, V & \sim & \mathcal{N}\left(\frac{1}{V}\left(\boldsymbol{B}(t_k)\right)^T \boldsymbol{c}; \tau^{-1}\right) & \forall k \in \{1, \ldots, n\} \\ \pi\left(\boldsymbol{c}|\gamma, k_e, V\right) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{M_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) & \\ \gamma & \sim & \mathcal{G}\left(a_\gamma; b_\gamma\right) & \\ \tau & \sim & \mathcal{G}\left(a_\tau; b_\tau\right) & \\ \pi\left(k_e, V\right) & \propto & \mathbb{I}\left\{k_e > 0, V > 0\right\} & \end{array}\right.$$

where $\boldsymbol{M_1} = \boldsymbol{R}(k_e, V, \gamma) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Eq. (4) for the specification of $\boldsymbol{R}(k_e, V, \gamma)$). The first component of $\boldsymbol{\mu_c}$ is set equal to D and the corresponding precision in ${\boldsymbol{\Sigma_c}}^{-1}$ to 10^{12}. The other components in $\boldsymbol{\mu_c}$ and ${\boldsymbol{\Sigma_c}}^{-1}$ are null.
For the FES and BES methods, the concentration of drug has an explicit expression: $C(t) = \frac{D}{V}\exp\left(-k_e t\right)$. For the FES method, PK-parameters k_e and V and precision τ of measurements are estimated by maximizing the likelihood. For the BES approach, we consider the following Bayesian model:

$$\left\{\begin{array}{rcll} y_k|\tau, k_e, V & \sim & \mathcal{N}\left(C\left(t_k\right); \tau^{-1}\right) & \forall k \in \{1, \ldots, n\} \\ \tau & \sim & \mathcal{G}\left(a_\tau; b_\tau\right) & \\ \pi\left(k_e, V\right) & \propto & \mathbb{I}\left\{k_e > 0, V > 0\right\} & \end{array}\right.$$

Simulation results

Table 1 summarizes the results of the simulation study. It gives estimation for the relative bias (in percent), the coverage probabilities at level 80% and 95% and the root mean squared error for volume V and constant of elimination k_e for the three different values of σ for the FGPE, FES, BGPE and BES methods.
For volume V, the relative bias seems to be the same for all methods. For medium and large value of σ, BGPE and BES methods slightly over-estimated it. For the FES, BGPE and the BES approaches, root mean squared error for V are similar, suggesting a smaller variability in Bayesian estimates. For the FGPE method, the number of outliers for the relative error is high, especially for $\sigma = 0.5$ that is translated in a larger RMSE. The estimated coverage probabilities for FGPE are well below nominal values, whereas for other approaches, the nominal levels are almost reached: this may be due to the approximation of the variance of the parameter estimate that is made in the FGPE approach.
Whatever σ, the relative bias for the constant of elimination k_e is small for all four methods except for BGPE and BES when $\sigma = 0.5$ (relative bias superior at 5%). RMSE for BES and BGPE approaches is similar to the FES one suggesting again a smaller dispersion of the Bayesian estimates. The root mean squared error in the FGPE approach is markedly larger than in other approaches essentially due to a large number of outliers. Coverage probabilities for FES and BES credible regions are in agreement with their nominal values. For BGPE and FGPE approaches, the estimated coverage probabilities are larger than expected, especially for FGPE approach.

		$\sigma = 0.1$ ($\tau = 100$)				$\sigma = 0.3$ ($\tau \approx 11.11$)				$\sigma = 0.5$ ($\tau = 4$)			
		RB	CP 80%	CP 95%	RMSE	RB	CP 80%	CP 95%	RMSE	RB	CP 80%	CP 95%	RMSE
$V = 5L$	FGPE	-0.0	43.3	58.4	0.11	0.1	42.1	60.3	0.33	1.0	44.0	58.5	0.56
	FES	-0.0	77.0	**93.8**	0.10	0.1	**79.1**	92.6	0.30	0.5	**79.0**	92.8	0.50
	BGPE	0.1	**78.7**	**95.2**	0.10	0.8	**80.9**	**93.9**	0.31	2.9	**80.1**	**94.1**	0.56
	BES	0.1	**78.9**	**95.6**	0.10	0.9	**81.2**	**94.2**	0.31	2.9	**80.1**	**94.6**	0.55
$k_e = 5h^{-1}$	FGPE	-0.2	94.6	99.2	0.30	-0.8	93.7	98.9	0.61	-2.4	90.7	97.3	1.02
	FES	0.2	**77.6**	93.4	0.16	0.4	**78.1**	93.1	0.49	1.7	**78.4**	92.8	0.83
	BGPE	0.4	84.7	97.4	0.17	2.3	85.0	97.1	0.52	7.3	82.9	96.7	0.98
	BES	0.3	**79.0**	**94.9**	0.16	1.6	**80.8**	**94.2**	0.50	5.3	**79.5**	**95.1**	0.90

Table 1: One-compartment model with IV bolus - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP 80%), 90% (CP %) and root mean squared error (RMSE) for FGPE, FES, BGPE and BES methods for the three different levels of standard deviation σ. Estimated coverage probabilities in bold characters are in line with their nominal values.

5.3 Two-compartment model

Data

Each sample corresponds to 81 measurements made on a single subject at equidistant time points between 0h and 2h. At each time point t_k, the concentration y_k is generated using the explicit solution of the two-compartments model with oral dosing and some measurement error:

$$y_k = \frac{D}{V}\frac{k_a}{k_a - k_e}\left(\exp\left(-k_e t_k\right) - \exp\left(-k_a t_k\right)\right) + \epsilon_k$$

where $D = 20mg$, $k_a = 8h^{-1}$, $k_e = 2h^{-1}$, $V = 1L$ and $\epsilon \sim \mathcal{N}\left(0; \sigma_e^2\right)$. We consider three different cases for the standard deviation σ_e: $\sigma_e = 0.3$, 0.6 and 0.9. These three levels correspond to low, medium and large levels of noise in the data measurement. In each configuration, 1000 datasets were generated.

Models

For the FGPE and the BGPE methods, we approximate the quantity $Q_a(t)$ of drug in the stomach by $\widetilde{Q_a}(t) = \left(\boldsymbol{B_a}(t)\right)^T \boldsymbol{c_a}$ and the concentration $C_e(t)$ of drug in the central compartment by $\widetilde{C_e}(t) = \left(\boldsymbol{B_e}(t)\right)^T \boldsymbol{c_e}$ using a 5-order B-spline expansion with inner knots at each tenth between 0 and 2.
For the FGPE method, the first spline coefficient c_{a1} is fixed to the dose D and c_{e1} is fixed to 0. PK-parameters k_a, k_e and V, the spline coefficients $\boldsymbol{c_a}_{2:K_a}$ and $\boldsymbol{c_e}_{2:K_e}$ and precision τ of measurements are estimated by minimizing the criterion J presented in Eq. (2).
The Bayesian model of the BGPE approach is:

$$\begin{cases} y_k|\boldsymbol{c_e}, \tau_e & \sim & \mathcal{N}\left(\left(\boldsymbol{B_e}(t_k)\right)^T \boldsymbol{c_e}; \tau_e^{-1}\right) & \forall k \in \{1, \ldots, n\} \\ \pi\left(\boldsymbol{c}|\gamma_1, \gamma_2, k_a, k_e, V\right) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{M_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) & \\ \gamma_1 & \sim & \mathcal{G}\left(a_{\gamma_1}; b_{\gamma_1}\right) & \\ \gamma_2 & \sim & \mathcal{G}\left(a_{\gamma_2}; b_{\gamma_2}\right) & \\ \tau_e & \sim & \mathcal{G}\left(a_{\tau_e}; b_{\tau_e}\right) & \\ \pi\left(k_a, k_e, V\right) & \propto & \mathbb{I}\left\{k_a > k_e > 0, V > 0\right\} & \end{cases}$$

where $\boldsymbol{M_1} = \boldsymbol{R}\left(k_a, k_e, V, \gamma_1, \gamma_2\right) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Eq. (5) for the expression of $\boldsymbol{R}\left(k_a, k_e, V, \gamma_1, \gamma_2\right)$). The first component of $\boldsymbol{\mu_c}$ is set equal to D, the $(K_a + 1)$-th component of $\boldsymbol{\mu_c}$ to 0 and the corresponding precision in ${\boldsymbol{\Sigma_c}}^{-1}$ to 10^{12}. The other components in $\boldsymbol{\mu_c}$ and ${\boldsymbol{\Sigma_c}}^{-1}$ are set to zero.
For the FES and BES methods, the concentration of drug in the central compartment has an explicit expression: $C_e(t) = \frac{D}{V}\frac{k_a}{k_a - k_e}\left(\exp\left(-k_e t\right) - \exp\left(-k_a t\right)\right)$. For the FES method, PK-parameters k_a, k_e and V and precision τ_e of measurements are estimated by maximizing the likelihood. For the BES approach, we consider the following Bayesian model:

$$\begin{cases} y_k|\tau_e, k_a, k_e, V & \sim & \mathcal{N}\left(C_e(t_k); \tau_e^{-1}\right) & \forall k \in \{1, \ldots, n\} \\ \tau_e & \sim & \mathcal{G}\left(a_{\tau_e}; b_{\tau_e}\right) & \\ \pi\left(k_a, k_e, V\right) & \propto & \mathbb{I}\left\{k_a > k_e > 0, V > 0\right\} & \end{cases}$$

Simulation results

Table 2 summarizes results the simulation study. It gives relative bias, the coverage probabilities at level 80% and 95% and root mean squared error for the constant of absorption k_a, the constant of elimination k_e and the volume V for the three different values of σ_e for the FGPE, FES, BGPE and BES methods.
For the constant of absorption k_a, in term of relative bias, all methods perform equally well when the level of noise is low. For large levels of noise, FGPE approach performs poorly: the relative bias is slightly positive with dispersion greater than for other methods (see root mean squared error). Coverage probabilities of the estimators are in agreement with the nominal values, except for the FGPE approach where it tends to be lower than the targeted value: this could be due to a bad approximation of the variance of the estimate, a bad choice of the distribution of estimator and possible adverse selection of the ODE-adhesion parameters.
For the constant of elimination k_e, performances are similar for the four methods in term of RMSE. Like in the first simulation study, some bias is introduced by the Bayesian methods. It is compensated by a smaller dispersion of the estimators. Coverage probabilities for FES, BGPE and BES approaches are similar and in agreement with their

nominal values. For the FGPE approach, the estimated coverage probabilities are below nominal value.
For the volume V of the central compartment, all methods perform equally in term of RMSE. BGPE and BES approachs under-estimate a little the volume. This result was expected due to the fact that the volume and the constant of elimination are negatively correlated. For the FGPE approach, estimated coverage probabilities are again below nominal levels. Again, for FES, BGPE and BES approaches, estimated coverage probabilities are almost in agreement with their nominal values.

		$\sigma_e = 0.3$ ($\tau_e \approx 11.11$)				$\sigma_e = 0.6$ ($\tau_e \approx 2.78$)				$\sigma_e = 0.9$ ($\tau_e \approx 1.23$)			
		RB	CP 80%	CP 95%	RMSE	RB	CP 80%	CP 95%	RMSE	RB	CP 80%	CP 95%	RMSE
$k_a = 8h^{-1}$	FGPE	0.5	74.7	90.4	0.43	2.0	72.9	91.9	0.95	4.4	**77.9**	**94.1**	1.53
	FES	0.2	**80.3**	**93.8**	0.36	0.6	**80.0**	**94.4**	0.72	0.4	**80.3**	**94.8**	1.08
	BGPE	0.3	**79.8**	**94.8**	0.36	0.6	**79.0**	**93.8**	0.75	-0.1	**78.4**	93.1	1.15
	BES	0.2	**80.6**	**94.6**	0.36	0.4	**80.3**	**94.5**	0.75	-0.3	**79.5**	**94.1**	1.16
$k_e = 2h^{-1}$	FGPE	-0.0	73.2	90.4	0.68	-0.2	71.1	91.6	0.14	-0.1	76.9	92.1	0.21
	FES	0.0	**79.1**	**93.7**	0.57	0.2	**78.8**	**94.7**	0.11	1.1	**82.3**	**95.2**	0.17
	BGPE	0.0	**78.6**	**94.3**	0.58	0.9	**79.3**	**94.6**	0.12	3.4	**78.5**	**94.2**	0.21
	BES	0.0	**79.5**	**94.2**	0.85	1.0	**80.2**	**95.2**	0.12	3.5	**79.2**	**94.4**	0.21
$V = 1L$	FGPE	0.1	73.7	90.1	0.03	0.4	71.9	91.8	0.05	0.7	77.1	92.6	0.08
	FES	0.0	**79.4**	93.4	0.02	0.1	**80.2**	**94.6**	0.04	-0.5	**80.3**	**96.0**	0.07
	BGPE	-0.1	**79.2**	**94.3**	0.02	-0.2	**79.6**	**94.4**	0.05	-1.5	77.0	**94.2**	0.07
	BES	-0.1	**80.0**	**94.4**	0.02	-0.3	**80.6**	**95.2**	0.05	-1.5	**78.4**	**94.7**	0.07

Table 2: Two-compartment models - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP 80%), 90% (CP %) and root mean squared error (RMSE) for FGPE, FES, BGPE and BES methods for the three different levels of standard deviation σ_e. Estimated coverage probabilities in bold characters are in line with their nominal values.

6 Application

6.1 Data

Arteriogenesis is important for the prevention and the recovery of tissue ischemia caused by arterial occlusive disease. Several studies can be made to study this phenomenon. In one of them (van Weel et al., 2007), ischemia on the left hind limb of mice was induced by electrocoagulation of the left common femoral artery proximal to the bifurcation of superficial and deep femoral artery. The blood flow was measured at baseline, directly after the femoral artery occlusion and over the 28 next days on the two hind limbs. The perfusion ratio $x(t)$ is expressed as the ratio of the ischemic to the non-ischemic blood flow.
In the study of van Weel et al. (2007), 113 mice were enrolled. We focuss on the specific C57BL/6 mice's group to illustrate the Bayesian generalized profiling estimation for ODE in the hierarchical case. In this specific group, perfusion ratio was measured 6 times over a 28 days period in 12 mice after a femoral artery occlusion (see Fig. 2). Originally, this dataset was analised using a nonlinear model that assumes an exponential approach to an asymptote whose differential equation model and explicit solution are presented in Section 2.3.

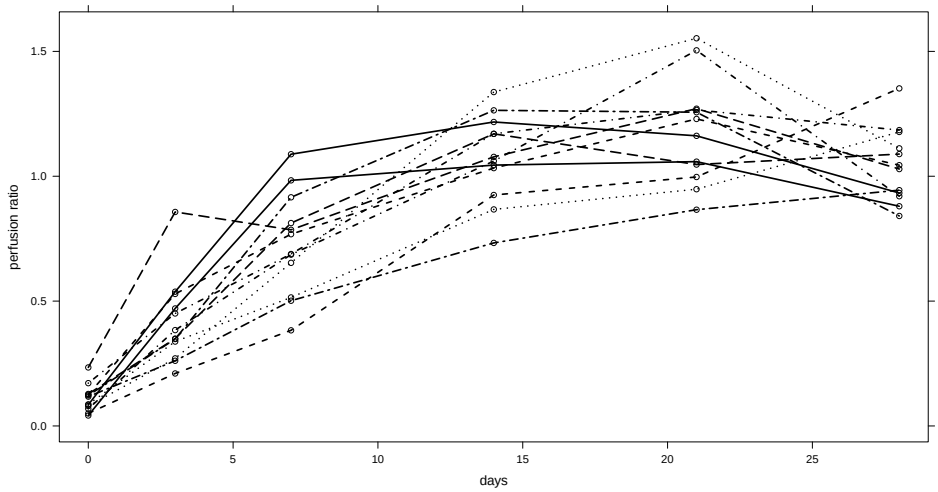


Figure 2: Ischemic/non-ischemic perfution ratio.

6.2 Model

The perfusion ratio $x_i(t)$ for the subject i is approximated by $\widetilde{x}_i(t) = (\boldsymbol{B}(t))^T \boldsymbol{c_i}$ using 5-order B-spline expansion with 57 equidistant knots between 0 and 28 days. Note that the same B-spline basis is used for all the subject and that only the spline coefficients are subject-specific. The Bayesian model used to analyze this data is:

$$\begin{cases} y_{ik}|\boldsymbol{c_i},\tau & \sim & \mathcal{N}\left((\boldsymbol{B}(t_{ik}))^T \boldsymbol{c_i};\tau^{-1}\right) & \forall i \in \{1,\dots,I\} \quad \forall k \in \{1,\dots,n\} \\ \pi(\boldsymbol{c_i}|\gamma,\alpha,\beta_i) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c_i}^T \boldsymbol{M_{i1}} \boldsymbol{c_i} - 2\boldsymbol{c_i}^T \boldsymbol{v_{i1}}\right)\right) & \\ \gamma;\tau & \sim & \mathcal{G}\left(1;10^{-8}\right) & \\ l\alpha;l\beta & \sim & \mathcal{N}\left(0;10^{8}\right) & \\ l\beta_i|l\beta,\tau_{l\beta} & \sim & \mathcal{N}\left(l\beta;\tau_{l\beta}^{-1}\right) & \forall i \in \{1,\dots,I\} \\ \tau_{l\beta} & \sim & \mathcal{G}\left(1;10^{-8}\right) & \end{cases}$$

where $\boldsymbol{M_{i1}} = \boldsymbol{R}(\alpha,\beta_i,\gamma) + {\boldsymbol{\Sigma_{c_i}}}^{-1}$ and $\boldsymbol{v_{i1}} = -\boldsymbol{r}(\alpha,\beta_i,\gamma) + {\boldsymbol{\Sigma_{c_i}}}^{-1}\boldsymbol{\mu_{c_i}}$ (see Section 2.3 for the expression of matrix $\boldsymbol{R}(\alpha,\beta_i,\gamma)$ and vector $\boldsymbol{r}(\alpha,\beta_i,\gamma)$). We fix the first component of each $\boldsymbol{\mu_{c_i}}$ to 0.25, i.e. to the empirical mean of the ischemic/non-ischemic perfusion ratio measured just after the surgery on all the mices except those from the group C57BL/6 and the corresponding precision in ${\boldsymbol{\Sigma_{c_i}}}^{-1}$ to 64, i.e. to the inverse of the empirical variance of the same measured perfusion ratio. The other components in $\boldsymbol{\mu_{c_i}}$ and ${\boldsymbol{\Sigma_{c_i}}}^{-1}$ are set to zero. This choice allows us to express

the fact that the artery was probably not completely occluded just after the surgery.
The model was selected from four possible models considering parameters $l\alpha$ and $l\beta$ either as a fixed parameter or as a random effect using the penalized deviance information criterion (Spiegelhalter et al., 2002).

6.3 Results

For the Metropolis-Hastings algorithms with adaptive proposals, we generate 10 parallel chains (10 parallel chains is the minimum recommended in Cai et al. (2008) in the case of mixture of triangular and exponential distributions for the adaptive proposals), with 1000 MCMC iterations and discard the first 300 realizations. We use this first sample to reparametrize the problem by applying a translation and a rotation to the parameter vector $\boldsymbol{\theta} = (l\alpha, l\beta)^T$. Then 7500 MCMC iterations were generated and 5000 were discarded. Figure 3 (resp. figure 4) shows traces and histograms of $\sigma = 1/\sqrt{\tau}$, $\log_{10}(\gamma)$ and $\tau_{l\beta}$ (resp. $\log(\alpha)$ and $\log(\beta)$). For each parameter chains, the Gelman-Rubin statistic was computed and was less than reference value 1.05 suggesting convergence of the MCMC algorithm.
Table 3 reports the posterior mean, 2.5%, 50% and 97.5% quantile of the ODE parameters, precision of measurement and ODE-adhesion parameter. The posterior 2.5% quantile for the parameter $l\alpha$ seems to mean that the perfusion ratio will tend to a value above 1 after the surgery. The posterior probability that $l\alpha$ is greater than 0 is virtually equal to 1 : arteriogenesis seems to have occurred. The increase of the perfusion ratio is defined, in the log scale, by $l\alpha + l\beta$. This increase is subject-specific due to the individual parameters $l\beta_i$. The precision $\tau_{l\beta}$ of random effect suggests a high variability of the recovery of the ischemia.
Figure 5 shows in (a) the estimated perfusion ratio with 95% credibility sets for the subject 8, in (b) the credibility interval for the posterior predictive distribution of the perfusion ratio for the same subject, in (c) the mean population perfusion ratio with 95% credibility sets and in (d) the credibility interval for the posterior predictive distribution of the mean population perfusion ratio. To obtain figure 5 (a), the conditional mean $\widetilde{x_8}(t)$ of the perfusion ratio for the subject 8 is reconstructed on a grid of time for each realization of the posterior distribution of the spline coefficients for this subject. Then, at each point of the grid of time, 2.5%, 50% and 97.5% quantiles are computed. On the same grid of time, the posterior preditive distribution of the perfusion ratio is computed by sampling from a Gaussian distribution with mean $\widetilde{x_8}(t)$ and variance τ^{-1} at each realization from the posterior distribution. Figure 5 (b) is then obtained by reporting the 2.5%, 50% and 97.5% quantiles as function of the time. For the population figures 5 (c) and (d), the procedure is the same as above but we have first to sample from a multivariate Gaussian distribution with mean $\boldsymbol{M_1}^{-1}\boldsymbol{v_1}$ and variance $\boldsymbol{M_1}^{-1}$ to obtain a sample of the posterior distribution of the mean population spline coefficients.
For each mice, the posterior probability that individual mean perfusion ratio is greater than 1 is higher 0.95, except for subject 12 where this probability is around 0.8. At the end of the study, i.e. at 28 days, the posterior probability that the mean population perfusion ratio is greater than 1 is higher than 0.95. The 95% credibility interval of the posterior predictive distribution is $(0.82; 1.49)$ and the posterior predictive probability that a perfusion ratio is higher that 1 after 28 days is equal to 0.8: this suggests recovery of mice from group C57BL/6 in 80% of the case after 28 days.

7 Discussion

We have presented a full Bayesian model based on ODE-penalized B-spline approach to jointly estimate the state functions and parameters of a system of affine differential equations. Simulations suggest that the proposed Bayesian ODE-penalized B-spline approach is superior of the frequentist one in term of estimation of the ODE parameters, coverage probabilities and prediction. It also provides similar results to the frequentist and Bayesian approach based on the explicit solution to the ODE.
We have also seen that the basic Bayesian ODE-penalized B-spline approach was simply extended to a hierarchical setting only by specifying the inter-individual variability of the ODE parameters. This hierarchical model offers the advantage to fit individual state functions for each subject.
The possible use of prior information on the ODE parameters and on the precision of measurement is a definite advantage. The proposed Bayesian model also enables to include uncertainty with respect to the initial condition of the state functions. This was not the case in the corresponding frequentist approach where the initial condition must either be fixed or estimated.
Contrary to the frequentist approach, the selection of the ODE-adhesion parameter is now automatic and does not depend on the choice of a specific criterion (e.g. BIC) or on the strategy used to compute the number of effective parameters: the selection of the ODE-adhesion parameter is include directly in the model.
In addition, as we work in a Bayesian setting, uncertainty measures are readily available through the credibility sets computed from the MCMC chains: the variance of parameter estimates is no more to approximate and confidence

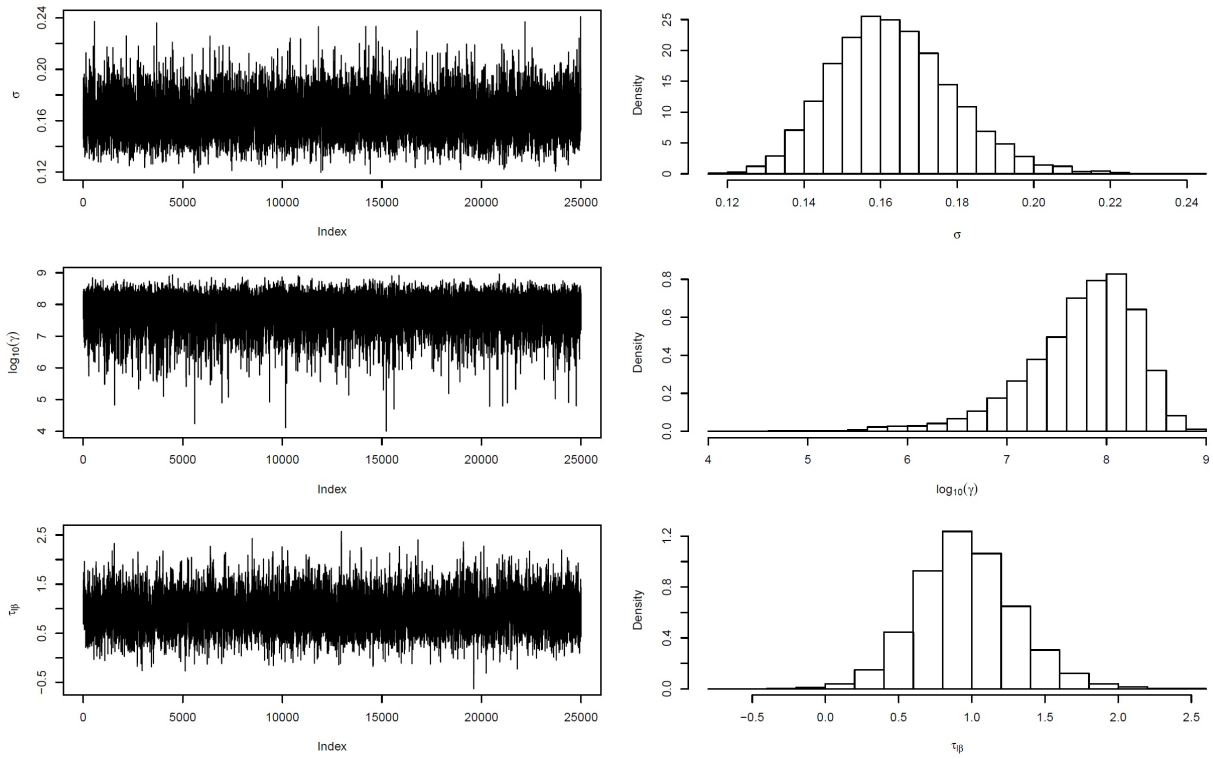


Figure 3: Traces and histograms for σ, $\log_{10}(\gamma)$ and $\tau_{l\beta}$

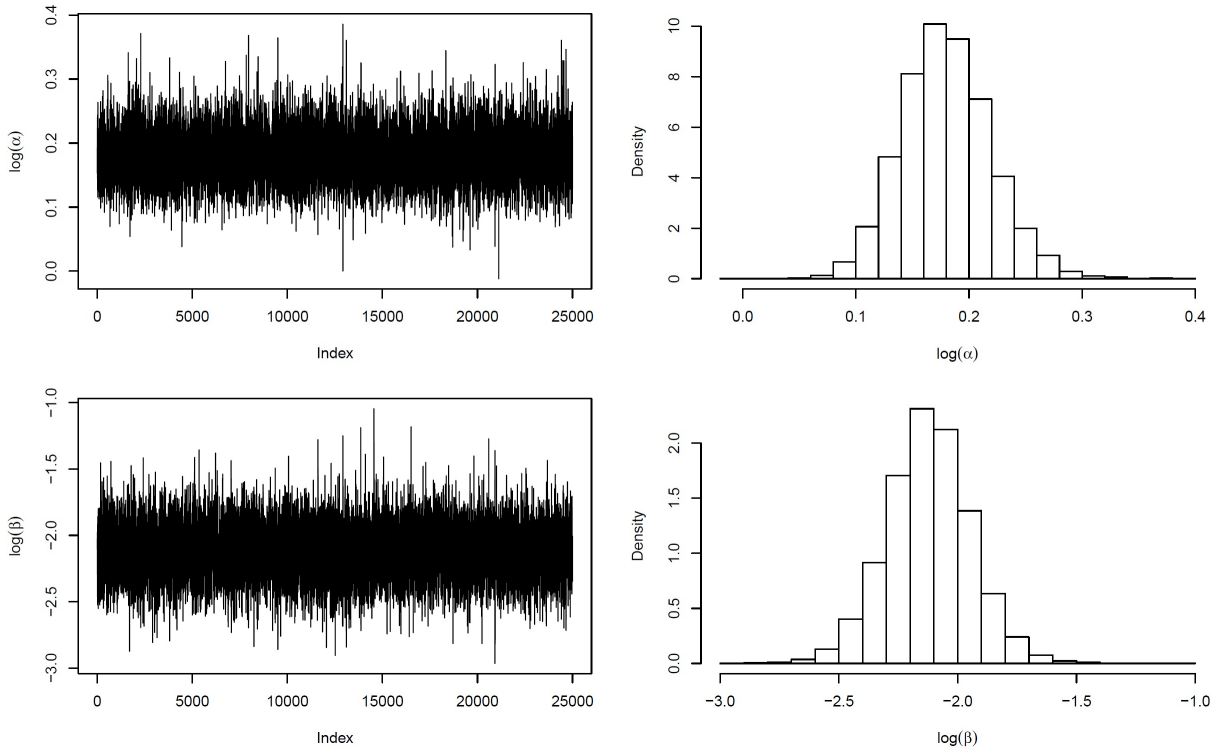


Figure 4: Traces and histograms for $\log(\alpha)$ and $\log(\beta)$

Parameters	Mean	2.5% quantile	50% quantile	97.5% quantile
$\log_{10}(\gamma)$	7.740	6.388	7.834	8.565
τ	38.50	25.515	38.043	54.545
$l\alpha$	0.179	0.104	0.177	0.260
$l\beta$	-2.122	-2.467	-2.122	-1.782
$\tau_{l\beta}$	13.806	2.347	9.330	53.041
$l\beta_1$	-2.033	-2.603	-2.034	-1.465
$l\beta_2$	-2.372	-2.798	-2.371	-1.943
$l\beta_3$	-2.401	-2.888	-2.400	-1.922
$l\beta_4$	-2.081	-2.541	-2.084	-1.617
$l\beta_5$	-1.934	-2.498	-1.949	-1.282
$l\beta_6$	-1.978	-2.482	-1.979	-1.468
$l\beta_7$	-1.826	-2.370	-1.831	-1.253
$l\beta_8$	-2.083	-2.582	-2.089	-1.556
$l\beta_9$	-1.988	-2.391	-1.991	-1.568
$l\beta_{10}$	-2.053	-2.503	-2.055	-1.575
$l\beta_{11}$	-2.096	-2.586	-2.097	-1.612
$l\beta_{12}$	-2.611	-3.148	-2.611	-2.083

Table 3: Estimated parameters with credibility sets

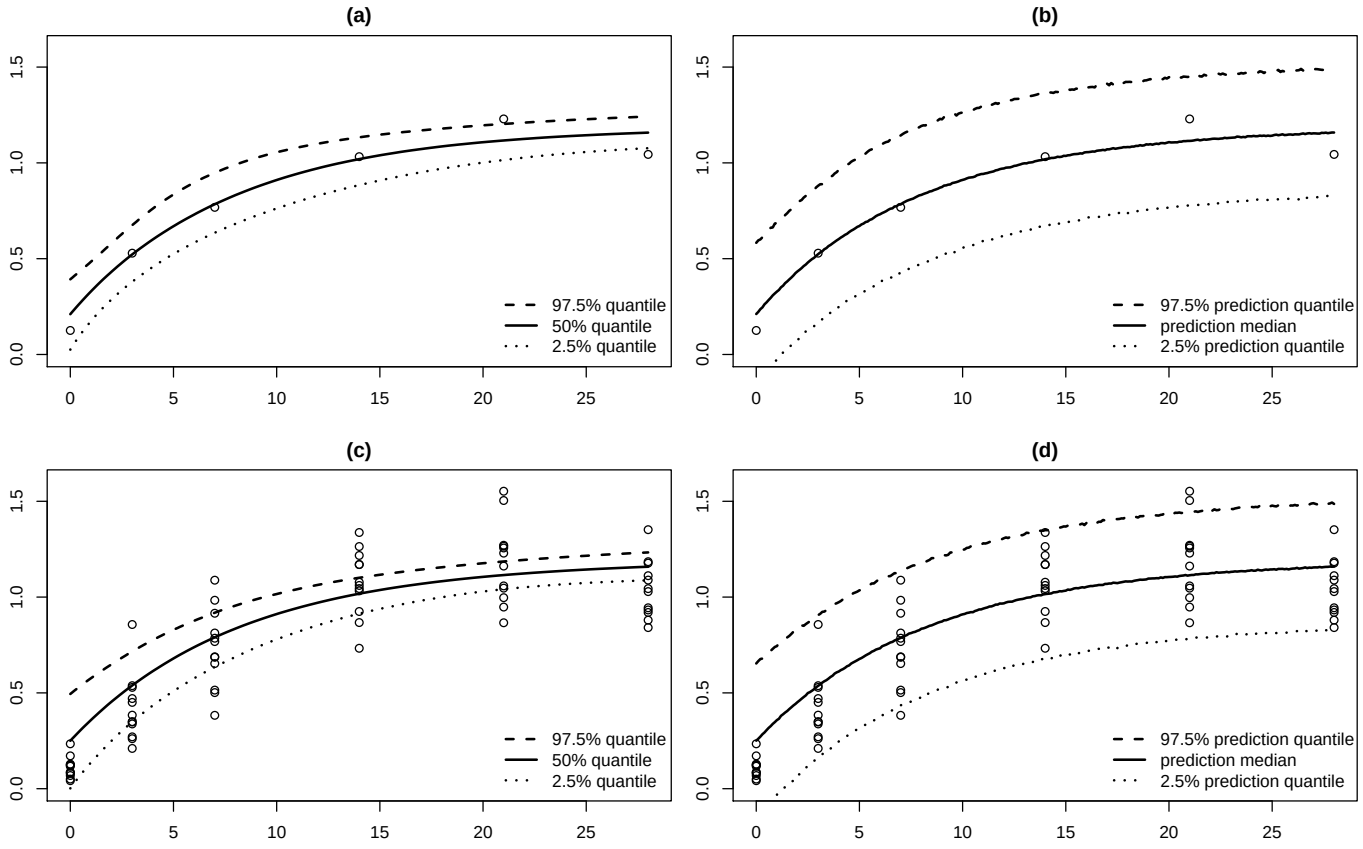


Figure 5: (a): Credibility interval for the perfusion ratio for subject 8. (b): Credibility interval for the posterior predictive distribution of perfusion ratio for subject 8. (c): Credibility interval for the mean population perfusion ratio. (d): Credibility interval for the posterior predictive distribution of perfusion ratio

intervals no more depend on the choice of a specific quantile.
Some extensions are desirable. The Bayesian ODE-penalized B-spline approach assumes that the error distribution is Gaussian with homogeneous variance. The case where the variance of the Gaussian error distribution is non-homogeneous has to be considered. It corresponds to cases where the variance of the error depends on time and/or on the state function that is measured. In the case where the error term depends on the state function, the marginalization of the joint posterior distribution with respect to the spline coefficients will be difficult. Therefore, it will be complicated to get rid of the inconvenient posterior correlation between the generated spline coefficients and the differential equation parameters.
The assumption of Gaussian error distribution is most of the time inappropriate and other distributions (such as truncated Gaussian, gamma or lognormal distribution) should be investigated. The choice of other distributions will be again challenging as the marginalization of the joint posterior distribution with respect to the spline coefficients will be no more feasible.
Before applying parameter estimation procedure, one has also to care about the design. Attention should be given to the choice of experiemental design to avoid poor parameter estimation and thus unreliable results. Optimal design are based on criteria that involve the Fisher information matrix (FIM). Actually, to construct this matrix, one has to approximate the solution of the ODE using a first order Taylor series expansion (in the affine case as well as in the nonlinear ODE case). Using the ODE penalized B-spline approach, the FIM will not require a Taylor expansion, but design optimality criterion will continue to depend on a priori value of the parameters involved in the model.
Finaly, in this paper, only systems of affine differential equations have been considered but a large part of the ODE models are nonlinear. These nonlinear cases are more challenging in the sense that the derivation of the prior distribution (and therefore the posterior distribution) of the spline coefficients will be more complicated than the multivariate Gaussian distribution we got in the affine case.

Appendix 1 - Prior distribution for the spline parameters

In order to understand the limite case in the prior distribution for the spline parameter, we will consider the simpliest dynamic system, namely the exponential decline problem:

$$\left\{ \begin{array}{rcl} \frac{dx}{dt}(t) & = & -\theta x(t) \\ x(0) & = & x_0 \end{array} \right.$$

In this case, the penalty matrix $\boldsymbol{R}(\theta,\gamma)$ has a simple expression:

$$\boldsymbol{R}(\theta,\gamma) = \gamma\left(\boldsymbol{P^{11}} + \theta\left(\boldsymbol{P^{10}} + \boldsymbol{P^{01}}\right) + \theta^2\boldsymbol{P^{00}}\right)$$

where $\boldsymbol{P^{ij}} = \int \boldsymbol{B}^{(i)}(t)\left(\boldsymbol{B}^{(j)}(t)\right)^T dt$, $i,j = 0,1$.
The additional elements in the prior distribution for the spline parameters are:

$$\boldsymbol{\mu_c}^T = (x_0, 0, \ldots, 0)$$

$$\boldsymbol{\Sigma_c}^{-1} = \begin{pmatrix} \tau_{x_0} & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}$$

As explained in Section 3, the prior distribution for the spline parameters is here a multivariate normal distribution with mean $\boldsymbol{\mu_{c,prior}} = \left(\boldsymbol{R}(\boldsymbol{\theta},\boldsymbol{\gamma}) + \boldsymbol{\Sigma_c}^{-1}\right)^{-1}\boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}$ and variance-covariance matrix $\boldsymbol{V_{c,prior}} = \left(\boldsymbol{R}(\boldsymbol{\theta},\boldsymbol{\gamma}) + \boldsymbol{\Sigma_c}^{-1}\right)^{-1}$.
For the variance-covariance, a block matrix decomposition is usefull to study the limited case:

$$\begin{array}{rcl} \boldsymbol{V_{c,prior}} & = & \begin{pmatrix} \boldsymbol{R}_1^1 + \tau_{x_0} & \boldsymbol{R}_1^{2:m} \\ \boldsymbol{R}_{2:m}^1 & \boldsymbol{R}_{2:m}^{2:m} \end{pmatrix}^{-1} \\ & = & \begin{pmatrix} s_D^{-1} & -s_D^{-1}\boldsymbol{R}_1^{2:m}\left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} \\ -\left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1}\boldsymbol{R}_{2:m}^1 s_D^{-1} & \left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} + \left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1}\boldsymbol{R}_{2:m}^1 s_D^{-1}\boldsymbol{R}_1^{2:m}\left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} \end{pmatrix} \end{array}$$

where $s_D = \boldsymbol{R}_1^1 + \tau_{x_0} - \boldsymbol{R}_1^{2:m} \left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} \boldsymbol{R}_{2:m}^1$.

Note that, when τ_{x_0} tends to infinity then V tends to $\begin{pmatrix} 0 & 0 \\ 0 & \left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} \end{pmatrix}$.

The precision τ_{x_0} has therefore influences on each block of the variance covariance matrix. Furthermore, when τ_{x_0} tends to infinity, the limit of the variance-covariance matrix proves that the first component $\boldsymbol{c}_1$ of the spline parameters is fixed (variance is equal to zero and the correlation with the other spline parameters $\boldsymbol{c}_{2:m}$ is null).

Using the same block matrix decomposition, the expression of the mean of the prior distribution of the spline parameters is:

$$\mu = \begin{pmatrix} s_D^{-1} \tau_{x_0} x_0 \\ -\left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} \boldsymbol{R}_{2:m}^1 s_D^{-1} \tau_{x_0} x_0 \end{pmatrix}$$

Note that, when τ_{x_0} tends to infinity then μ tends to $\begin{pmatrix} x_0 \\ -\left(\boldsymbol{R}_{2:m}^{2:m}\right)^{-1} \boldsymbol{R}_{2:m}^1 x_0 \end{pmatrix}$.

The first component of the mean of the prior distribution tends to x_0 (main objective) when τ_{x_0} tends to infinity. It is also important to note that the other component of this prior mean depend on the initial condition x_0 and on the differential equation.

References

L Biegler, J Damiano, and G Blau. Nonlinear parameter estimation: A case study comparison. *AIChE Journal*, 32(1): 29–45, 1986. doi: http://dx.doi.org/10.1002/aic.690320105.

B Cai, R Meyer, and F Perron. Metropolis - Hastings algorithms with adaptive proposals. *Statistics and Computing*, 18(4):421–433, 2008. doi: http://dx.doi.org/10.1007/s11222-008-9051-5.

D Campbell. *Bayesian collocation tempering and generalized profiling for estimation of parameters from differential equation models*. PhD thesis, McGill University, 2007.

J Cao and H Zhao. Estimating dynamic models for gene regulation networks. *Bioinformatics*, 24(14):1619–1624, 2008. doi: http://dx.doi.org/10.1093/bioinformatics/btn246.

Carl de Boor. *A Practical Guide to Splines*. Springer-Verlag New York Inc, 2001.

P Eilers and B Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–121, 1996. doi: http://dx.doi.org/10.1214/ss/1038425655.

A Jullion and P Lambert. Robust specification of the roughness penalty prior distribution in spatially adaptative bayesian P-splines models. *Journal of Computational Statistics & Data Analysis*, 51:2542–2558, 2007. doi: http://dx.doi.org/10.1016/j.csda.2006.09.027.

P Lambert. Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics & Data Analysis*, 51(12):6307–6320, 2007. doi: http://dx.doi.org/10.1016/j.csda.2007.01.018.

S Lang and A Brezger. Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212, 2004. doi: http://dx.doi.org/10.1198/1061860043010.

D Lunn, N Best, A Thomas, J Wakefield, and D Spiegelhalter. Bayesian analysis of population PK/PD models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics*, 29(3):271–307, 2002. doi: http://dx.doi.org/10.1023/A:1020206907668.

A Poyton, M Varziri, K McAuley, P McLellan, and J O Ramsay. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & Chemical Engineering*, 30:698–708, 2006. doi: http://dx.doi.org/10.1016/j.compchemeng.2005.11.008.

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

J O Ramsay, G Hooker, D Campbell, and J Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society, Series B*, 69:741–796, 2007. doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00610.x.

D Spiegelhalter, N Best, B Carlin, and A van der Linde. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64:583–640, 2002.

V van Weel, R Toes, L Seghers, M Deckers, M De Vries, P Eilers, J Sipkens, A Schepers, D Eefting, V van Hinsbergh, J van Bockel, and P Quax. Natural killer cells and CD4+ t-cells modulate collateral artery development. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27:2310–2318, 2007. doi: http://dx.doi.org/10.1161/ATVBAHA.107.151407.

J Varah. A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal of Computational and Graphical Statistics*, 3:28–46, 1982. doi: http://dx.doi.org/10.1137/0903003.

J Ye. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 93:120–131, 1998. URL `http://www.jstor.org/stable/2669609`.