UCL

Université catholique de Louvain

THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE CATHOLIQUE DE LOUVAIN

Institut de Statistique, Biostatistique et Sciences Actuarielles

DOMAINE DE RECHERCHE : Sciences (orientation Statistique)

Présentée par

Jonathan Jaeger

Functional estimation in systems defined by differential equations using Bayesian smoothing methods

Soutenance privée le 28 juin 2012
Soutenance publique le 26 septembre 2012

JURY

M. Paul Eilers	Professeur (Erasmus MC)	Membre du jury
Mme. Bernadette Govaerts	Professeur (UCL)	Membre du jury
M. Philippe Lambert	Professeur (ULg, UCL)	Promoteur
Mme. Catherine Legrand	Professeur (UCL)	Membre du jury
M. Rainer von Sachs	Professeur (UCL)	Président du jury

Acknowledgments

Je voudrais en premier lieu remercier le Professeur Philippe Lambert pour son implication et ses nombreuses suggestions tout au long de la thèse. Merci pour ton fort soutien et ta constante bonne humeur durant ces quatre années.
I would also thank Professor Paul Eilers for his constructive comments and being part of the jury. Je souhaite également remercier le Professeur Bernadette Govaerts pour la confiance qu'elle m'a accordée, en particulier pour les missions SMCS. Merci au Professeur Catherine Legrand pour son implication et ses commentaires pertinents et au Professeur Rainer von Sachs pour sa disponibilité.
Mes remerciements vont bien entendu à tous les membres de l'Institut, personnels administratifs, académiques, scientifiques et aux membres du SMCS qui ont contribué à une très bonne atmosphère de travail.
Je tiens à remercier ma famille et mes amis de m'avoir apporté leur soutien. Merci pour les bons moments de détente qui ont agrémenté ces quatre années.
Merci à Edouard Ledent de m'avoir ouvert les yeux. Sans tes conseils, la statistique ne serait déjà plus qu'un vague souvenir.
Enfin, je remercie Sebastien de m'encourager, de me soutenir et d'être à mes côtés.

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

Contents

Context and goals of the thesis **1**

I General introduction **5**

I.1 Problem definition . . . 5

I.2 Datasets . . . 6

I.2.1 Perfusion ratio modeling after a femoral artery occlusion . . . 6

I.2.2 Pharmacokinetic datasets . . . 8

I.2.2.1 Pharmacokinetic studies and compartment models . . . 8

I.2.2.2 Indomethacin dataset . . . 8

I.2.2.3 Theophylline dataset . . . 10

I.3 Literature review . . . 12

I.3.1 Traditional parameter estimation in ordinary differential equation models 12

I.3.1.1 Single shooting approach . . . 12

I.3.1.2 Multiple shooting approach . . . 14

I.3.2 Kalman filter . . . 15

I.3.3 Collocation based approaches . . . 16

I.3.3.1 Orthogonal collocation . . . 16

I.3.3.2 Local kernel smoothing-based approach . . . 16

I.3.3.3 Spline least squares method . . . 17

I.3.3.4 Principal Differential Analysis . . . 17

I.3.3.5 ODE-penalized smoothing approaches . . . 18

I.3.3.5.1 Iterative Principal Differential Analysis . . . 18

I.3.3.5.2 Generalized profiling estimation procedure 18
I.3.3.5.3 Advantages and weaknesses of the frequentist approaches 19
I.3.3.5.4 Bayesian generalized profiling estimation procedure . . 19
I.4 Discussion 20

II Bayesian P-spline estimation in hierarchical models specified by systems of affine differential equations 23
II.1 Introduction 23
II.2 Bayesian ODE-penalized B-spline approach 24
II.2.1 B-spline definition and generalized profiling estimation procedure 24
II.2.2 Bayesian ODE-penalized B-spline approach 28
II.3 Posterior distribution and parameter estimation 30
II.3.1 Joint posterior distribution 30
II.3.2 Exploring the joint posterior distribution using MCMC 32
II.3.2.1 Conditional posterior distributions 32
II.3.2.2 Posterior correlation and marginal joint posterior distribution . 33
II.3.2.3 Metropolis-Hastings algorithm with adaptive proposals 34
II.3.2.4 Re-parametrization using rotation and translation 34
II.3.2.5 Implementation 36
II.4 Hierarchical Bayesian ODE-penalized B-spline approach for affine dynamic systems 36
II.4.1 Individual system of differential equations 37
II.4.2 B-spline basis functions expansion and ODE-penalty 37
II.4.3 Bayesian hierarchical model 38
II.5 Simulation study 41
II.5.1 Rate of the gamma prior distribution for the ODE-adhesion parameters . 41
II.5.1.1 Data 41
II.5.1.2 Model 41
II.5.1.3 Simulation results 42
II.5.2 Performance of the Bayesian ODE-penalized B-spline approach compared to traditional methods 45
II.5.2.1 Preliminary remarks 45
II.5.2.2 1-compartment intravenous bolus model 46
II.5.2.2.1 Data 46
II.5.2.2.2 Models 46
II.5.2.2.3 Simulation results 47

II.5.2.3 Oral 2-compartment model . . . 52
II.5.2.3.1 Data . . . 52
II.5.2.3.2 Models . . . 52
II.5.2.3.3 Simulation results . . . 53
II.6 Application . . . 59
II.6.1 Data and differential equation model . . . 59
II.6.2 Model . . . 59
II.6.3 Results . . . 60
II.7 Discussion . . . 66
II.A Penalty matrix in a general system of affine differential equations . . . 67
II.B Prior distribution for the spline parameters . . . 69
II.C Impact of the normalizing constant for the prior distribution of the spline coefficients . . . 70
II.D Markov chains Monte Carlo methods . . . 73
II.D.1 Metropolis algorithm . . . 73
II.D.2 Metropolis-Hastings algorithm . . . 74
II.D.3 Gibbs algorithm . . . 74

III Bayesian penalized smoothing approaches in models specified using affine differential equations with unknown error distributions 75
III.1 Introduction . . . 75
III.2 Bayesian ODE-penalized B-spline approach and penalized mixture of Gaussian distributions . . . 77
III.2.1 Model . . . 77
III.2.1.1 Assumption on the error distribution . . . 77
III.2.1.2 Complete likelihood function . . . 78
III.2.1.3 Prior for the transformed mixture weights . . . 79
III.2.1.4 Identification penalty . . . 80
III.2.1.5 Other prior distributions . . . 80
III.2.2 Posterior distribution and parameter estimation . . . 81
III.2.2.1 Joint posterior distribution . . . 81
III.2.2.2 Conditional posterior distributions . . . 82
III.2.2.3 MCMC algorithm . . . 83
III.2.2.3.1 MCMC algorithm . . . 83
III.2.2.3.2 Starting value . . . 84

III.2.2.3.3 Re-parameterization using rotation and translation . . 84
III.3 Bayesian ODE-penalized B-spline approach and penalized mixture of Student distributions . . . 84
III.3.1 Model . . . 85
III.3.1.1 Assumption on the error distribution . . . 85
III.3.1.2 Data augmentation . . . 85
III.3.1.3 Bayesian model . . . 86
III.3.2 Posterior distribution and parameter estimation . . . 86
III.3.2.1 Posterior distribution . . . 87
III.3.2.2 MCMC algorithm . . . 88
III.4 Simulation . . . 89
III.4.1 Robustness of the mixture of Student distributions . . . 89
III.4.1.1 Data and model specification . . . 89
III.4.1.2 Results . . . 90
III.4.2 Simulation study . . . 92
III.4.2.1 Scenario . . . 92
III.4.2.2 Results . . . 92
III.5 Application . . . 99
III.5.1 Data, objective and model specification . . . 99
III.5.2 Results . . . 99
III.6 Discussion . . . 105
III.A Bayesian ODE-penalized B-splines with Student error distribution . . . 106
III.A.1 The Bayesian model . . . 106
III.A.2 Simulation study . . . 108

IV On the use of adhesion parameters to validate models specified using systems of affine differential equations **111**
IV.1 Introduction . . . 111
IV.2 Bayesian ODE-penalized B-spline approach and ODE-model selection . . . 113
IV.2.1 The Bayesian ODE-penalized B-spline approach . . . 113
IV.2.2 Study of the posterior distribution for the ODE-adhesion parameter . . . 115
IV.2.2.1 Asymptotics of the marginalized posterior distribution for ODE-adhesion parameter . . . 115
IV.2.2.2 Divergence measure and ODE-model selection . . . 116
IV.3 Simulation . . . 118

IV.3.1 Sensitivity to the sample size . . . 119
IV.3.2 Sensitivity to a wrong specification of the ODE parameter or of the precision of measurement . . . 120
IV.3.3 Sensitivity to a misspecification of the error distribution . . . 124
IV.3.4 Sensitivity to a ODE-model misspecification . . . 125
IV.4 Application . . . 126
IV.4.1 Data and objective . . . 126
IV.4.2 Bayesian model . . . 127
IV.4.3 Results . . . 128
IV.5 Discussion . . . 136

Concluding comments **137**

List of Symbols **143**

List of Figures **145**

List of Tables **151**

References **155**

Context and goals of the thesis

A large part of the dynamic systems in physics, in chemical engineering and in life/social sciences is modeled by systems of ordinary differential equations (ODEs). In pharmacokinetics, concentration time curves can be modeled using kinetic models defined by systems of differential equations (Bonate, 2011). In biology, individual or population growth can be modeled by differential equations (Verhulst equation in Verhulst (1838), Lotka-Voltera equations). In chemical engineering, ordinary differential equations are used for determining reaction rates or heat transfer operations.

In this thesis, we assume that changes in a state functions $\boldsymbol{x}(t) \in \mathbb{R}^d$ are governed by a set of differential equations:

$$\frac{d\boldsymbol{x}}{dt}(t) = f(\boldsymbol{x}, t, \boldsymbol{\theta}), t \in [0; T]$$

where f is a known affine function of $\boldsymbol{x}$ and $\boldsymbol{\theta} \in \mathbb{R}^q$ a vector of unknown parameters. It is assumed that only a subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state functions are measured at time point t_{jk}, $j \in \mathcal{J}$, $k = 1, \ldots, n_j$ with additive measurement error ϵ_{jk}. We denote by $y_{jk} = \boldsymbol{x}_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$ the corresponding measurement ($j \in \mathcal{J}$). We further impose that $\mathbb{E}(\epsilon_j) = 0$ and $\mathbb{V}(\epsilon_j) = 1$ such that $x_j(t)$ and τ_j are respectively the conditional mean and the inverse-variance of y_j. Given the set of observations $\{(t_{jk}, y_{jk}), j \in \mathcal{J}, k = 1, \ldots, n_j\}$, the objective is to jointly estimate the ODE parameters $\boldsymbol{\theta}$, the state functions $\boldsymbol{x}(t)$, the precision of measurement $\boldsymbol{\tau} = \text{vec}(\tau_j; j \in \mathcal{J})$ and possibly also the error distributions.

The most common methods for estimating parameters from ODE models rely on frequentist or Bayesian nonlinear least squares (Lunn et al., 2002; Seber and Wild, 1989). Both of

these approaches depend on a likelihood involving a numerical solution of the ODE model. These traditional NLS methods can be computationally intensive and have stability and error-propagation issues.

Alternatively, approaches based on the approximation of the unknown state functions involved in the ODE model using a linear combination of B-splines with flexibility counter-balanced by a penalty term related to the system of differential equations have been recently proposed (Poyton et al., 2006; Ramsay and Silverman, 2005; Ramsay et al., 2007). These frequentist ODE-penalized smoothing approaches can be viewed as a trade off between fitting the data and solving the differential equation. They appear therefore robust to model misspecification and to disturbance of the dynamical system. A large part of the existing limitations of the NLS approaches are solved by these ODE-penalized smoothing approaches: they converge quickly, give less biased and more precise estimates and allow to estimate unobserved state function. But, two major drawbacks can be highlighted in these frequentist approaches. The first point is the delicate choice of the ODE-adhesion parameters that control the relative emphasis between goodness-of-fit and solving the differential equation. The second point is the approximation needed for the uncertainty measures about the parameter estimates.

The considerable advances made by Ramsay et al. (2007) are considered as the starting point of the thesis and the overall objective is to propose a fully Bayesian framework of this ODE-penalized smoothing approach when the system of differential equations is affine. Considering the ODE-penalized B-spline approach in a Bayesian framework solves the weaknesses of the frequentist approach: the selection of the ODE-adhesion parameters is now automatic and uncertainty measures about parameter estimates (or function of them) can simply be quantified using MCMC. Furthermore, prior information can be injected in the model specification and generalization from single to multi-subject study is naturally feasible.

The thesis is divided in four chapters:

- Chapter I details the background of the research and gives a review of the literature. The datasets used in the thesis are also presented there.
- Chapter II presents a fully Bayesian approach to jointly estimate the ODE parameters, the state functions and the precision of measurement when the system of differential equations is affine and the error distributions are homogeneous Gaussian. It is shown that the prior distribution elicitation for the spline coefficients has a major influence

in the model specification. Extension from single-subject to multi-subject study is also presented. This chapter is an extended version of the discussion paper DP2011/01 (Jaeger and Lambert, 2011).

- Chapter III is an extension of Chapter II where the error distributions are modeled in a flexible way by using penalized mixtures of Gaussian distributions or penalized mixtures of Student distributions. This approach has two major advantages: it avoids restrictive parametric assumptions on the error distribution and conditionally on some latent variables (in a Metropolis-within-Gibbs sampler), one is back to the Gaussian setting assumed in Chapter II. Working with error distributions modeled by a penalized mixtures of Student distributions allows to be at the same time flexible in the specification of the error distribution and also robust to outliers. This chapter is an extended version of the discussion paper DP2012/17 (Jaeger and Lambert, 2012a).

- Chapter IV proposes a methodology to validate an ODE model or to select competing ODE models. The suitability of a proposed system of differential equations to model the dynamics of a state function is assessed and quantified by the comparison of the posterior distribution of the ODE-adhesion parameter to its prior. This comparison can be made either by a visual inspection or using a divergence measures. To our best knowledge, it is the first time that penalty parameters are used in model validation/selection. This chapter is an extended version of the discussion paper DP2012/18 (Jaeger and Lambert, 2012b).

The thesis finishes with discussion about the current and further research directions.

Chapter I

General introduction

Abstract

In this chapter, the problem of parameter estimation in model defined by systems of affine differential equations is formulated. This parameter estimation problem is motivated by the presentation of the three datasets used in this thesis. Disadvantages of the traditional approaches used to solve these dynamic systems are pointed out. Collocation approaches, and especially the generalized profiling estimation procedure, are presented.

Keywords: Collocation approach; Dataset; Dynamic systems; Kalman filter; Multiple shooting approach; ODE parameter estimation; Single shooting approach.

I.1 Problem definition

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:

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{f}(\boldsymbol{x}, t, \boldsymbol{\theta}), t \in [0; T], \tag{I.1}$$

where $\boldsymbol{f}$ is a known affine function with respect to $\boldsymbol{x}$ and $\boldsymbol{\theta} \in \mathbb{R}^q$ an unknown vector of parameters. The number d of state functions involved in the ODE model is generally lower than 5.

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}. With this assumption, possibly not all the state functions are observed and the number of observation points could also differ between the observed state functions. We denote by

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

the corresponding measurement, where $x_j(t)$ and τ_j are respectively the conditional mean and the precision (i.e. inverse-variance) of y_j given the time.
Using the set of observations $\{(t_{jk}, y_{jk}); j \in \mathcal{J}, k = 1, \ldots, n_j\}$, the objective is to estimate the ODE parameter $\boldsymbol{\theta}$, the state function $\boldsymbol{x}(t)$, the precisions of measurement $\boldsymbol{\tau} = \text{vec}(\tau_j; j \in \mathcal{J})$ and possibly also the standardized error distributions if these are not standardized Gaussians.

I.2 Datasets

In this section, we motivate the ODE parameter estimation by the presentation of the datasets used for the applications in Chapters II, III and IV. One dataset is based on a recovery study after a femoral artery occlusion and two datasets are based on pharmacokinetic studies.

I.2.1 Perfusion ratio modeling after a femoral artery occlusion

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., 113 mice were enrolled. We focus on the specific C57BL/6 mice's group to illustrate the hierarchical Bayesian ODE-penalized B-spline approach presented in Chapter II. In this specific group, perfusion ratio was measured 6 times over a 28 day period in 12 mice after a femoral artery occlusion (see Figure I.1). Originally, this dataset was analyzed using a nonlinear model that assumes an exponential approach to an asymptote. The differential equation that describes the evolution of the perfusion ratio $x(t)$ is:

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

It is assumed that α and β are positive quantities. The parameter α corresponds to the asymptotic value of the ischemic to the non-ischemic perfusion ratio and $\alpha\beta$ corresponds to the initial increase of the perfusion ratio. Under this differential equation model, the perfusion

ratio $x(t)$ at time t can be shown to be:

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

Please note that working with this analytical solution for the state function forces the estimated state function to be equal to zero immediately after surgery.

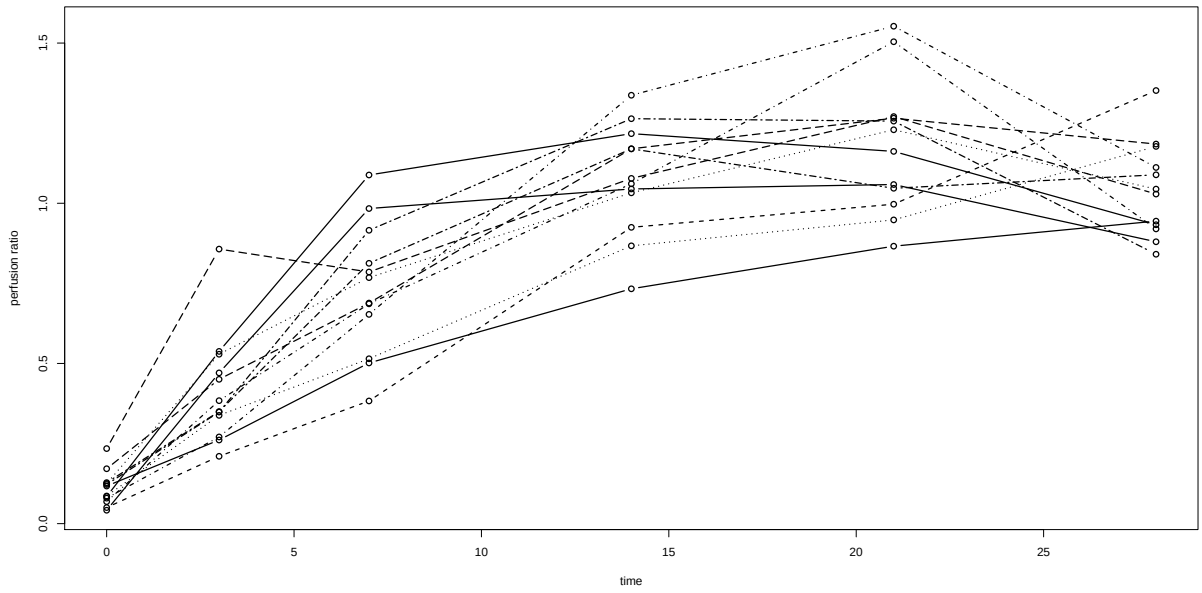


Figure I.1 – Ischemic/non-ischemic perfution ratio.

This dataset was found interesting for several reasons. First, its hierarchical structure is a challenge when it comes to inference in models specified using differential equations. It will be used to illustrate the hierarchical version of the Bayesian ODE-penalized B-spline model presented in Chapter II. A second difficulty comes from the initial conditions as the perfusion ratios directly after ischemia are not exactly zero, but instead randomly vary across mice. The possibility to set a prior distribution on the initial conditions will be found to be a major asset of the Bayesian approach. Finally, it will illustrate the stability of the proposed approach. Some convergence issues were indeed encountered in the original frequentist approach.
van Weel et al. (2007) are gratefully acknowledged for having provided this dataset.

I.2.2 Pharmacokinetic datasets

I.2.2.1 Pharmacokinetic studies and compartment models

The exposure of subjects to a new drug under investigation are needed in order to determine the pharmacokinetic (PK) profiles, i.e. the evolution of the drug concentration in plasma over time. In order to achieve this goal, blood samples are taken on several subjects in order to estimate population PK parameters and also the variability around these population parameters. 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 (Bonate, 2011).

I.2.2.2 Indomethacin dataset

Kwan et al. (1976) studied the pharmacokinetics of a drug named Indomethicin. It is a non-steroidal anti-inflammatory drug commonly used to reduce fever, pain and stiffness. In this study, six subjects received a single dose of Indomethacin (25 mg) by bolus intravenous injection. Plasma concentrations of the drug (in μg/mL) were measured 11 times from 15 minutes to 8 hours after injection. Figure I.2 shows the PK profiles. All the concentration curves have a similar shape but seem to differ across individuals. Two PK compartment models can potentially describe the dynamics of the observed concentrations.

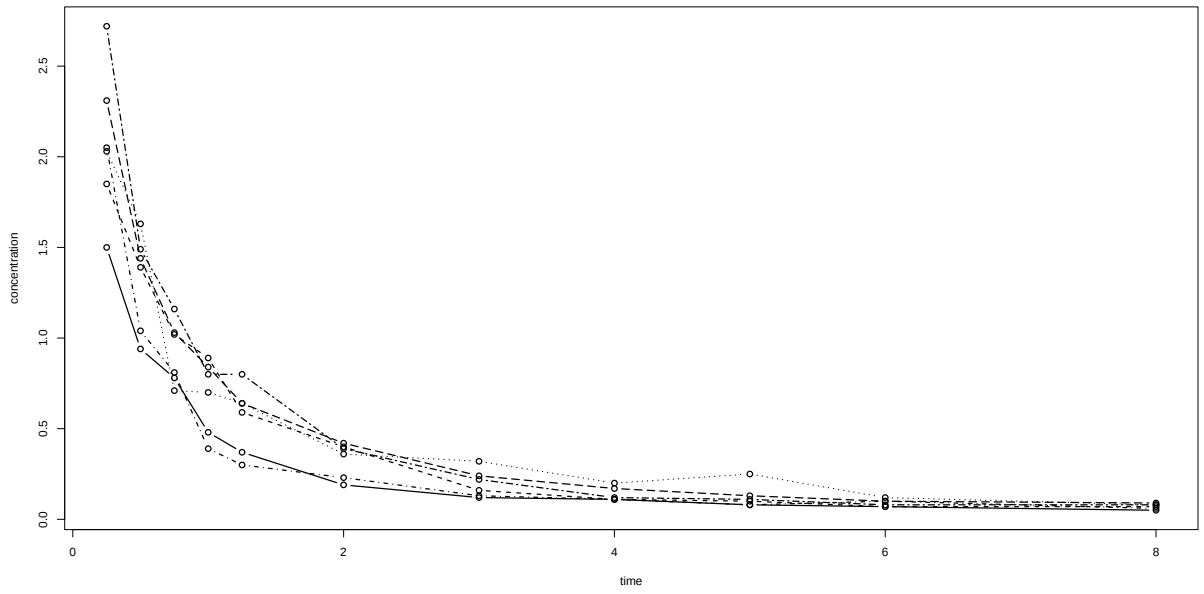


Figure I.2 – Pharmacokinetic profiles for the Indomethacin dataset.

The 1-compartment intravenous bolus model describes the concentration $x(t)$ of drug in blood after an instantaneous injection of a dose D of a drug into the blood (see Figure I.3). The entire body is here assimilated to one central compartment of volume $V > 0$ and the drug is eliminated at a rate $k_e > 0$.

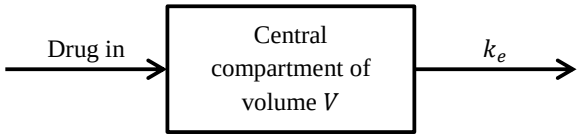


Figure I.3 – 1-compartment intravenous bolus model.

This 1-compartment intravenous bolus model can be mathematically translated in this differential equation model:

$$\begin{cases} \frac{dx}{dt}(t) &= -k_e x(t) \\ x(0) &= \frac{D}{V}. \end{cases}$$

The 2-compartment intravenous bolus model arises when a part of the drug in the central compartment passes through a peripheral compartment. Let $x_1(t)$ denote the concentration of drug in the central compartment of volume V and $x_2(t)$ the amount of drug in the peripheral compartment at time t. A part of the drug is directly eliminated from the central compartment with rate k_e, another part goes to the peripheral compartment at rate k_{cp} and comes back in the central compartment at rate k_{pc} (see Figure I.4). In this setting, only the concentration of drug in the central compartment is observed and the amount of drug in the peripheral compartment is unobserved.

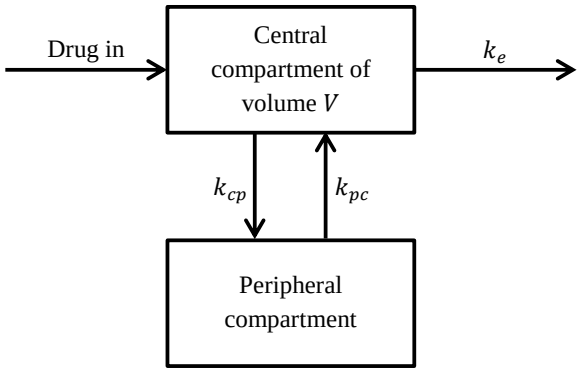


Figure I.4 – 2-compartment intravenous bolus model.

The corresponding system of differential equations with the initial conditions is:

$$\begin{cases} \dfrac{dx_1}{dt}(t) &= -(k_e + k_{cp})\, x_1(t) + \dfrac{k_{pc}}{V} x_2(t) \\ \dfrac{dx_2}{dt}(t) &= k_{cp} V x_1(t) - k_{pc} x_2(t) \\ x_1(0) &= \dfrac{D}{V} \\ x_2(0) &= 0. \end{cases}$$

For this model, it is supposed that k_e, k_{cp}, k_{pc} and V are positive with $k_{cp} > k_{pc}$. An analytic solution of this dynamic system exists and is available in Bonate (2011).
This dataset will be used in Chapter IV to illustrate the usefulness of the ODE-adhesion parameters in the selection of an adequate system of differential equations for the modelling of the dynamics of the concentration of Indomethacin in blood.

I.2.2.3 Theophylline dataset

Theophylline is a well-known anti-asthmatic agent, administrated orally. Twelve subjects are given oral dose at time 0 with blood samples taken at 11 time points over the following 25 hours (Davidian and Giltinan, 1995). The blood samples are assayed for the theophylline concentration. Individual profiles are presented on Figure I.5.

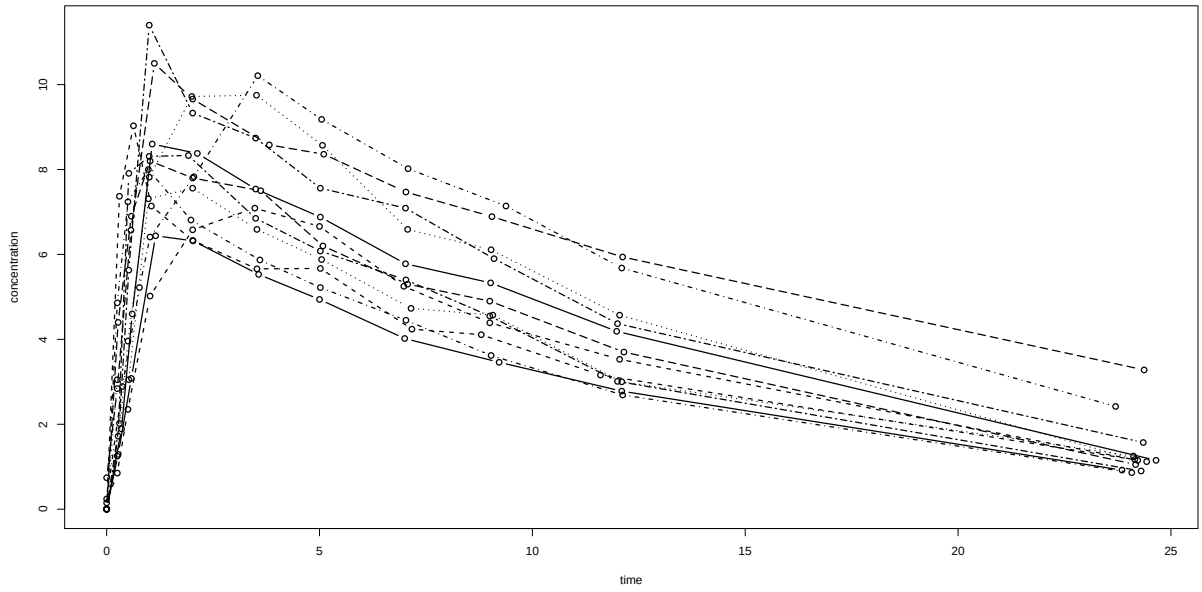


Figure I.5 – Pharmacokinetic profiles for the Theophylline dataset.

A common model for the kinetics of Theophylline after oral administration is the oral 2-

compartment model with first order absorption and elimination. This compartment model arises when drug at dose D is administrated orally in the 1-compartment intravenous bolus model. Let $x_1(t)$ denote the amount 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 $x_2(t)$ in this central compartment of volume V and is eliminated at rate k_e. Figure I.6 shows a diagram describing the evolution of drug in the body.

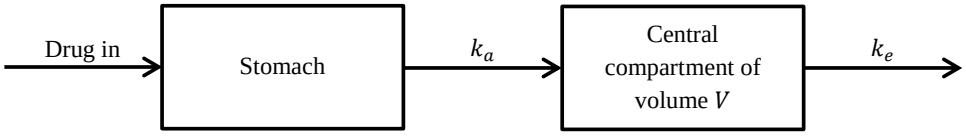


Figure I.6 – Oral 2-compartment model.

The system of differential equations describing the evolution of the amount of drug in the stomach and the concentration of drug in the blood is:

$$\begin{cases} \dfrac{dx_1}{dt}(t) &= -k_a x_1(t) \\ \dfrac{dx_2}{dt}(t) &= \dfrac{k_a}{V} x_1(t) - k_e x_2(t) \\ x_1(0) &= D \\ x_2(0) &= 0. \end{cases}$$

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:

$$\begin{cases} x_1(t) &= D\exp(-k_a t) \\ x_2(t) &= \dfrac{D}{V}\dfrac{k_a}{k_a - k_e}\left(\exp(-k_e t) - \exp(-k_a t)\right). \end{cases}$$

This dataset will be used in Chapter III to illustrate the combination of the Bayesian ODE-penalized B-spline approach for the estimation of the state functions and the penalized mixture of Gaussian distributions for the estimation of the standardized error distribution.

I.3 Literature review

In this section, a review of the literature for parameter estimation in systems defined by a set of differential equations is presented. First, traditional parameter estimation procedures are presented. They depend on a likelihood involving a numerical integration of the solution. Then, ODE parameter estimation combined with Kalman filter is presented. Finally, collocation approaches are presented. There, the state functions are approximated using a basis function expansion. It is mostly combined with a penalty term that constrains the variation of the approximated state function.

I.3.1 Traditional parameter estimation in ordinary differential equation models

In this section, we will review two widely used strategies in the estimation of parameters in models defined using ODE: the single shooting approach and the multiple shooting approach.

I.3.1.1 Single shooting approach

The most commonly used estimation procedures rely on nonlinear least squares (NLS) procedures (Biegler et al., 1986). It consists to minimize the sum of the squared deviations of the model predictions from the measured state function:

$$\sum_{k=1}^{n} \left(y_k - x\left(t_k\right)\right)^2, \tag{I.2}$$

with dynamic of the state function given by Eq. (I.1). This approach depends on an analytical solution when the ODE model is affine or relies on a numerical procedure when no explicit solution is available (as often in systems of nonlinear differential equations). As an initial condition is needed for the numerical integration, this approach is also called single shooting method or initial value problem.

The estimation procedure is iterative: first the ODE system is solved numerically given the current estimate of the ODE parameter $\boldsymbol{\theta}$ and the initial value $\boldsymbol{x}\left(0\right)$. Then the numerical solution is compared with the measured responses, suggesting an update of the estimate for $\boldsymbol{\theta}$. Iterations between computation of the numerical approximation of the solution and parameter updating continue until some convergence criteria for the parameter estimate is met. If the initial condition of the state function is not known, it should be included in the set of parameters to be estimated.

The numerical approximation to the solution consists to iteratively reconstruct the state function on a (equidistant) grid of time by extrapolating the state function to the next time-step using some already available information on the state function at previous time-steps. Several approaches have been developed, but the fourth order Runge-Kutta algorithm (Griffiths and Higham, 2010) is usually advocated. Given the state function $x(t)$ at time t defined by Eq. (I.1), the state function is approximated at time $t+h$ by:

$$x(t+h) = x(t) + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4),$$

where

$$\begin{aligned} K_1 &= f(x, t, \boldsymbol{\theta}) \\ K_2 &= f\left(x + \frac{h}{2}K_1, t + \frac{h}{2}, \boldsymbol{\theta}\right) \\ K_3 &= f\left(x + \frac{h}{2}K_2, t + \frac{h}{2}, \boldsymbol{\theta}\right) \\ K_4 &= f(x + hK_3, t + h, \boldsymbol{\theta}). \end{aligned}$$

Each coefficient $\{K_i; i = 1, \ldots, 4\}$ has a simple graphical interpretation: for example, K_1 is the derivative of the state function at time t and K_4 is the derivative of the state function at time $t+h$ using $x(t) + hK_3$. Note that the fourth order Runge-Kutta algorithm provides significant improvement over lower order Runge-Kutta algorithms.

This NLS procedure has many disadvantages. First of all, numerical approximation of the state function has to be computed for each update of the parameter estimate which can be time consuming. Furthermore, the accuracy of the parameter estimate is also highly dependent on the numerical integration. Finally, an important problem arises in this NLS approach when misspecification of the initial condition of the state function is propagated and amplified by the numerical integration. This problem is illustrated on the oral 2-compartment model presented in Section I.2.2.3. In this kind of pharmacokinetic studies, increasing oral dose can be periodically administrated. As the rate of absorption is generally large, the amount of the drug in the stomach can be considered to be equal to the new dose at the new administration time. At the opposite, especially when the constant of elimination is small, the concentration of the drug in blood at the end of a phase can be under the limit of quantification but should not be considered to be equal to zero. This is illustrated on Figure I.7. In this setting, five increasing oral administrations are considered (+50% every two hours). The true solution and the numerical solution differ more and more at the beginning of each new phase with

a decreasing error propagation of this difference within two consecutive administrations. It highlights the necessity to be able to express some uncertainty with respect to initial conditions.

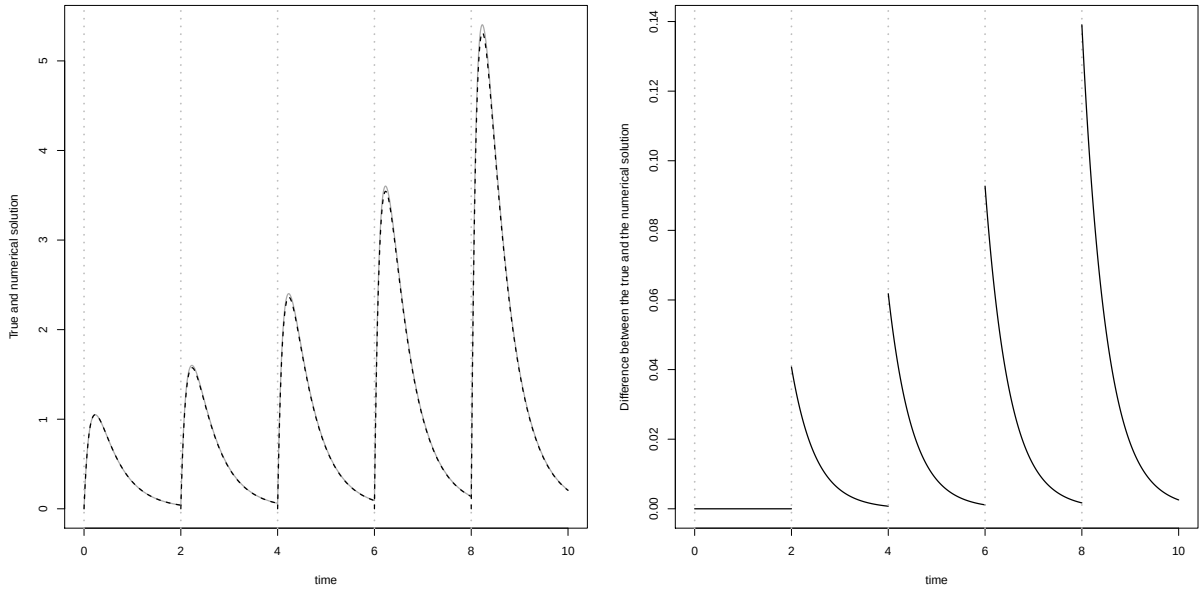


Figure I.7 – Error propagation between the true concentration (solid grey curve) and numerical approximation (using four order Runge-Kutta algorithm, dashed black curve) in a oral 2-compartment model after periodical administration (dotted light gray vertical lines).

Some stability issues can also be encountered for more complex ODE models, especially with input function and/or when function $\boldsymbol{f}$ is nonlinear with respect to $\boldsymbol{x}$.
Bayesian approaches of these NLS procedures have been proposed in Lunn et al. (2002). It allows to include some prior information about the parameters involved in the model and can therefore possibly stabilize the estimation procedure. Moreover, inference is directly based on the desired model rather than on approximations as usually practiced in frequentist approaches (approximation of the matrix of variance-covariance of the parameter estimate, confidence intervals, ...). But the posterior distribution of ODE parameter $\boldsymbol{\theta}$ has no explicit form due to the numerical approximation of the solution to the ODE. The most troublesome disadvantage of this Bayesian approach is that the solution of the system of differential equations must be reevaluated numerically at each iteration of the MCMC algorithm.

I.3.1.2 Multiple shooting approach

The influence of the initial condition of the state function on the numerical integration is less problematic in the multiple shooting approach (Baake et al., 1992). Its basic idea is to consider

the system of differential equations not as an initial value problem but as a multipoint boundary-value problem and to treat this as a nonlinear constraint in the optimization process. In this approach, the interval of study $[0;T]$ is splitted into m multiple subintervals using shooting nodes $0 = \zeta_1 < \zeta_2 < \ldots < \zeta_m < \zeta_{m+1} = T$ and the system of differential equations is considered on each subinterval $[\zeta_l; \zeta_{l+1}]$, $l = 1, \ldots, m$:

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt}(t) &= f(\boldsymbol{x}, t, \boldsymbol{\theta}) \\ \boldsymbol{x}(\zeta_l) &= \boldsymbol{x}_{0,l}, \end{cases}$$

where the local initial values $\boldsymbol{x}_{0,l} = \boldsymbol{x}(\zeta_l)$, $l = 1, \ldots, m$, are introduced as additional parameters. The ODE parameter $\boldsymbol{\theta}$ and the additional local initial values are estimated by minimizing the sum of all sum of squared deviations of the model predictions from the measured state function on each sub-intervals (see Eq. (I.2)). Whereas the ODE parameter $\boldsymbol{\theta}$ is optimized over the entire interval $[0;T]$, the local initial values are optimized separately in each subinterval. Note that standard numerical approximation or collocation-based approaches (presented in Section I.3.3) can be used to approximate the ODE in each subinterval. This strategy leads to a discontinuous approximation of the the state function that usually turns to be closer to the true solution than when using a single shooting approach. However, that extra accuracy has an important computational price as several extra initial conditions are added to the set of parameters to be estimated.

I.3.2 Kalman filter

The Kalman filter is a iterative predictive filter, originally introduced for linear dynamic systems, that is based on a recursive algorithm. It estimates the trajectory of the state function using a two steps procedure. For each new observation, the state function is first predicted given the dynamic model and the last available estimate of the state function. This predicted value is called a priori state estimate as it does not include observation information from the current timestep. Then, the a priori state estimate is combined with the observation at current time in order to refined it. This updated state estimate is called a posteriori state estimate.
Maximum likelihood estimation is used for parameter estimation (Durbin and Koopman, 2001). Alternatively, the EM algorithm can be adapted for parameter estimation, but the convergence rate of the EM algorithm compared with the Newton-based procedure procedure is slow.
Kalman filter has been generalized to the nonlinear case (extended Kalman filter) using linearization but that approach needs time consuming calculations.

Note that Kalman filters allow processes noise in the dynamic model and also time varying model parameters.

I.3.3 Collocation based approaches

Alternatively, the estimation of the ODE parameter $\boldsymbol{\theta}$ and of the state function $\boldsymbol{x}(t)$ can be managed using collocation-based approaches. The basic idea consists to approximate the state function using some basis decomposition (essentially orthogonal Legendre polynomials or B-splines).

I.3.3.1 Orthogonal collocation

Orthogonal collocation applies a polynomial approximation to the system of differential equations and requires satisfaction of these equations at discrete collocation points, the zeros of orthogonal polynomials (e.g. Legendre polynomials in Biegler (1984)). The state function is approximated by:

$$\widetilde{x}(t) = x(0) + \sum_{l=1}^{m} a_l P_{l-1}(t),$$

where $\{a_l; l = 1, \ldots, m\}$ are unknown coefficients and $P_{l-1}(t)$ are $(l-1)$-order Legendre polynomials. The coefficients a_l, $l = 1, \ldots, m$ are determined by substituting the polynomial approximation in the initial value problem given in Eq. (I.1) evaluated at the zeros of the Legendre polynomials. Therefore, the resolution of the system of differential equations and the parameter estimation problem reduces to a constrained nonlinear optimization problem. Parameter and state functions are simultaneously estimated by minimizing the parameter-estimation objective function, subject to the discretized model equations. This approach is limited by the accuracy of the collocation procedure, i.e. depends on the accuracy of the polynomial approximation.

I.3.3.2 Local kernel smoothing-based approach

Liang and Wu (2008) proposed a parameter estimation method based on a local kernel smoothing approach. There, the state function is estimated by minimizing a weighted least squares criterion that involves a symmetric kernel function and a bandwidth. Then, the ODE parameter $\boldsymbol{\theta}$ is estimated through the minimization of the squared deviation from the ODE model (in one dimension):

$$\sum_{k=1}^{n} \left(\frac{d\widetilde{x}}{dt}(t_k) - f(\widetilde{x}, t_k, \boldsymbol{\theta}) \right)^2,$$

where $\widetilde{x}(t_k)$ is the smooth estimate of the state function evaluated at time point t_k.
This approach suffers from several weaknesses. First, the state function is estimated without taking the ODE model into account and depends on a bandwidth selection. Furthermore, the ODE parameter is estimated based on a discrete criterion and not on a criterion based on the ODE model discrepancy on the whole interval of observation. Finally, all the state functions have to be observed in order to estimate the ODE parameter.

I.3.3.3 Spline least squares method

Varah (1982) introduces a spline based approach 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 the ODE parameter $\boldsymbol{\theta}$ is estimated through the minimization of the sum of all the squared deviation from the ODE model, evaluated at selected time points $\left\{\hat{t}_l; l = 1, \ldots, m\right\}$:

$$\sum_{j=1}^{d}\sum_{l=1}^{m}\left(\frac{d\widetilde{x}_j}{dt}\left(\hat{t}_l\right) - f_j\left(\widetilde{\boldsymbol{x}}, \hat{t}_l, \boldsymbol{\theta}\right)\right)^2,$$

where $\widetilde{x}_j(t)$ is the cubic B-spline basis function expansion used to approximate the jth state function.
Varah's approach is interesting when the ODE model is linear in the ODE parameters as the second step consists to a linear least squares problem. But this approach suffers from several weaknesses. First of all, the number and the position of the knots are adjusted adaptively using visual inspection, until a good spline fit is obtained. The number and the position of the set of points $\left\{\hat{t}_l; l = 1, \ldots, m\right\}$ is also chosen arbitrary without any guideline. Finally, this spline least squares approach is highly dependent on the spline fit: each state function has to be observed in order to be approximated and a poor spline fit can lead to poorly estimated ODE parameters.

I.3.3.4 Principal Differential Analysis

Ramsay (1996) proposes to approximate the state functions using a B-spline basis function expansion using L-spline. In this two step procedure, called Principal Differential Analysis (PDA), the spline coefficients are estimated by the minimization of a criterion involving a penalty defined using a linear differential operator

$$L = w_0 I + w_1 D + \ldots + w_{m-1} D^{m-1} + D^m$$

where D^l denotes the l-order differential operator. The weights parameter $\{w_l; l = 0, \ldots, m-1\}$ are then estimated by the minimization of the integrated squares of the errors in the predicted derivative. This approach can be applied to the resolution of ODE when the dynamics of the observed state function is defined by a linear differential equation.

I.3.3.5 ODE-penalized smoothing approaches

I.3.3.5.1 Iterative Principal Differential Analysis Poyton et al. (2006) proposed a generalization of this PDA approach, called iPDA for iterative Principal Differential Analysis, that jointly estimates the spline coefficients and the ODE parameter $\boldsymbol{\theta}$ using (in 1 dimension) the ODE-model based penalty in the fitting criterion:

$$\sum_{k=1}^{n} \left(y_k - \widetilde{x}\left(t_k\right)\right)^2 + \lambda PEN, \tag{I.3}$$

where $\widetilde{x}\left(t\right)$ is the B-spline approximation of the state function evaluated at time t. The penalty term PEN corresponds to the integral of the squared deviations of the smoothed curves from the ODE model. The penalty parameter λ enables to reach a balance between goodness-of-fit and solving the ODE equations, large values of λ increasing the relative importance of the 2nd component. In this iterative two step procedure, the first step of the PDA approach is replaced by the estimation of the spline coefficients through the minimization of the penalized fitting criterion using the last updated value of the ODE parameter. The ODE parameter $\boldsymbol{\theta}$ is then updated by minimizing the ODE-penalty term PEN.

I.3.3.5.2 Generalized profiling estimation procedure Ramsay et al. (2007) proposed a variation of the iPDA approach. Instead of having an inner and outer optimization based on the same ODE-penalized fitting criterion, the generalized smoothing approach (that will be latter mostly called *ODE-penalized B-spline* approach in this thesis) consists first to estimate the spline coefficients using the complete ODE-penalized fitting criterion given in Eq. (I.3). Then the ODE parameter $\boldsymbol{\theta}$ is updated by the minimization (outer optimization) of the sum of the squared errors:

$$\sum_{k=1}^{n} \left(y_k - \widetilde{x}\left(t_k\right)\right)^2.$$

With this outer optimization, the spline fit is assumed to be adequately described by the set of differential equations. Iteration between the inner and the outer optimization continue until some convergence criteria is met. This approach is considered as the starting point of the thesis,

see Chapter II for more details.

I.3.3.5.3 Advantages and weaknesses of the frequentist approaches The use of an ODE-penalized fitting criterion in Poyton et al. (2006) and Ramsay et al. (2007) is a trade off between fitting the data and solving the ODE. These methods converge quickly. It can be explained by the compactness of the support of each B-spline basic functions. They give also less biased and more precise estimates than Varah's approach. They are less sensitive to the initial condition of the state function and they allow estimating unobserved state functions. Furthermore, the approximation of the solution and the estimation of parameters involved in the ODE model are not influenced anymore by the number of basis elements provided that this number is large enough to describe the shape and the variations of the state function.
Nevertheless, in these approaches, two critical points have to be highlighted. The first point is the delicate choice of the smoothing parameter that control the relative emphasis between goodness-of-fit and solving the differential equation. Ramsay et al. (2007) proposed either to adjust this parameter manually or to select it automatically using a model selection criterion. In both cases, the adjusted/selected value of the smoothing parameter does not statistically validate or discard the proposed ODE model. Secondly, the approximation made for the variance of the ODE parameter estimate turns to have a strong influence on the coverage probability of confidence intervals.

I.3.3.5.4 Bayesian generalized profiling estimation procedure Based on this frequentist approach proposed by Ramsay et al. (2007), two Bayesian specifications of this ODE-penalized B-spline approach were given by Campbell (2007).
The first proposal aims to generalize the frequentist ODE-penalized B-spline approach by using the ODE-model based penalty to construct the prior distribution for the spline coefficients:

$$p\left(\boldsymbol{c}|\boldsymbol{\theta},\lambda\right) \propto \exp\left(-\lambda PEN\right),$$

where $\boldsymbol{c}$ denotes the vector of spline coefficient used for the spline approximation of the state functions. This approach was not successful to adapt the standard Bayesian P-spline approach (Lang and Brezger, 2004) to the resolution of ODEs for at least three reasons. First of all, an important normalizing constant was erroneously neglected in the derivation of the proposed prior distribution for the spline coefficients, leading to an incorrect joint posterior distribution for the spline coefficients, the ODE parameter and the smoothing parameters. Secondly, the prior distribution for the smoothing parameter λ gave too much prior weight to small values,

indicating no prior confidence on the proposed ODE model to describe the dynamic of the state function. Finally the initial conditions of the state functions were not integrated in the modelling process.
The second proposal made by Campbell combines the ODE penalty term with other prior information on the ODE parameter to construct the prior distribution for the ODE parameters and relies on a sequence of parallel MCMC chains corresponding to increasing values of ODE-adhesion parameters. There, two sources of uncertainty are neglected. The first one is related to the estimation of the spline coefficients as point estimates for these are plugged-in for given values of the ODE-parameters and of the ODE-adhesion parameters. The second one is a consequence of the selection of the ODE-adhesion parameters. Thus, that method is not fully Bayesian, leading to under-estimation of standard errors and potentially misleading credible intervals for the ODE parameters and for the state functions.

I.4 Discussion

The collocation approach has the advantage to capture the shape of the observed data in order to estimate the trajectory of state functions. In contrast to the single shooting approach or the multiple shooting approach, initial conditions of the state functions are not necessarily require in order to approximate the trajectories of the state functions. In terms of computational time, the ODE-penalized B-spline approach has the advantage of being less demanding.

Some negative points about the frequentist ODE-penalized B-spline approach (Ramsay et al., 2007) and the actual Bayesian approaches (Campbell, 2007) have already been mentioned in this chapter. The following requirements can easily be dealt in a fully Bayesian framework and will be treated throughout this thesis:

- Some uncertainty will be included in the specification of the initial condition of the state function, i.e. the initial condition of the state function will be questioned using a prior distribution that will either describe the complete confidence about the initial conditions or some doubt about them.
- Uncertainty measures about the model parameters will be quantified without any approximation.
- The selection of the ODE-adhesion parameters will be automatic and will no more be related to a selection criterion thanks to a prior distribution for these parameters.

- Hierarchical structure will be easily managed by specifying interindividual variability of the ODE parameters.
- Inference on the ODE parameters will be done without necessarily requiring the rebuild of the state function at each iteration of the MCMC procedure.
- The ODE-adhesion parameters will to be used in order to quantify the suitability of a proposed differential equations to model the dynamics of the observed state functions. Selection of competing ODE models with uncertainty measure with respect to this selection will also be introduced using ODE-adhesion parameters.

Chapter II

Bayesian P-spline estimation in hierarchical models specified by systems of affine differential equations

Abstract

A full Bayesian approach based on ODE-penalized B-splines is proposed to jointly estimate ODE-parameters and state function from affine 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. The basic model is extended 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: Bayesian ODE-penalized B-spline; Dynamic systems; ODE-model selection; Ordinary differential equations; Parameter estimation.

II.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:

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{f}(\boldsymbol{x}, t, \boldsymbol{\theta}), t \in [0; T], \tag{II.1}$$

where $\boldsymbol{f}$ is a known affine 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 $y_{jk} = x_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$ the corresponding measurement. This means, that in this setting not all state functions are necessarily observed and that for each observed state function, the number of observation points could differ.

In this chapter, we propose a fully Bayesian approach to jointly estimate the ODE parameters and the state functions involved in affine systems of differential equations when the error distributions are supposed to be Gaussians. We deliberately focus on the affine case to assess and to highlight the important components in the model specification.

In Section II.2, we first give a brief overview of the generalized profiling estimation procedure for ODE models (Ramsay et al., 2007). 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 II.3 presents the strategies used to explore the joint posterior distribution with Markov Chain Monte Carlo (MCMC). Section II.4 generalizes the Bayesian ODE-penalized B-spline approach to multi-subject studies. Section II.5 gives the results of some simulations. It justifies the specific choice for the prior distribution for the ODE-adhesion parameter and it compares the performance of our approach with those of traditional methods. Application on real data are explained and illustrated in Section II.6. Conclusions and generalizations of the proposed method are discussed in Section II.7.

II.2 Bayesian ODE-penalized B-spline approach

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 explain how to generalize this approach in a full Bayesian framework.

II.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, Cao and Zhao (2008); Poyton et al. (2006); Ramsay et al. (2007) proposed to use a B-spline basis function expansion combined with a penalty related to the system of differential equations.

Denote by $\boldsymbol{B_j}(t)$ the K_j-vector of B-spline basis functions of order p evaluated at time t defined using m_j inner knots $\eta_{j,1} \leq \ldots \leq \eta_{j,m_j}$ and two p-multiple knots $\eta_{j,0}$ and η_{j,m_j+1} corresponding

respectively to the lower and the upper bound of the study interval. These $m_j + 2$ knots are typically placed equidistantly on the time interval of the study. Each basis function 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 II.1 presents a B-spline basis of order 4 (solid grey curves), also called cubic B-splines, with knots at each tenth between 0 and 1 (light grey vertical dotted lines).

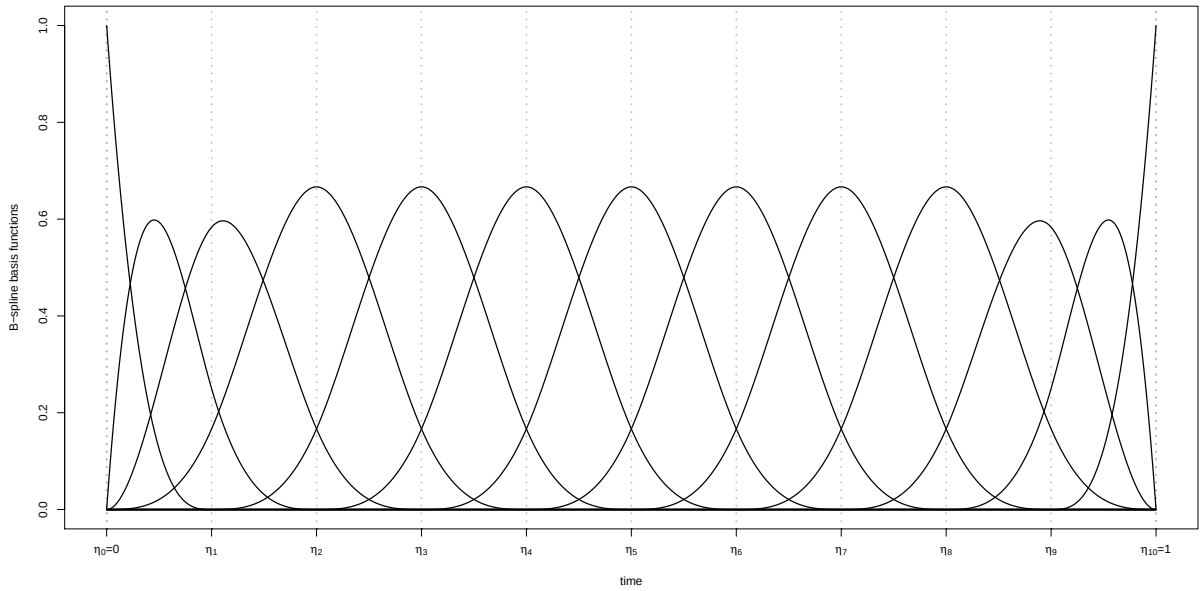


Figure II.1 – Cubic B-splines (solid curves) with inner knots (dotted grey vertical lines) at each tenth between $[0; T]$.

This B-spline basis vector is used to approximate the jth component of the state function $\boldsymbol{x}(t)$. The approximation $\widetilde{x}_j(t)$ of $x_j(t)$ is expressed as a linear combination of these B-spline basis functions:

$$\begin{aligned} \widetilde{x}_j(t) &= \sum_{k=1}^{K_j} B_{jk}(t)\, c_{jk} \\ &= \left(\boldsymbol{B_j}(t)\right)^T \boldsymbol{c_j}, \end{aligned}$$

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 $x_j(t)$ (with the possibility to increase the order p or the number m_j of inner knots). Note that in most of the case, the same B-spline basis is used for the approximation of each state function. That choice for the spline basis has many advantages. First it permits to directly link the initial condition of the state function

(i.e. $x_j(0)$) to the first component of the basis function expansion (i.e. 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 $x_j(t)$. More details on B-spline properties are available in de Boor (2001).
Let $\boldsymbol{\epsilon_j}$ denote the vector of standardized errors associated to the observed variable $\boldsymbol{y_j}^T = \left(y_{j1},\ldots,y_{jn_j}\right)$, $j \in \mathcal{J}$. In practice, one usually assumes independent Gaussian errors with mean 0 and standard deviation equal to 1. Then, the data fitting criterion can be taken to be the log-likelihood

$$H(\boldsymbol{c},\boldsymbol{\tau}|\boldsymbol{y}) = \sum_{j\in\mathcal{J}}\left\{\frac{n_j}{2}\log(\tau_j) - \frac{\tau_j}{2}\left\|\boldsymbol{y_j}-\boldsymbol{B_j}\boldsymbol{c_j}\right\|^2\right\},$$

where $\boldsymbol{B_j}$ is the $n_j \times K_j$-matrix with $(\boldsymbol{B_j}(t_{jk}))^T$ on kth line $(k=1,\ldots,n_j)$. 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. These two extreme cases are illustrated in Figure II.2. In this setting, observations are simulated using the oral 2-compartment model (presented in Section I.2.2.3), with additive Gaussian noise. On the left part of the figure, too few basis functions are considered, leading to an inefficient approximation of the state function. On the right part of the figure, the B-spline approximation just interpolates 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 system of differential equations, a differential equation operator is introduced:

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

The proximity between the approximation of the state function, $\widetilde{x}_j(t)$, and the solution $x_j(t)$ of the corresponding jth 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 the interval $[0,T]$ interval which contains all times of measurement. Note that for the affine system of differential equations that we consider here, this jth penalty term is an homogeneous polynomial of degree 2 in the B-spline coefficients. The full fidelity-

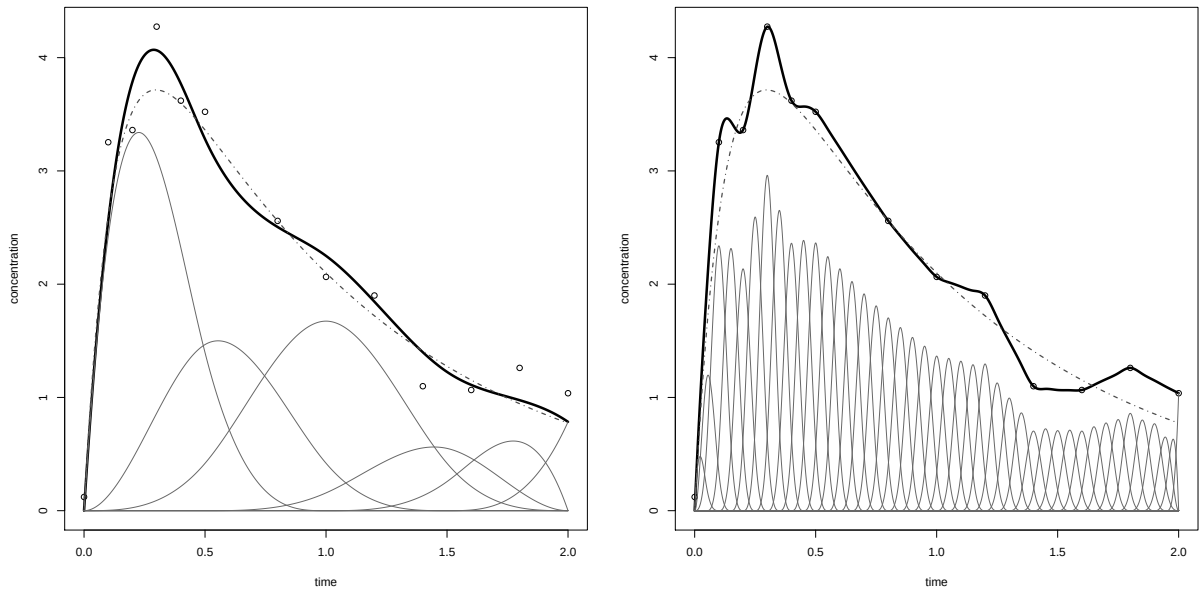


Figure II.2 – Influence of the number of functions in the B-spline basis. Left part: knots at each half-integer are placed between 0 and 2. Right part: knots at each twentieth are placed between 0 and 2. Grey dotdashed curve corresponds to the true curve, black solid curve corresponds to the B-spline approximation and the dark grey solid curves correspond to the B-spline basis function times the corresponding spline coefficient.

to-ODE measure is then given by:

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

where $\boldsymbol{c} = \left({\boldsymbol{c_1}}^T, \ldots, {\boldsymbol{c_d}}^T\right)^T$ is the vector of all spline coefficients of length $K = \sum_{j=1}^{d} K_j$, $\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)$ is a $K \times K$ supermatrix that is constructed by placing the corresponding penalty matrices involved in each $PEN_j\left(\tilde{x}\right)$, $\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)$ is a penalty vector of length K and $l\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)$ is a penalty constant (see Appendix II.A for the automatic construction of these quantities).
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 system of differential equations to describe the dynamics in the system. For a given vector $\boldsymbol{\gamma}$, the spline coefficient estimates are implicitly defined using the current estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$, i.e. as maximizers of the fitting criterion

$$J\left(\boldsymbol{c}|\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{\gamma}\right) = \sum_{j\in\mathcal{J}} \left\{ \frac{n_j}{2}\log\left(\tau_j\right) - \frac{\tau_j}{2}\left\|\boldsymbol{y_j} - \boldsymbol{B_j}\boldsymbol{c_j}\right\|^2 \right\} - \frac{1}{2} PEN\left(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma}\right). \tag{II.3}$$

Then, the data fitting criterion H is optimized with respect to $(\boldsymbol{\theta}, \boldsymbol{\tau})$. As explained in more details in Ramsay et al. (2007), the dependence of H on $(\boldsymbol{\theta}, \boldsymbol{\tau})$ is twofold: directly by its definition, but also implicitly through the spline parameter estimates. Note that, if large values are taken for the γ_js, 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 γ_js 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 $\boldsymbol{c}$. This is illustrated on Figure II.3 where the observations are simulated using the oral 2-compartment model (presented in Section I.2.2.3), with additive Gaussian noise.

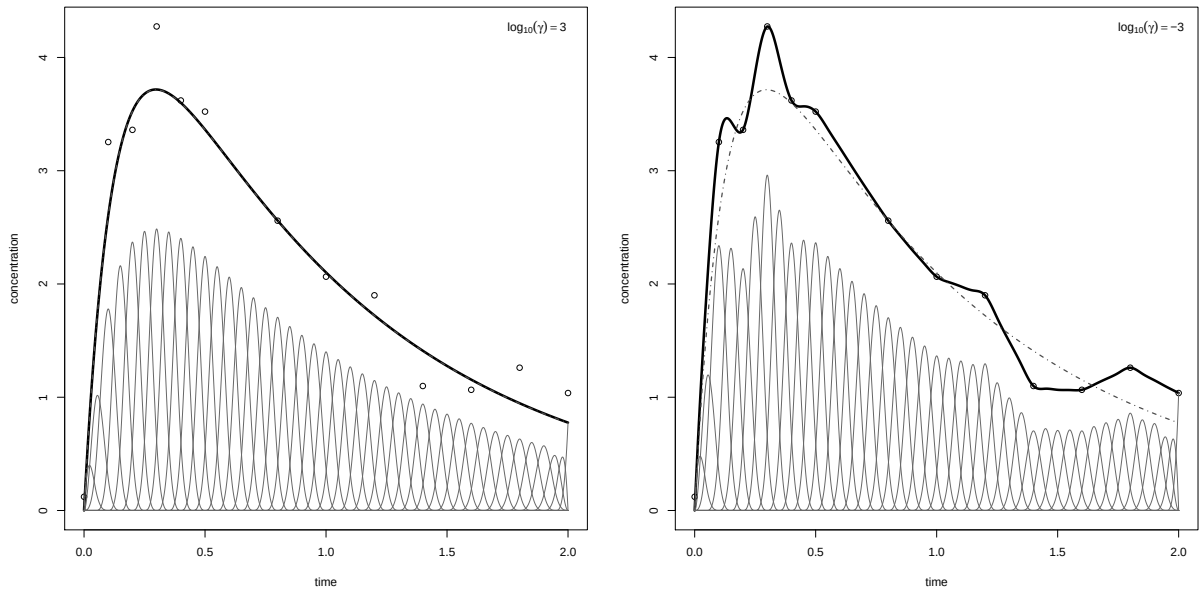


Figure II.3 – Influence of the ODE-adhesion parameter on the B-spline approximation. Grey dotdashed curve corresponds to the true curve, black solid curve corresponds to the B-spline approximation and the dark grey solid curves correspond to the B-spline basis function times the corresponding spline coefficient. Value of the ODE-adhesion parameter is given on the top right corner.

II.2.2 Bayesian ODE-penalized B-spline approach

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

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

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

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

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, we propose to work with the following alternative model specification:

$$\begin{cases} y_{jk}|\boldsymbol{c_j}, \tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B_j}\left(t_{jk}\right)\right)^T \boldsymbol{c_j}; \tau_j^{-1}\right) & \forall j \in \mathcal{J}, \forall k \in \{1, \ldots, n_j\} \\ p\left(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\gamma}\right) & \propto & \exp\left(-\frac{1}{2} PEN\left(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma}\right) - \frac{1}{2}\left(\boldsymbol{c} - \boldsymbol{\mu_c}\right)^T {\boldsymbol{\Sigma_c}}^{-1}\left(\boldsymbol{c} - \boldsymbol{\mu_c}\right)\right), \end{cases} \tag{II.4}$$

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 $x_j\left(0\right)$ of the state function. The corresponding precision in ${\boldsymbol{\Sigma_c}}^{-1}$ allows to express uncertainty with respect to this initial condition. In Section II.3, the corresponding prior distribution for $\boldsymbol{c}$ (including the normalizing constant ignored in (Campbell, 2007) but required in the derivation of the joint posterior distribution) will be given and the case where the precision tends to infinity will be considered in a simple case in Appendix II.B.

Further priors are required to have a full model specification, namely for the precision parameters $\boldsymbol{\tau} = \text{vec}\left(\tau_j; j \in \mathcal{J}\right)$, for the vector $\boldsymbol{\gamma} = \text{vec}\left(\gamma_j; j = 1, \ldots, d\right)$ 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:

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

Note that this prior distribution is a conditional conjugate prior distribution. The prior for 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). This choice for the prior distributions of the ODE-adhesion parameters signals our prior confidence in the specification of the system of differential equations to describe the dynamics in the system. Such a prior for γ_j is rather flat although it puts slightly more weight on value of $\log_{10}(\gamma_j)$ around $\log_{10}\left(b_{\gamma_j}^{-1}\right)$. Choice for the value of the rate is considered in Section II.5.1 and full motivation for this prior distribution will be given in Section IV.2.2. For the vector $\boldsymbol{\theta}$ of differential equation parameters, the chosen prior will depend on the context. Let us denote this prior distribution by $p(\boldsymbol{\theta})$.

II.3 Posterior distribution and parameter estimation

In this section, it will be clearly demonstrated that the key point of this approach is the construction of the normalization constant for the prior distribution for the spline coefficients. We will then explain how to explore the joint posterior distribution using MCMC techniques.

II.3.1 Joint posterior distribution

In order to obtain the joint posterior distribution, it is mandatory to derive the normalization constant for the prior distribution for the spline coefficients $\boldsymbol{c}$ given in Eq. (II.4). Note that

$$\log\left(p\left(\boldsymbol{c}|\boldsymbol{\theta},\boldsymbol{\gamma}\right)\right) \doteq -\frac{1}{2}\left\{\boldsymbol{c}^T\left(\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)+{\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}\right)\boldsymbol{c}-2\boldsymbol{c}^T\left(-\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)+{\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}\boldsymbol{\mu}_{\boldsymbol{c}}\right)+l\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)\right\},$$

is composed of two parts. The first corresponds to the log-density of a multivariate normal distribution with mean $\left(\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)+{\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}\right)^{-1}\left({\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}\boldsymbol{\mu}_{\boldsymbol{c}}-\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)\right)$ and variance-covariance matrix $\left(\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)+{\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}\right)^{-1}$. 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:

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

where $\boldsymbol{v_1}=-\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)+{\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}\boldsymbol{\mu}_{\boldsymbol{c}}$ and $\boldsymbol{V_1}=\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right)+{\boldsymbol{\Sigma}_{\boldsymbol{c}}}^{-1}$. That normalizing constant is required in the derivation of the joint posterior of $(\boldsymbol{c},\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}|\boldsymbol{y})$ and it has a major influence on the posterior distribution. It was missing in Campbell's proposal. It partly explains why a more sophisticated approach based on parallel MCMC chains with a plugged-in point estimate for the spline coefficients was advocated there. In contrast, our approach is simpler and fully

Bayesian. The properties of our estimates are examined in simulation studies in Section II.5 (see also Appendix II.C for a plot of the determinant of the penalty matrix). It suggests that the coverages of our credible regions are in agreement with their nominal value. It also reveals the absolute necessity of the normalizing constant. Unfortunately, an analytic expression for that constant is only available for affine systems of differential equations. Therefore, numerical approximations to that constant will be required to handle nonlinear systems.
We remind that $\mathcal{J}$ represents the subset of $\{1, \dots, d\}$ 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:

$$\begin{aligned}
\log\left(p\left(\boldsymbol{c}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta} | \boldsymbol{y}\right)\right) \doteq & \sum_{j \in \mathcal{J}} \left\{ \frac{n_j}{2} \log\left(\tau_j\right) - \frac{\tau_j}{2} \left\| \boldsymbol{y_j} - \boldsymbol{B_j} \boldsymbol{c_j} \right\|^2 \right\} \\
& - \frac{1}{2} \left\{ \boldsymbol{c}^T \boldsymbol{V_1} \boldsymbol{c} - 2 \boldsymbol{c}^T \boldsymbol{v_1} + \boldsymbol{v_1}^T \boldsymbol{V_1}^{-1} \boldsymbol{v_1} - \log\left(\det\left(\boldsymbol{V_1}\right)\right) \right\} \\
& + \sum_{j \in \mathcal{J}} \left\{ \left(a_{\tau_j} - 1\right) \log\left(\tau_j\right) - b_{\tau_j} \tau_j \right\} \\
& + \sum_{j=1}^{d} \left\{ \left(a_{\gamma_j} - 1\right) \log\left(\gamma_j\right) - b_{\gamma_j} \gamma_j \right\} \\
& + \log\left(p\left(\boldsymbol{\theta}\right)\right) \\
\doteq & \sum_{j \in \mathcal{J}} \left\{ \frac{n_j}{2} \log\left(\tau_j\right) - \frac{\tau_j}{2} \boldsymbol{y_j}^T \boldsymbol{y_j} \right\} - \frac{1}{2} \left\{ \boldsymbol{v_1}^T \boldsymbol{V_1}^{-1} \boldsymbol{v_1} - \log\left(\det\left(\boldsymbol{V_1}\right)\right) \right\} \\
& - \frac{1}{2} \boldsymbol{c}^T \left(\boldsymbol{V_1} + \operatorname{diag}\left(\boldsymbol{Z_j}; j = 1, \dots, d\right)\right) \boldsymbol{c} + \boldsymbol{c}^T \left(\boldsymbol{v_1} + \operatorname{vec}\left(\boldsymbol{z_j}; j = 1, \dots, d\right)\right) \\
& + \sum_{j \in \mathcal{J}} \left\{ \left(a_{\tau_j} - 1\right) \log\left(\tau_j\right) - b_{\tau_j} \tau_j \right\} \qquad \text{(II.5)} \\
& + \sum_{j=1}^{d} \left\{ \left(a_{\gamma_j} - 1\right) \log\left(\gamma_j\right) - b_{\gamma_j} \gamma_j \right\} \\
& + \log\left(p\left(\boldsymbol{\theta}\right)\right),
\end{aligned}$$

where $\operatorname{diag}\left(\boldsymbol{Z_j}; j = 1, \dots, d\right)$ 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 jth state function is observed and the null $K_j \times K_j$-matrix otherwise. The vector $\operatorname{vec}\left(\boldsymbol{z_j}; j = 1, \dots, d\right)$ corresponds to the concatenation of vector $\boldsymbol{z_j}$ that is equal to $\tau_j \boldsymbol{B_j}^T \boldsymbol{y_j}$ if the jth state function is observed and the null K_j-vector otherwise.

II.3.2 Exploring the joint posterior distribution using MCMC

In this section, we explain how to use and how to implement Markov Chain Monte Carlo (MCMC) technique in order to generate samples from posterior distributions in a Bayesian framework. Appendix II.D gives an overview of Metropolis(-Hastings) and Gibbs algorithms.

II.3.2.1 Conditional posterior distributions

From Eq. (II.5), the log conditional posterior distribution for the vector spline coefficients is a given by:

$$\begin{aligned}\log\left(p\left(\boldsymbol{c}|\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{y}\right)\right) &\doteq -\frac{1}{2}\boldsymbol{c}^T\left(\boldsymbol{V_1}+\operatorname{diag}\left(\boldsymbol{Z_j};j=1,\ldots,d\right)\right)\boldsymbol{c}+\boldsymbol{c}^T\left(\boldsymbol{v_1}+\operatorname{vec}\left(\boldsymbol{z_j};j=1,\ldots,d\right)\right)\\ &\doteq -\frac{1}{2}\left\{\boldsymbol{c}^T\boldsymbol{V_2}\boldsymbol{c}-2\boldsymbol{c}^T\boldsymbol{v_2}\right\},\end{aligned}$$

where

$$\boldsymbol{v_2}=\boldsymbol{v_2}\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta},y\right)=\boldsymbol{v_1}+\operatorname{vec}\left(\boldsymbol{z_j};j=1,\ldots,d\right)$$

and

$$\boldsymbol{V_2}=\boldsymbol{V_2}\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}\right)=\boldsymbol{V_1}+\operatorname{diag}\left(\boldsymbol{Z_j};j=1,\ldots,d\right).$$

Therefore, the conditional posterior distribution for the vector spline coefficients is a multivariate normal distribution :

$$\boldsymbol{c}|\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{y}\sim\mathcal{N}_K\left(\boldsymbol{V_2}^{-1}\boldsymbol{v_2};\boldsymbol{V_2}^{-1}\right).$$

For each precision τ_j, $j\in\mathcal{J}$, the log conditional posterior is given by:

$$\log\left(p\left(\tau_j|\boldsymbol{c},\boldsymbol{y}\right)\right) \doteq \left(\frac{n_j}{2}+a_{\tau_j}-1\right)\log\left(\tau_j\right)-\left(\frac{\left\|\boldsymbol{y_j}-\boldsymbol{B_j}\boldsymbol{c_j}\right\|^2}{2}+b_{\tau_j}\right)\tau_j.$$

This conditional posterior distribution 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 $\boldsymbol{V_1}$ and $\boldsymbol{v_1}$. Therefore, a Metropolis-Hastings step was also

used to generate a random sample of $\boldsymbol{\theta}$ from the conditional posterior distribution.

II.3.2.2 Posterior correlation and marginal joint posterior distribution

Strong posterior correlation occurs between the spline coefficients chains and the differential equation parameters chains. It could have an impact in the mixing of the MCMC chains. This strong posterior correlation is illustrated on Figure II.4 using scatterplots of parameter samples generated from the joint posterior using MCMC. The underlying data were simulated from the oral 2-compartment model presented in Section I.2.2.3 with additive Gaussian noise. This figure shows on the left side the posterior correlation between the second spline coefficient used for the approximation of the amount of drug in the stomach and the constant of absorption. On the right side, the figure shows the posterior correlation between the second spline coefficient used for the approximation of the concentration of drug in the blood and the constant of absorption. In order to avoid being forced to deal with this strong posterior conditional correlation, the joint posterior distribution can be marginalized with respect to the spline coefficients. The log of it is given by:

$$\begin{aligned}
\log\left(p\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}|\boldsymbol{y}\right)\right) &\doteq \log\left(\int p\left(\boldsymbol{c},\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}|\boldsymbol{y}\right)d\boldsymbol{c}\right)\\
&\doteq \sum_{j\in\mathcal{J}}\left\{\frac{n_j}{2}\log\left(\tau_j\right)-\frac{\tau_j}{2}\boldsymbol{y_j}^T\boldsymbol{y_j}\right\}\\
&\quad -\frac{1}{2}\left\{\boldsymbol{v_1}^T\boldsymbol{V_1}^{-1}\boldsymbol{v_1}-\log\left(\det\left(\boldsymbol{V_1}\right)\right)-\boldsymbol{v_2}^T\boldsymbol{V_2}^{-1}\boldsymbol{v_2}+\log\left(\det\left(\boldsymbol{V_2}\right)\right)\right\}\\
&\quad +\sum_{j\in\mathcal{J}}\left\{\left(a_{\tau_j}-1\right)\log\left(\tau_j\right)-b_{\tau_j}\tau_j\right\}+\sum_{j=1}^{d}\left\{\left(a_{\gamma_j}-1\right)\log\left(\gamma_j\right)-b_{\gamma_j}\gamma_j\right\}\\
&\quad +\log\left(p\left(\boldsymbol{\theta}\right)\right).
\end{aligned}$$

As $\boldsymbol{V_1}=\boldsymbol{V_1}\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)$, $\boldsymbol{v_1}=\boldsymbol{v_1}\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)$, $\boldsymbol{V_2}=\boldsymbol{V_2}\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}\right)$ and $\boldsymbol{v_2}=\boldsymbol{v_2}\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{y}\right)$ are function of $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$, the conditional distribution of these parameter 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 $\boldsymbol{V_2}^{-1}\boldsymbol{v_2}$ and conditional variance-covariance $\boldsymbol{V_2}^{-1}$. Thanks to this marginalization, inference on the ODE parameters, on the precisions of measurement and on the ODE-adhesion parameters can be done without having to reconstruct the state function at each iteration of the MCMC algorithm.

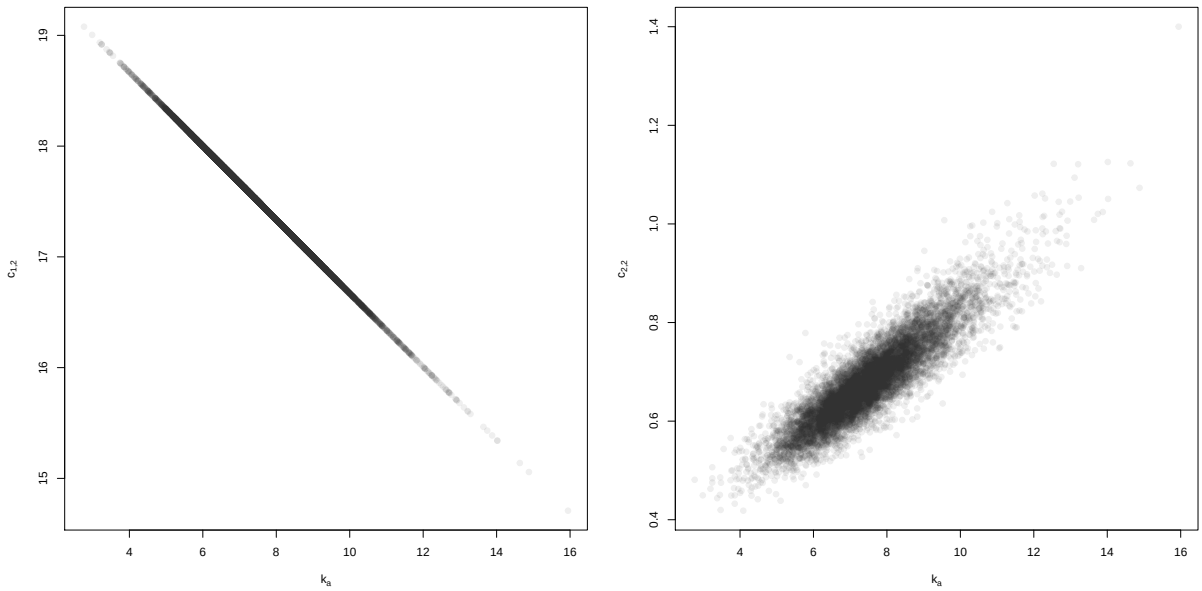


Figure II.4 – Posterior correlation between spline and PK parameters in an oral 2-compartment model: scatterplots of parameter samples generated from the joint posterior using MCMC.

II.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 $p(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\tau}|\boldsymbol{y})$, 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).

II.3.2.4 Re-parametrization using rotation and translation

Posterior correlations also arise among the components of the ODE parameter $\boldsymbol{\theta}$. This posterior correlation is illustrated on Figure II.5. In this setting, data were simulated from the oral 2-compartment model presented in Section I.2.2.3 with additive Gaussian noise. Negative correlations are observed between the constant of absorption and the constant of elimination

and between the constant of elimination and the volume, while positive correlation is observed between the constant of absorption and the volume.

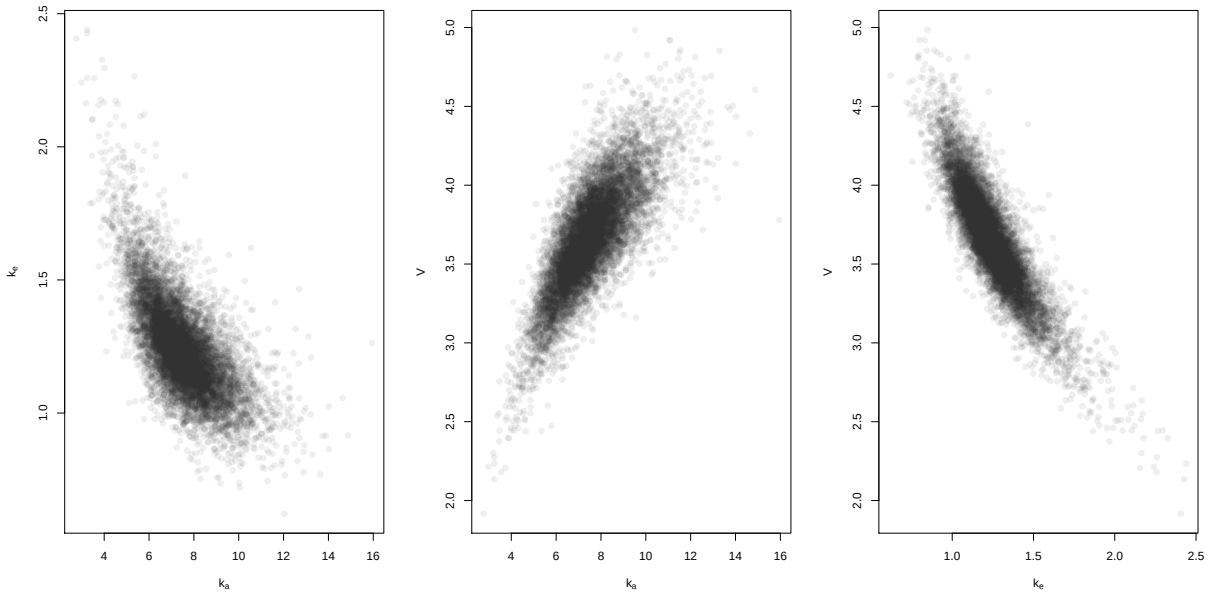


Figure II.5 – Scatter plot between samples from the posterior distribution for the ODE parameters after the analysis of a simulated dataset (oral 2-compartment model)

Therefore, in order 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, reparameterize 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 reparameterize the posterior distribution using $\boldsymbol{\beta}$ such that:

$$\boldsymbol{\theta} = \boldsymbol{L}\boldsymbol{\beta} + \bar{\boldsymbol{\theta}}, \tag{II.6}$$

where $\boldsymbol{L}$ is the lower triangular matrix in the Cholesky decomposition of $\boldsymbol{S}$ with $\boldsymbol{S} = \boldsymbol{L}\boldsymbol{L}^T$. Then we use a Metropolis-Hastings algorithm with adaptive proposals to generate a sample from the posterior distribution for $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$. A sample from the posterior distribution for $\boldsymbol{\theta}$ is obtained by applying Eq. (II.6) on the $\boldsymbol{\beta}$ sample. 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 $\boldsymbol{V_2}^{-1}\boldsymbol{v_2}$ and conditional variance-

covariance $\boldsymbol{V_2}^{-1}$.

II.3.2.5 Implementation

The implementation was done by a combination of `R` code (R Core Team, 2012) and `C` code (Kernighan and Ritchie, 1988). `R` was used for its simplicity and `C` for its execution speed.
For the `C` code, generic codes were programmed for each kind of model. There, all the instructions are given in order to update and then to store the samples of the joint posterior distribution. Several header files were called at the beginning of the `C` code. It is either some already available routines (e.g. header files programmed for `R` functions) or programmed routines (e.g. the log posterior distribution called in the Metropolis steps or the proposal distributions). The arguments of the `C` code (that are also the outputs) are: the data, the design matrix based on the B-spline basis functions, the elementary penalty matrices, the hyperparameters, the initial condition of the state function with the precision that express our prior uncertainty, the initial conditions for the MCMC algorithm, the number of iteration, the burn-in, the thinning rate and some arrays in order to store the samples from the joint posterior distribution.
In `R`, all the elementary elements are setting up. Then the `C` code is called using `.C` function. After execution of the `C` code, traditional checks are made on the samples from the joint posterior distribution: visual checking using trace plots and convergence diagnostics (Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992), Geweke's convergence diagnostic, ...). If the MCMC has converged, then the results can be presented using summary statistics (mean, median, variance, confidence intervals, highest posterior density intervals, ...) and traditional plots (plot of the posterior densities, visual predictive checks, ...).
Implementation and execution of the `R` code, combined with the `C` code was done on a Intel Core i7-860 Processor (2.80GHz) computer.

II.4 Hierarchical Bayesian ODE-penalized B-spline approach for affine dynamic systems

In this section, we introduce the concept of hierarchical Bayesian ODE-penalized B-spline approach for systems of affine differential equations. 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.

II.4.1 Individual system of differential equations

Assume now that observations are collected on I subjects and that the changes in the state $\boldsymbol{x_i}(t) \in \mathbb{R}^d$, $(i = 1, \ldots, I)$ of each subject i is governed by an affine system of differential equations:

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

where $\boldsymbol{\theta_i} \in \mathbb{R}^q$ is the unknown individual vector of ODE-parameters and $\boldsymbol{f}$ 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 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 standardized measurement error ϵ_{ijk}. We denote by $y_{ijk} = x_{ij}(t_{ijk}) + \tau_j^{-1/2}\epsilon_{ijk}$ the corresponding measurement. As in the standard case, it is assumed that the measurement errors ϵ_{ijk} are independent and distributed according to a standardized Gaussian distribution.

II.4.2 B-spline basis functions expansion and ODE-penalty

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

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

where $\boldsymbol{B_{ij}}(t)$ is the K_{ij}-vector of B-spline basis functions evaluated at time t and $\boldsymbol{c_{ij}}^T = \text{vec}(c_{ijk}; k = 1, \ldots, K_{ij})$ the K_{ij}-vector of B-spline coefficients. Note that in most of the case, the same B-spline basis is used for the approximation of each state function for each subject. The total ODE-penalty term is constructed in a three steps procedure. First the jth penalty term for the ith subject is constructed, then the total penalty term for the ith 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 $\widetilde{x}_{ij}(t)$ and the solution $x_{ij}(t)$ is given by:

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

where $L_{i,j,\boldsymbol{\theta_i}}\left(\widetilde{\boldsymbol{x_i}}\right) = \dfrac{d\widetilde{x}_{ij}}{dt}\left(t\right) - f_j\left(\widetilde{\boldsymbol{x_i}}, t, \boldsymbol{\theta_i}\right)$ is the differential equation operator for the jth state function of the ith subject. Then the total penalty term for the ith subject is given by:

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

The vector $\boldsymbol{c_i}$ corresponds to the concatenation of all spline coefficients for the ith subject. The matrix $\boldsymbol{R_i}\left(\boldsymbol{\theta}_i, \boldsymbol{\gamma}\right)$, the vector $\boldsymbol{r_i}\left(\boldsymbol{\theta}_i, \boldsymbol{\gamma}\right)$ and the constant $l_i\left(\boldsymbol{\theta}_i, \boldsymbol{\gamma}\right)$ are constructed in the same way as the penalty terms presented in the standard case (see Appendix II.A). As the same function $\boldsymbol{f}$ 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:

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

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

II.4.3 Bayesian hierarchical model

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

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

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 $\boldsymbol{V_1}^{-1}\boldsymbol{v_1}$ and variance-covariance matrix $\boldsymbol{V_1}^{-1}$ where $\boldsymbol{v_1} = \boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c} - \boldsymbol{r}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})$ and $\boldsymbol{V_1} = \boldsymbol{R}(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma}) + \boldsymbol{\Sigma_c}^{-1}$ and the second part corresponds to a term independent of $\boldsymbol{c}$. Therefore, the normalizing constant

$$(\det(\boldsymbol{V_1}))^{1/2} \exp\left(\frac{1}{2} l(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\gamma})\right) \exp\left(-\frac{1}{2}\boldsymbol{v_1}^T \boldsymbol{V_1}^{-1}\boldsymbol{v_1}\right),$$

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 $p(\boldsymbol{\theta})$. 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. This distribution is generalization to multiple dimensions of the gamma distribution and is traditionally used in the estimation of precision matrix. In the proposed prior distribution, the matrix $\boldsymbol{V}$ corresponds to the prior expected value for the variance-covariance matrix of the random effects and s quantifies the confidence in that specific value, large values of s translating a large prior confidence. Note that s has to be larger than q (dimension of the ODE parameter $\boldsymbol{\theta}$). This Wishart distribution is a conditional conjugate prior distribution.

The strategy used to generate a sample from the joint posterior distribution is similar to the one used in the standard approach, see Section II.3.2.

First, the conditional posterior distribution for the vector $\boldsymbol{c}$ is a multivariate Gaussian distribution with mean $\boldsymbol{V_2}^{-1}\boldsymbol{v_2}$ and variance-covariance matrix $\boldsymbol{V_2}^{-1}$ where:

$$\boldsymbol{v_2} = \boldsymbol{v_1} + \text{vec}\left(\text{vec}\left(\boldsymbol{z_{ij}}; j = 1, \ldots, d\right); i = 1, \ldots, I\right)$$

and

$$\boldsymbol{V_2} = \boldsymbol{V_1} + \text{diag}\left(\text{diag}\left(\boldsymbol{Z_{ij}}; j = 1, \ldots, d\right); i = 1, \ldots, I\right).$$

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. For each $i = 1, \ldots, I$, $j \in \mathcal{J}$, the matrix $\boldsymbol{B_{ij}}$ is a $n_{ij} \times K_{ij}$-matrix with $(\boldsymbol{B_{ij}}(t_{ijk}))^T$ on its kth line ($k = 1, \ldots, n_{ij}$).
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. The log of this marginal joint posterior distribution is equal to:

$$
\begin{aligned}
& \log\left(p\left(\boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_I}, \boldsymbol{\theta}, \boldsymbol{P_\theta} | \boldsymbol{y}\right)\right) \\
\doteq\ & \frac{1}{2} \sum_{i=1}^{I} \sum_{j \in \mathcal{J}} \left\{ n_{ij} \log\left(\tau_j\right) - \tau_j \boldsymbol{y_{ij}}^T \boldsymbol{y_{ij}} \right\} \\
& - \frac{1}{2} \left\{ \boldsymbol{v_1}^T \boldsymbol{V_1}^{-1} \boldsymbol{v_1} - \log\left(\det\left(\boldsymbol{V_1}\right)\right) - \boldsymbol{v_2}^T \boldsymbol{V_2}^{-1} \boldsymbol{v_2} + \log\left(\det\left(\boldsymbol{V_2}\right)\right) \right\} \\
& + \sum_{j=1}^{d} \left\{ \left(a_{\gamma_j} - 1\right) \log\left(\gamma_j\right) - b_{\gamma_j} \gamma_j \right\} \\
& + \sum_{j \in \mathcal{J}} \left\{ \left(a_{\tau_j} - 1\right) \log\left(\tau_j\right) - b_{\tau_j} \tau_j \right\} \\
& + \frac{I}{2} \log\left(\det\left(\boldsymbol{P_\theta}\right)\right) - \frac{1}{2} \sum_{i=1}^{I} \left(\boldsymbol{\theta_i} - \boldsymbol{\theta}\right)^T \boldsymbol{P_\theta} \left(\boldsymbol{\theta_i} - \boldsymbol{\theta}\right) \\
& + \log\left(p\left(\boldsymbol{\theta}\right)\right) \\
& + \frac{s - q - 1}{2} \log\left(\det\left(\boldsymbol{P_\theta}\right)\right) - \frac{1}{2} \mathrm{tr}\left(\boldsymbol{V} \boldsymbol{P_\theta}\right).
\end{aligned}
$$

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

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

The other conditional posterior distributions cannot be identified: Metropolis-Hastings steps with adaptive proposals have been used 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 $\boldsymbol{c}$.

II.5 Simulation study

In this part, we perform some simulations inspired by pharmacokinetic models presented in Section I.2.2. The objective is twofold. The first simulation study aims to provide some guidelines and motivation in the choice of the hyperparameters in the gamma prior for the ODE-adhesion parameters. The last two simulation studies compare the performances of existing estimation strategies with our proposal.

II.5.1 Rate of the gamma prior distribution for the ODE-adhesion parameters

As mentioned in Section II.2.2, the prior distribution adopted for the ODE-adhesion parameters is a gamma prior distribution $\mathcal{G}(a_\gamma, b_\gamma)$ with shape a_γ equal to one and rate b_γ set equal to a small value. The aim of this simulation study is to propose a level of the rate under which some summary statistics (average 95% HPD interval width, mean posterior variance) based on samples from the posterior distribution for the ODE parameter $\boldsymbol{\theta}$ and the vector of precision of measurement $\boldsymbol{\tau}$ are no more altered. The retained pharmacokinetic model is the 1-compartment intravenous bolus model presented in Section I.2.2.2

II.5.1.1 Data

Each simulated dataset corresponds to n measurements made on a single subject at equidistant time points between 0 and 1 ($n = 10, 15, 20, 30, 50, 100, 200$). The simulated observations are:

$$y_k = x(t_k) + \tau^{-1/2}\epsilon_k.$$

The state function is given by:

$$x(t) = \frac{D}{V}\exp(-k_e t),$$

where $D = 1mg$, $k_e = 5h^{-1}$ and $V = 1L$. The precision of measurement τ is set to 100. The distribution for the standardized error term ϵ_k is given by a standard Gaussian distribution. For each simulation, 1000 datasets were generated.

II.5.1.2 Model

For the analysis of the simulated dataset, only the constant of elimination k_e and the precision of measurement are estimated. The volume is set to its simulation value, i.e. $V = 1L$. The state

function $x(t)$ of drug is approximated by $\widetilde{x}(t) = (\boldsymbol{B}(t))^T \boldsymbol{c}$ using a 5-order B-spline expansion with inner knots at each tenth between 0 and 1. The following Bayesian model is considered for the analysis of each simulated dataset:

$$\begin{cases} y_k|\boldsymbol{c},\tau & \sim & \mathcal{N}\left((\boldsymbol{B}(t_k))^T \boldsymbol{c}; \tau^{-1}\right) & \forall k \in \{1,\ldots,n\} \\ p(\boldsymbol{c}|\gamma,k_e) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T\boldsymbol{V_1}\boldsymbol{c} - 2\boldsymbol{c}^T\boldsymbol{v_1}\right)\right) & \\ \tau & \sim & \mathcal{G}(1;10^{-6}) & \\ \gamma & \sim & \mathcal{G}(1;b_\gamma) & \\ p(k_e) & \propto & \mathbb{I}\{k_e > 0\}, & \end{cases}$$

where $\boldsymbol{V_1} = \boldsymbol{R}(k_e,\gamma) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Appendix II.A for the specification of $\boldsymbol{R}(k_e,\gamma)$). The first component of $\boldsymbol{\mu_c}$ is set equal to $D = 1$ and the corresponding precision in ${\boldsymbol{\Sigma_c}}^{-1}$ to 10^{10}. The other components in $\boldsymbol{\mu_c}$ and ${\boldsymbol{\Sigma_c}}^{-1}$ are null. A grid of rate coefficients (namely 9 equidistant values for $-\log_{10}(b_\gamma)$ between 2 and 10) for the prior distribution for the ODE-adhesion parameter are considered for the analysis of each simulated dataset.

II.5.1.3 Simulation results

Figure II.6 shows the mean posterior variance for k_e and τ with respect to $-\log_{10}(b_\gamma)$ and Figure II.7 shows the average 95% HPD interval width for k_e and τ. Whatever the sample size, the mean posterior variances are decreasing to plateaus and are not influenced by the rate b_γ of the gamma prior distribution is $-\log_{10}(b_\gamma)$ is larger than 6. This is confirmed on the average 95% HPD interval width for k_e and τ that are reaching plateaus for value of $-\log_{10}(b_\gamma)$ over 6. We therefore propose to work with rate b_γ equal (or smaller) to 10^{-8}. Note that with shape equal to one and rate equal to 10^{-8} for the gamma prior distribution, the 2.5% quantile is equal to 6.40 and the 97.5% quantile is equal to 8.57.

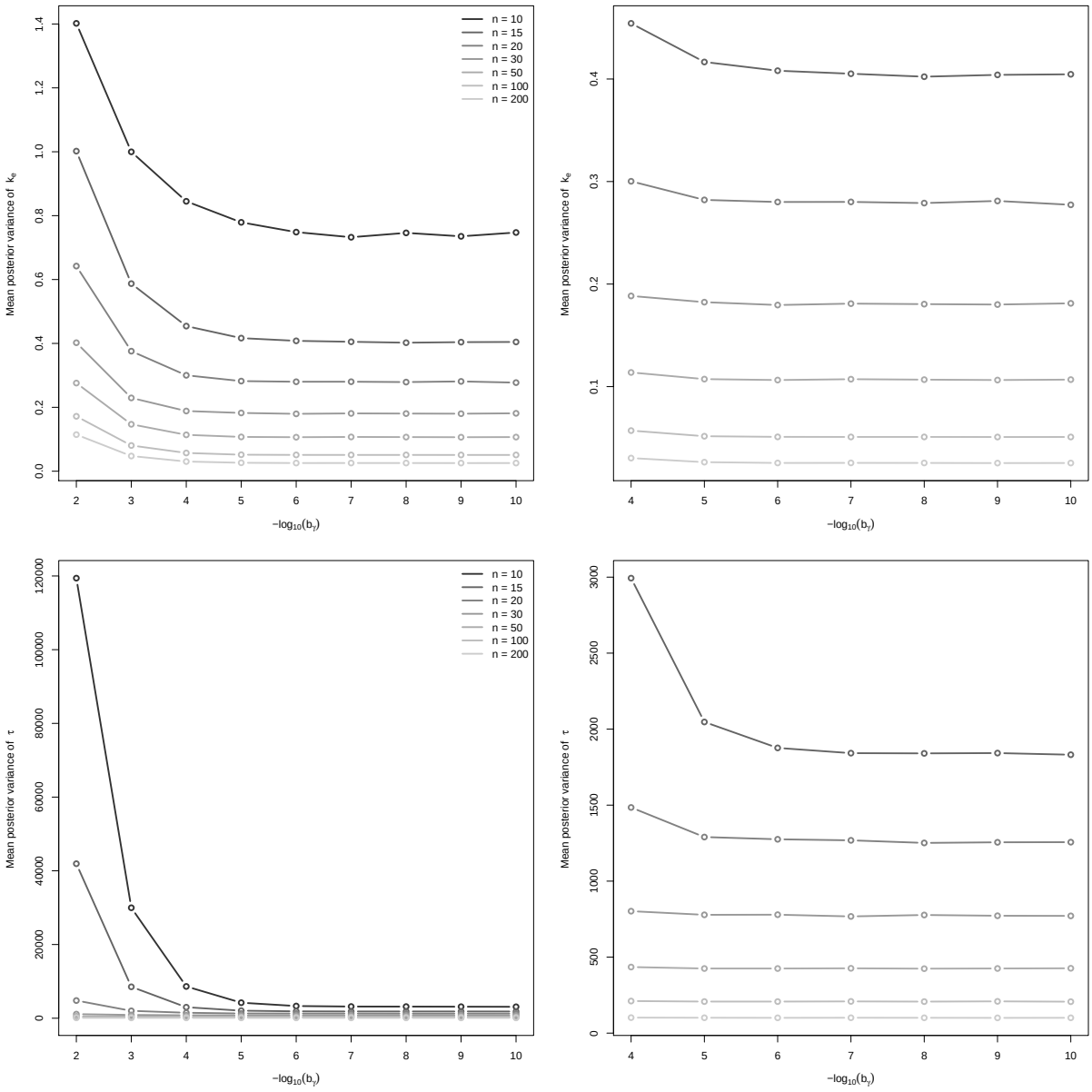


Figure II.6 – 1-compartment intravenous bolus model - Mean posterior variance of k_e (top figures) and τ (bottom figures) with respect to $-\log_{10}(b_\gamma)$ with sample size equal to 10, 15, 20, 50, 100 and 200.

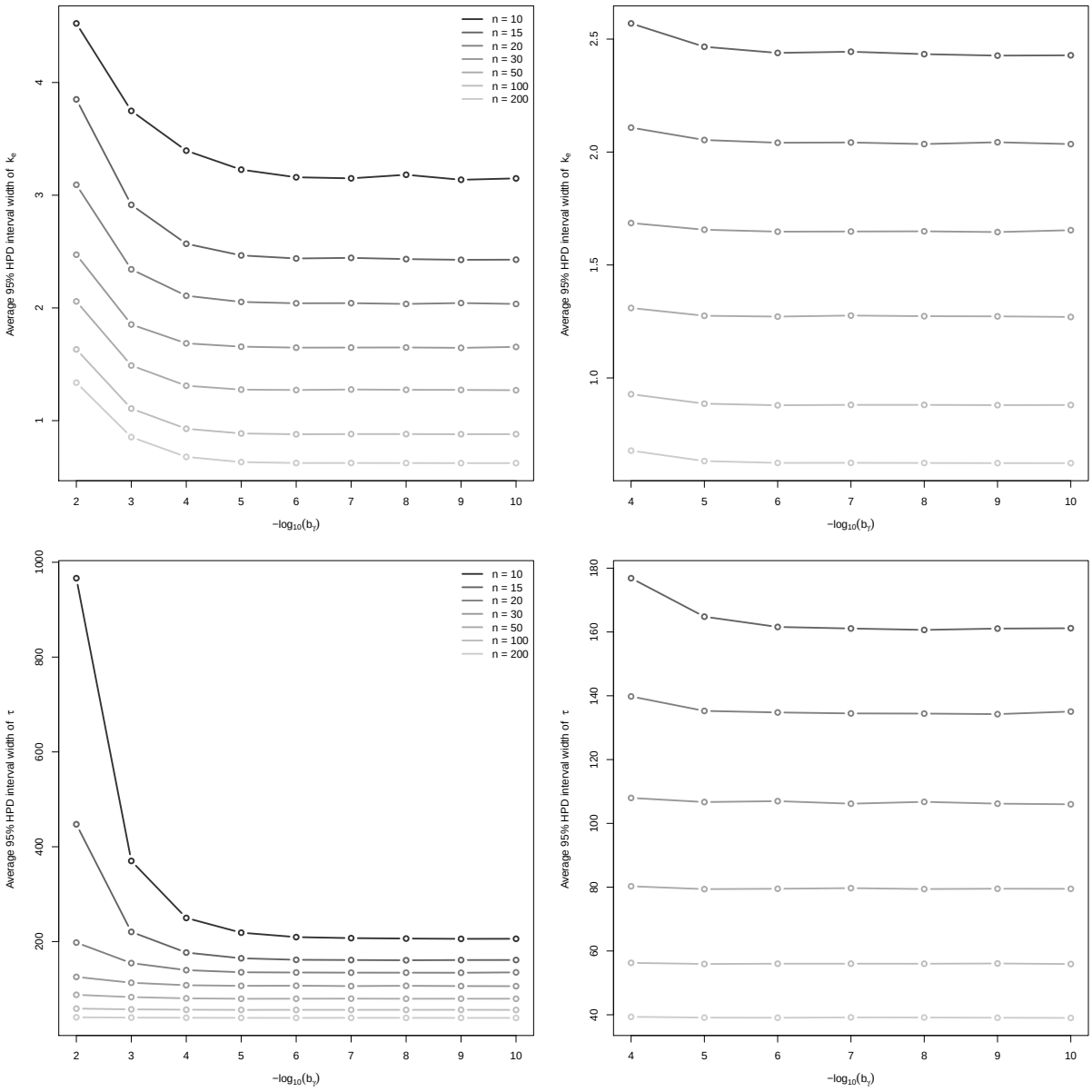


Figure II.7 – 1-compartment intravenous bolus model - Average 95% HPD interval width of k_e (top figures) and τ (bottom figures) with respect to $-\log_{10}(b_\gamma)$ with sample size equal to 10, 15, 20, 50, 100 and 200.

II.5.2 Performance of the Bayesian ODE-penalized B-spline approach compared to traditional methods

The aim of the two next simulation studies is to compare the performance of the Bayesian ODE-penalized B-spline (B-ODE-P-spline) approach with traditional approaches: the frequentist general profiling estimation (FGPE) method (Ramsay et al., 2007), the Bayesian general profiling estimation (BGPE) method proposed by Campbell (2007), the standard Bayesian (BES) and frequentist (FES) methods using the explicit solution to the system of differential equations.

II.5.2.1 Preliminary remarks

In order to compare all the considered approaches, 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. (II.3), modified in order to take into account that the first component is fixed. A search grid is done for each ODE-adhesion parameter γ, starting from small values and gradually increasing. The selected ODE-adhesion parameters correspond to the minimizer of the BIC criterion:

$$BIC = -2H\left(\boldsymbol{c}, \boldsymbol{\tau}|\boldsymbol{y}\right) + npar \log\left(n\right),$$

where $npar$ is the number of effective parameters and n the sample size. Note that the method proposed in Ye (1998) was used to compute the number of effective parameters at each grid point.
For the BGPE method, the first component of each vector of splines coefficients $\boldsymbol{c_j}, j = 1, \ldots, d$ is also fixed to the initial condition of the corresponding state function. The model description given in Campbell (2007) is modified in order to take into account that the first component of each vector of splines coefficients is fixed.
Concerning our B-ODE-P-spline method, the model specification in Section II.2.2 is used (and, thus, includes $\boldsymbol{\mu_c}$ 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 B-ODE-P-spline, BGPE and BES methods, the posterior mean, 80% and 95% highest

posterior density 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.

Finally, 95% credibility interval for the estimated coverage probabilities that contain the nominal values (80% or 95%) are considered compatible with their nominal values (see results in Sections II.5.2.2.3 and II.5.2.3.3).

II.5.2.2 1-compartment intravenous bolus model

The 1-compartment intravenous bolus model was presented in Section I.2.2.2. For this simulation, we consider that the concentration is only observed on a single subject.

II.5.2.2.1 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 $x(t)$ of the 1-compartment intravenous bolus model and some measurement error:

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

where $D = 20mg$, $k_e = 5h^{-1}$, $V = 5L$ and $\epsilon \sim \mathcal{N}(0; \sigma^2)$. 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.

II.5.2.2.2 Models For the FGPE, BGPE and the B-ODE-P-spline methods, the concentration $x(t)$ of drug is approximated by $\tilde{x}(t) = \frac{1}{V}\left(\boldsymbol{B}(t)\right)^T \boldsymbol{c}$ using a 5-order B-spline expansion with inner knots at each tenth between 0 and 1. Therefore, the number of B-spline basis functions is equal to $K = 14$.
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. (II.3). For the BGPE method, the first spline coefficient c_1 is fixed to the dose D. The Bayesian model is an adaptation of the one proposed in Campbell

(2007) with the following non-informative prior distributions: $\gamma \sim \mathcal{G}(1; 10^{-8})$, $\tau \sim \mathcal{G}(1; 10^{-8})$ and $p(k_e, V) \propto \mathbb{I}\{k_e > 0, V > 0\}$.
For the B-ODE-P-spline approach, the following Bayesian model is considered:

$$\begin{cases} y_k|\boldsymbol{c}, \tau, V & \sim & \mathcal{N}\left(V^{-1}\left(\boldsymbol{B}(t_k)\right)^T \boldsymbol{c}; \tau^{-1}\right) & \forall k \in \{1, \ldots, n\} \\ p(\boldsymbol{c}|\gamma, k_e, V) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{V_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) & \\ \gamma, \tau & \sim & \mathcal{G}(1; 10^{-8}) & \\ p(k_e, V) & \propto & \mathbb{I}\{k_e > 0, V > 0\}, & \end{cases}$$

where $\boldsymbol{V_1} = \boldsymbol{R}(k_e, V, \gamma) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Appendix II.A 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^{10}. 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: $x(t) = \frac{D}{V}\exp(-k_e t)$. 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:

$$\begin{cases} y_k|\tau, k_e, V & \sim & \mathcal{N}(x(t_k); \tau^{-1}) & \forall k \in \{1, \ldots, n\} \\ \tau & \sim & \mathcal{G}(1; 10^{-8}) & \\ p(k_e, V) & \propto & \mathbb{I}\{k_e > 0, V > 0\}. & \end{cases}$$

For each Bayesian approach, the Metropolis-Hastings algorithm with adaptive proposals has been implemented with 40 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). 300 MCMC iterations (with the first 100 discarded) were used to reparameterize the problem by applying a translation and a rotation to the ODE-parameters (see Section II.3.2.4). Then, for each parallel chain, 1000 MCMC iterations were generated (with the first 500 discarded). For information only, the analysis of one simulated dataset takes up to 5 seconds for BGPE and B-ODE-P-spline approaches on an Intel Core i7-860 Processor (2.80GHz).

II.5.2.2.3 Simulation results Table II.1, Table II.2 and Table II.3 summarize 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 the volume V and the constant of elimination k_e for the three different values of σ when the FGPE, FES, BGPE, B-ODE-P-spline and BES methods are used.

For volume V, the relative bias seems to be the same for all methods except for the BGPE approach. For medium and large value of σ, B-ODE-P-spline and BES methods slightly overestimated it. For the B-ODE-P-spline and the BES approaches, root mean squared error for V are similar but slightly larger than the one observed in the FES approach. This is due to the choice of the posterior mean as a point estimate in the Bayesian approaches while the marginal posteriors are slightly skewed. For the FGPE method, the number of outliers for the relative error is large, especially when $\sigma = 0.5$. The estimated coverage probabilities for FGPE are well below nominal values, whereas for the FES, B-ODE-P-spline and BES 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 the FGPE, FES, B-ODE-P-spline and BES methods. For the BGPE approach, the bias explodes when σ is larger than 0.1. RMSE for BES and B-ODE-P-spline approaches are similar and slightly smaller than with FES, suggesting a smaller dispersion of the fully 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. Only the coverage probabilities for FES, B-ODE-P-spline and BES credible regions are in agreement with their nominal values.
Table II.4 summarizes the average 80% and 95% HPD interval width and average posterior variance for volume V and constant of elimination k_e for the B-ODE-P-splines and BES approaches. It simply appears that all quantities are indistinguishable.
The first learning point of this simulation study concerns the absolute necessity to use the complete constant of normalization in the prior distribution of the spline coefficients for the derivation of the joint posterior: it was partly missing in Campbell's (BGPE) proposal. Furthermore, BES and B-ODE-P-spline approaches appear to give similar results.

Table II.1 – 1-compartment intravenous bolus model - $\sigma = 0.1$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for FGPE, FES, BGPE, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values.

		$\sigma = 0.1$ ($\tau = 100$)			
		RB	CP80%	CP95%	RMSE
$V = 5L$	FGPE	-0.0	43.3	58.4	0.11
	FES	-0.0	77.0	**93.8**	0.10
	BGPE	-0.9	76.3	93.0	0.16
	B-ODE-P-spline	0.3	**77.5**	**94.7**	0.16
	BES	0.3	77.1	**94.2**	0.16
$k_e = 5h^{-1}$	FGPE	-0.2	94.6	99.2	0.30
	FES	0.2	**77.6**	93.4	0.16
	BGPE	1.1	73.8	92.7	0.12
	B-ODE-P-spline	0.1	**77.5**	**94.9**	0.10
	BES	0.1	**77.8**	**94.8**	0.10

Table II.2 – 1-compartment intravenous bolus model - $\sigma = 0.3$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for FGPE, FES, BGPE, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values.

		$\sigma = 0.3$ ($\tau \approx 11.11$)			
		RB	CP80%	CP95%	RMSE
$V = 5L$	FGPE	0.1	42.1	60.3	0.33
	FES	0.1	**79.1**	92.6	0.30
	BGPE	-54.2	6.4	97.7	2.81
	B-ODE-P-spline	1.5	**78.4**	**94.0**	0.50
	BES	1.5	**78.7**	**93.8**	0.49
$k_e = 5h^{-1}$	FGPE	-0.8	93.7	98.9	0.61
	FES	0.4	**78.1**	93.1	0.49
	BGPE	26270	9.3	13.9	1556
	B-ODE-P-spline	0.8	**79.1**	93.6	0.30
	BES	0.8	**78.9**	**93.8**	0.30

Table II.3 – 1-compartment intravenous bolus model - $\sigma = 0.5$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for FGPE, FES, BGPE, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values.

		$\sigma = 0.5$ ($\tau = 4$)			
		RB	CP80%	CP95%	RMSE
$V = 5L$	FGPE	1.0	44.0	58.5	0.56
	FES	0.5	**79.0**	92.8	0.50
	BGPE	-58.7	0.0	100.0	2.94
	B-ODE-P-spline	5.0	77.4	93.5	0.90
	BES	5.0	**77.7**	**93.8**	0.90
$k_e = 5h^{-1}$	FGPE	-2.4	90.1	97.4	1.02
	FES	1.7	**78.4**	92.8	0.83
	BGPE	53875	0.0	0.1	2761
	B-ODE-P-spline	2.6	**79.1**	**93.8**	0.55
	BES	2.7	**79.5**	**93.8**	0.55

Table II.4 – 1-compartment intravenous bolus model - Average 80%-HPD interval width (P1), average 95%-HPD interval width (P2) and average posterior variance (P3) for B-ODE-P-spline and BES methods for the three different levels of standard deviation σ.

		$\sigma = 0.1$			$\sigma = 0.3$			$\sigma = 0.5$		
		P1	P2	P3	P1	P2	P3	P1	P2	P3
V	B-ODE-P-spline	0.396	0.622	2.565e-2	1.195	1.890	2.439e-1	2.099	3.376	8.480e-1
	BES	0.396	0.622	2.562e-2	1.194	1.890	2.436e-1	2.098	3.375	8.493e-1
k_e	B-ODE-P-spline	0.247	0.389	1.002e-2	0.743	1.174	9.397e-2	1.296	2.081	3.204e-1
	BES	0.247	0.389	1.001e-2	0.743	1.175	9.392e-2	1.297	2.082	3.209e-1

II.5.2.3 Oral 2-compartment model

The oral 2-compartment model was presented in Section I.2.2.3. For this simulation, we suppose that the concentration is observed on a single subject. Note that in this setting, one unobserved state function (the amount of drug in the stomach) is involved in the ODE model.

II.5.2.3.1 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 oral 2-compartment model and some measurement error:

$$y_k = x_2(t_k) + \epsilon_k,$$

where $D = 20mg$, $k_a = 8h^{-1}$, $k_e = 2h^{-1}$, $V = 1L$ and $\epsilon \sim \mathcal{N}(0; \sigma^2)$. We consider three different cases for the standard deviation σ: $\sigma = 0.3$, 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.

II.5.2.3.2 Models First of all, the BGPE approach is not considered here for two main reasons. Firstly, the preceding simulation study has clearly shown that the Bayesian approach proposed by Campbell (2007) is not efficient. Secondly, there is no proposed extension in Campbell (2007) of BGPE that permits to consider ODE models with several differential equations. For the FGPE and the B-ODE-P-spline methods, we approximate the amount $x_1(t)$ of drug in the stomach by $\widetilde{x_1}(t) = (\boldsymbol{B_1}(t))^T \boldsymbol{c_1}$ and the concentration $x_2(t)$ of drug in the central compartment by $\widetilde{x_2}(t) = (\boldsymbol{B_2}(t))^T \boldsymbol{c_2}$ using a 5-order B-spline expansion with inner knots at each tenth between 0 and 2. Therefore, the dimension of vectors $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ is equal to $K_1 = K_2 = 24$. For the FGPE method, the first spline coefficient $c_{1,1}$ is fixed to the dose D and $c_{2,1}$ is fixed to 0. PK-parameters k_a, k_e and V, the spline coefficients $\boldsymbol{c_{1,2:K_1}}$ and $\boldsymbol{c_{2,2:K_2}}$ and precision τ of measurements are estimated by minimizing the criterion J presented in Eq. (II.3).
The Bayesian model of the B-ODE-P-spline approach is:

$$\begin{cases} y_k|\boldsymbol{c_2}, \tau & \sim & \mathcal{N}\left(\left(\boldsymbol{B_2}(t_k)\right)^T \boldsymbol{c_2}; \tau^{-1}\right) & \forall k \in \{1, \ldots, n\} \\ p(\boldsymbol{c}|\gamma_1, \gamma_2, k_a, k_e, V) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{V_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) & \\ \gamma_1, \gamma_2, \tau & \sim & \mathcal{G}(1; 10^{-8}) & \\ p(k_a, k_e, V) & \propto & \mathbb{I}\{k_a > k_e > 0, V > 0\}, & \end{cases}$$

where $\boldsymbol{V_1} = \boldsymbol{R}(k_a, k_e, V, \gamma_1, \gamma_2) + {\boldsymbol{\Sigma_c}}^{-1}$ and $v_1 = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Appendix II.A for the automatic construction of $\boldsymbol{R}(k_a, k_e, V, \gamma_1, \gamma_2)$). The first component of $\boldsymbol{\mu_c}$ is set equal to D, the (K_1+1)th component of $\boldsymbol{\mu_c}$ to 0 and the corresponding precision in ${\boldsymbol{\Sigma_c}}^{-1}$ to 10^{10}. 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: $x_2(t) = \frac{D}{V}\frac{k_a}{k_a - k_e}(\exp(-k_e t) - \exp(-k_a t))$. For the FES method, PK-parameters k_a, 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:

$$\begin{cases} y_k|\tau, k_a, k_e, V & \sim & \mathcal{N}(x_2(t_k); \tau^{-1}) & \forall k \in \{1, \ldots, n\} \\ \tau & \sim & \mathcal{G}(1; 10^{-8}) & \\ p(k_a, k_e, V) & \propto & \mathbb{I}\{k_a > k_e > 0, V > 0\}, & \end{cases}$$

For each Bayesian approach, the Metropolis-Hastings algorithm with adaptive proposals has been implemented with 40 parallel chains. 300 MCMC iterations (with the first 100 discarded) were used to reparameterize the problem by applying a translation and a rotation to the ODE-parameters. Then, for each parallel chain, 1000 MCMC iterations were generated (with the first 500 discarded). The analysis of one simulated dataset takes up to 50 seconds for the B-ODE-P-spline approach on an Intel Core i7-860 Processor (2.80GHz).

II.5.2.3.3 Simulation results Table II.5, Table II.6 and Table II.7 summarize results of the simulation study. It gives the relative bias, the coverage probabilities at level 80% and 95% and the root mean squared error for the estimation of the constant of absorption k_a, the constant of elimination k_e and the volume V for three values of σ for the FGPE, FES, B-ODE-P-spline 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. Coverage probabilities of the estimators are in agreement with the nominal values, except for the FGPE approach where it tends to be smaller than the targeted value: this could be due to a bad approximation of the variance of the estimate, a poor choice for the estimator distribution 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 due to the choice of the posterior mean as a point estimate for the parameter (with a skewed

posterior distribution). It is compensated by a smaller dispersion of the estimators. Coverage probabilities for FES, B-ODE-P-spline and BES approaches are similar and in agreement with their nominal values. For the FGPE approach, the estimated coverage probabilities are below the nominal value.

For the volume V of the central compartment, all methods perform equally in term of RMSE. B-ODE-P-spline and BES approaches 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, B-ODE-P-spline and BES approaches, estimated coverage probabilities are almost in agreement with their nominal values.

Table II.8 reports the average 80% and 95% HPD interval width and average posterior variance for the constant of absorption k_a, the constant of elimination k_e and the volume V for the B-ODE-P-splines and BES approaches. As in 1-compartment intravenous bolus model, nothing can differentiate clearly the two Bayesian approaches.

This simulation study confirms the results previously obtained: FES, B-ODE-P-spline and BES approaches perform equally and outperform the FGPE approach, especially in term of coverage probabilities.

Table II.5 – Oral 2-compartment model - $\sigma = 0.3$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for F-ODE-P-spline, FES, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values.

		$\sigma = 0.3$ ($\tau \approx 11.11$)			
		RB	CP80%	CP95%	RMSE
$k_a = 8h^{-1}$	FGPE	0.5	74.7	90.4	0.43
	FES	0.2	**80.3**	**93.8**	0.36
	B-ODE-P-spline	0.2	**79.9**	**93.7**	0.36
	BES	0.2	**79.4**	**93.8**	0.36
$k_e = 2h^{-1}$	FGPE	-0.0	73.2	90.4	0.07
	FES	0.0	**79.1**	**93.7**	0.06
	B-ODE-P-spline	0.2	**78.8**	**94.3**	0.06
	BES	0.2	**78.7**	**94.2**	0.06
$V = 8h^{-1}$	FGPE	0.1	73.7	90.1	0.03
	FES	0.0	**79.4**	93.4	0.02
	B-ODE-P-spline	-0.0	**79.0**	**93.7**	0.02
	BES	-0.1	**78.7**	**94.0**	0.02

Table II.6 – Oral 2-compartment model - $\sigma = 0.6$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for F-ODE-P-spline, FES, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values.

		$\sigma = 0.6$ ($\tau \approx 2.78$)			
		RB	CP80%	CP95%	RMSE
$k_a = 8h^{-1}$	FGPE	2.0	73.0	92.9	0.95
	FES	0.6	**80.0**	**94.4**	0.72
	B-ODE-P-spline	0.4	**79.1**	**94.1**	0.74
	BES	0.4	**79.9**	**94.6**	0.74
$k_e = 2h^{-1}$	FGPE	-0.2	71.1	91.6	0.14
	FES	0.2	**78.8**	**94.7**	0.11
	B-ODE-P-spline	0.9	**79.1**	**95.5**	0.12
	BES	0.9	**79.0**	**95.6**	0.12
$V = 8h^{-1}$	FGPE	0.4	71.9	91.8	0.05
	FES	0.1	**80.2**	**94.6**	0.04
	B-ODE-P-spline	-0.2	**80.0**	**94.9**	0.05
	BES	-0.2	**80.1**	**94.9**	0.05

Table II.7 – Oral 2-compartment model - $\sigma = 0.9$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for F-ODE-P-spline, FES, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values.

		$\sigma = 0.9$ ($\tau \approx 1.23$)			
		RB	CP80%	CP95%	RMSE
$k_a = 8h^{-1}$	FGPE	4.4	77.9	94.1	1.53
	FES	0.4	**80.3**	**94.8**	1.08
	B-ODE-P-spline	-0.2	**79.4**	**93.8**	1.15
	BES	-0.2	**79.2**	**93.7**	1.15
$k_e = 2h^{-1}$	FGPE	-0.1	76.9	92.1	0.21
	FES	1.1	**82.3**	**95.2**	0.17
	B-ODE-P-spline	3.4	**78.5**	**95.4**	0.21
	BES	3.4	**79.2**	**95.3**	0.21
$V = 8h^{-1}$	FGPE	0.6	77.1	92.7	0.08
	FES	-0.5	**80.3**	**96.0**	0.07
	B-ODE-P-spline	-1.5	**80.8**	**95.1**	0.07
	BES	-1.5	**80.7**	**94.7**	0.07

Table II.8 – Oral 2-compartment model - Average 80%-HPD interval width (P1), average 95%-HPD interval width (P2) and average posterior variance (P3) for B-ODE-P-spline and BES methods for the three different levels of standard deviation σ.

		$\sigma = 0.3$			$\sigma = 0.6$			$\sigma = 0.9$		
		P1	P2	P3	P1	P2	P3	P1	P2	P3
k_a	B-ODE-P-spline	8.991e-1	1.396	1.285e-1	1.828	2.858	5.414e-1	2.835	4.477	1.332
	BES	9.003e-1	1.398	1.286e-1	1.829	2.859	5.417e-1	2.837	4.477	1.332
k_e	B-ODE-P-spline	1.405e-1	2.199e-1	3.237e-3	2.899e-1	4.599e-1	1.503e-2	4.755e-1	7.873e-1	4.915e-2
	BES	1.406e-1	2.188e-1	3.165e-3	2.899e-1	4.597e-1	1.504e-2	4.755e-1	7.865e-1	4.902e-2
V	B-ODE-P-spline	5.502e-2	8.567e-2	4.822e-4	1.126e-1	1.768e-1	2.086e-3	1.787e-1	2.864e-1	5.542e-3
	BES	5.503e-2	8.552e-2	4.790e-4	1.126e-1	1.768e-1	2.084e-3	1.788e-1	2.861e-1	5.535e-3

II.6 Application

In this section, we propose one application of the B-ODE-P-spline approach that focuses on the modeling of perfusion ratio after an ischemia on mice.

II.6.1 Data and differential equation model

The perfusion ratio dataset (van Weel et al., 2007) was presented in Section I.2.1. We recall that we focus on the ischemic to the non-ischemic perfusion ratio $x(t)$ of $I = 12$ mice from a specific group called C57BL/6. This perfusion ratio was measured $n = 6$ times over a 28 day period (see Figure I.1).
The dynamics of the perfusion ratio can be described by this differential equation:

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

As explained in Chapter I, some convergence issues can be encountered by using the explicit solution in a frequentist framework. Note also that the occlusion was certainly not perfect at the surgery time. Therefore, some doubt should be taken into account in the specification of the initial condition of the perfusion ratio.

II.6.2 Model

The perfusion ratio $x_i(t)$ for subject i is approximated by $\widetilde{x}_i(t) = (\boldsymbol{B}(t))^T \boldsymbol{c_i}$, a linear combination of 5-order B-splines from a basis corresponding to 57 equidistant knots between 0 and 28 days. Note that the same B-spline basis is used for all subjects and that only the spline coefficients are subject-specific. The ODE parameters α and β are considered in the log scale: $l\alpha$ and $l\beta$ denote the logarithm of α and β. The Bayesian model used to analyze these data is:

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

where $\boldsymbol{V_{i1}} = \boldsymbol{R}(\alpha, \beta_i, \gamma) + {\boldsymbol{\Sigma_{c_i}}}^{-1}$ and $\boldsymbol{v_{i1}} = -\boldsymbol{r}(\alpha, \beta_i, \gamma) + {\boldsymbol{\Sigma_{c_i}}}^{-1}\boldsymbol{\mu_{c_i}}$ (see Appendix II.A 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 mice except those from 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 ratios. 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 out of four possible models considering parameters $l\alpha$ and $l\beta$ either as a fixed parameter or as a random effect using the deviance information criterion (Spiegelhalter et al., 2002). In this selected model, $l\alpha$ and $l\beta$ are the mean mean population parameters and $\tau_{l\beta}$ is the precision are random effect for parameter $l\beta$ that describes the interindividual variability.

II.6.3 Results

For the Metropolis-Hastings algorithms with adaptive proposals, we generate 10 parallel chains with 1000 MCMC iterations and discard the first 300 realizations. We use this first sample to reparameterize 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 II.8 (resp. Figure II.9) shows traces and histograms of $\sigma = 1/\sqrt{\tau}$, $\log_{10}(\gamma)$ and $\tau_{l\beta}$ (resp. $l\alpha$ and $l\beta$). For each parameter chains, the Gelman-Rubin statistic was computed and was less than the reference value 1.05 (Gelman and Rubin, 1992), suggesting convergence of the MCMC algorithm.
Table II.9 reports the posterior mean, 2.5%, 50% and 97.5% quantiles of the ODE parameters, precision of measurement and ODE-adhesion parameter. The posterior 2.5% quantile for parameter $l\alpha$ suggests that the perfusion ratio will tend to a value above 1 after 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 quantified on 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 II.10 shows the fitted individual perfusion ratio with 80% and 95% credibility sets. To obtain this figure, the conditional mean $\widetilde{x_i}(t)$ of the perfusion ratio for the mouse i is reconstructed on a grid of times for each realization of the posterior distribution for the spline coefficients for this mouse. Then, at each point of the grid of times, 2.5%, 10%, 50%, 90% and 97.5% quantiles are computed. Figure II.11 shows the credibility interval for the posterior

predictive distribution of the perfusion ratio for each mouse. Three observed perfusion ratio are outside the 95% predictive credibility interval and 10 observed perfusion ratio are outside the 80% credibility interval which is compatible with the expected coverages.
For each mouse, the posterior probability that the individual mean perfusion ratio is greater than 1 is larger than 0.95, except for subject 12 where this probability is around 0.8. At the end of the study, i.e. at day 28, the posterior probability that the mean population perfusion ratio is above 1 is larger 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 larger than 1 after day 28 is equal to 0.8: this suggests recovery of a large majority of mice by the end of the study.

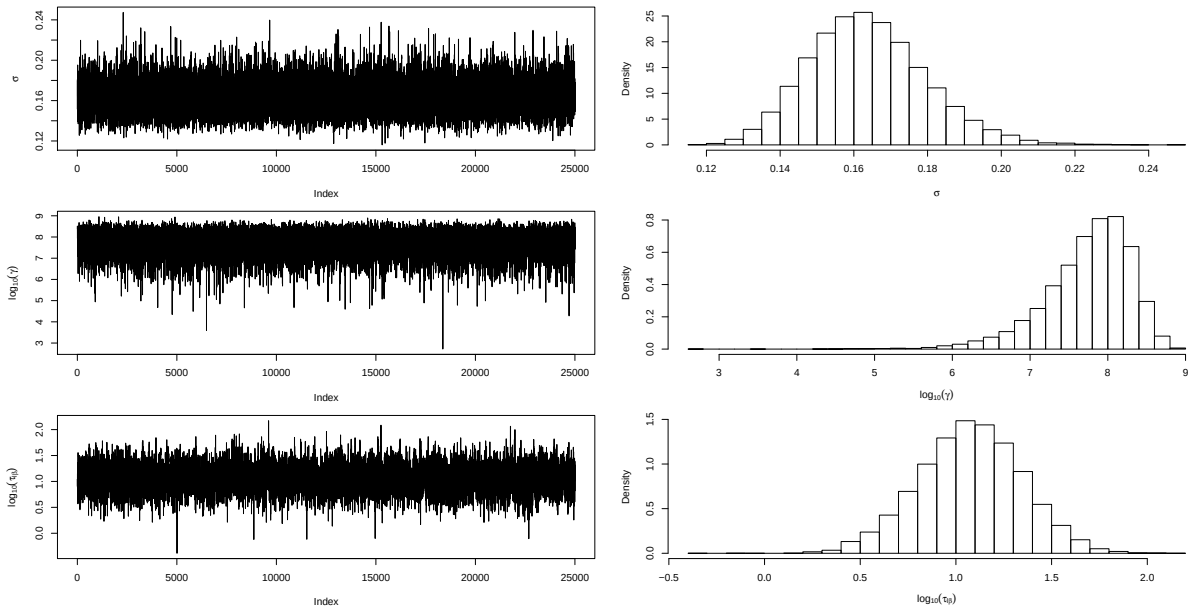


Figure II.8 – Traces and histograms for σ, $\log_{10}(\gamma)$ and $\log_{10}(\tau_{l\beta})$

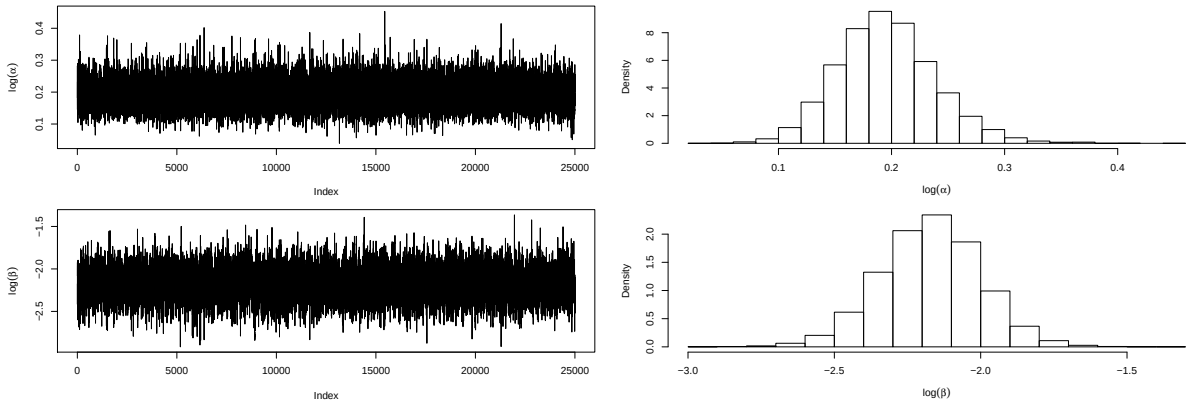


Figure II.9 – Traces and histograms for $l\alpha = \log(\alpha)$ and $l\beta = \log(\beta)$

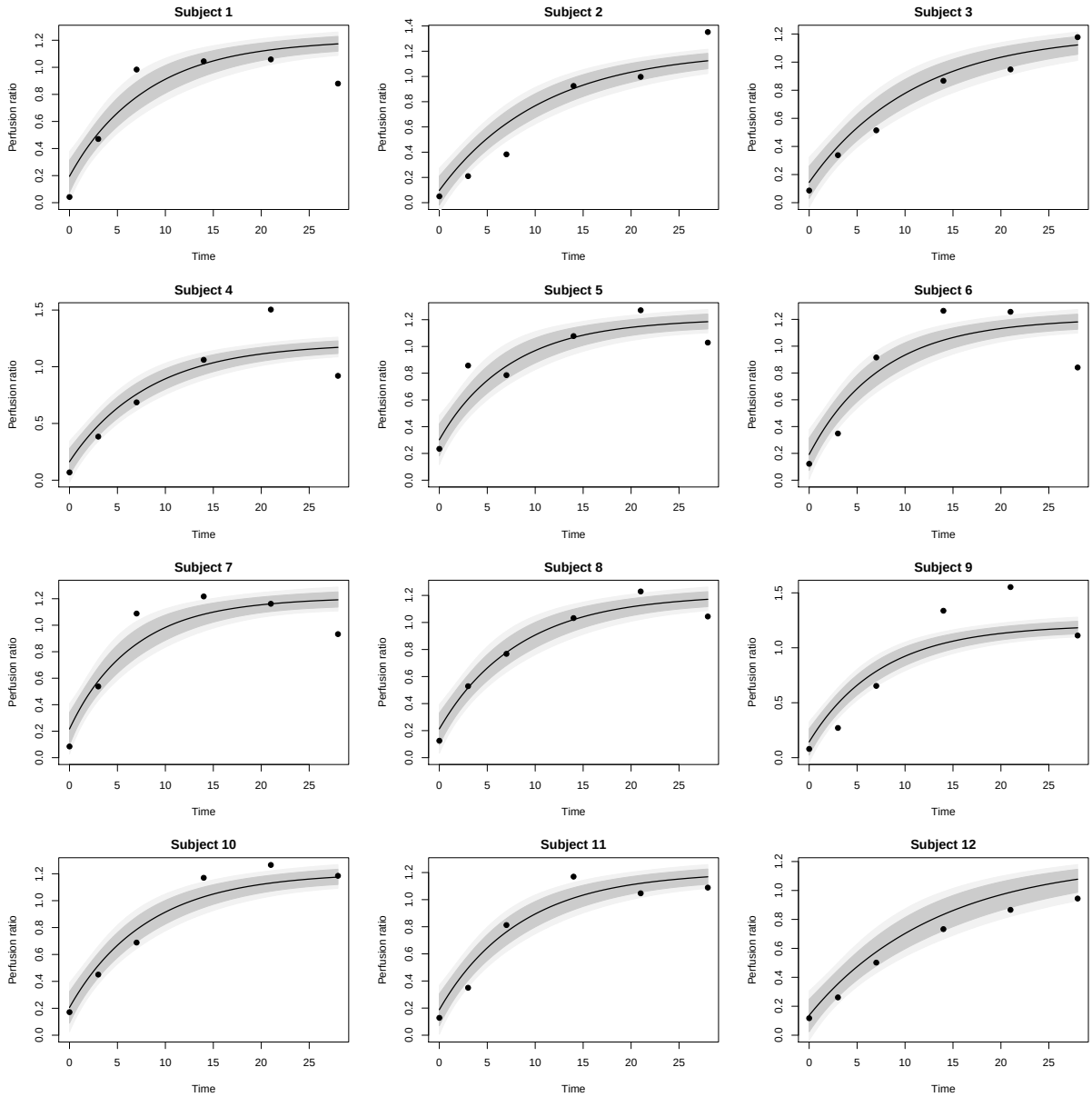


Figure II.10 – Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted perfusion ratio with pointwise median (black solid curve) and observed perfusion ratio (solid circles) versus time for the twelve mice.

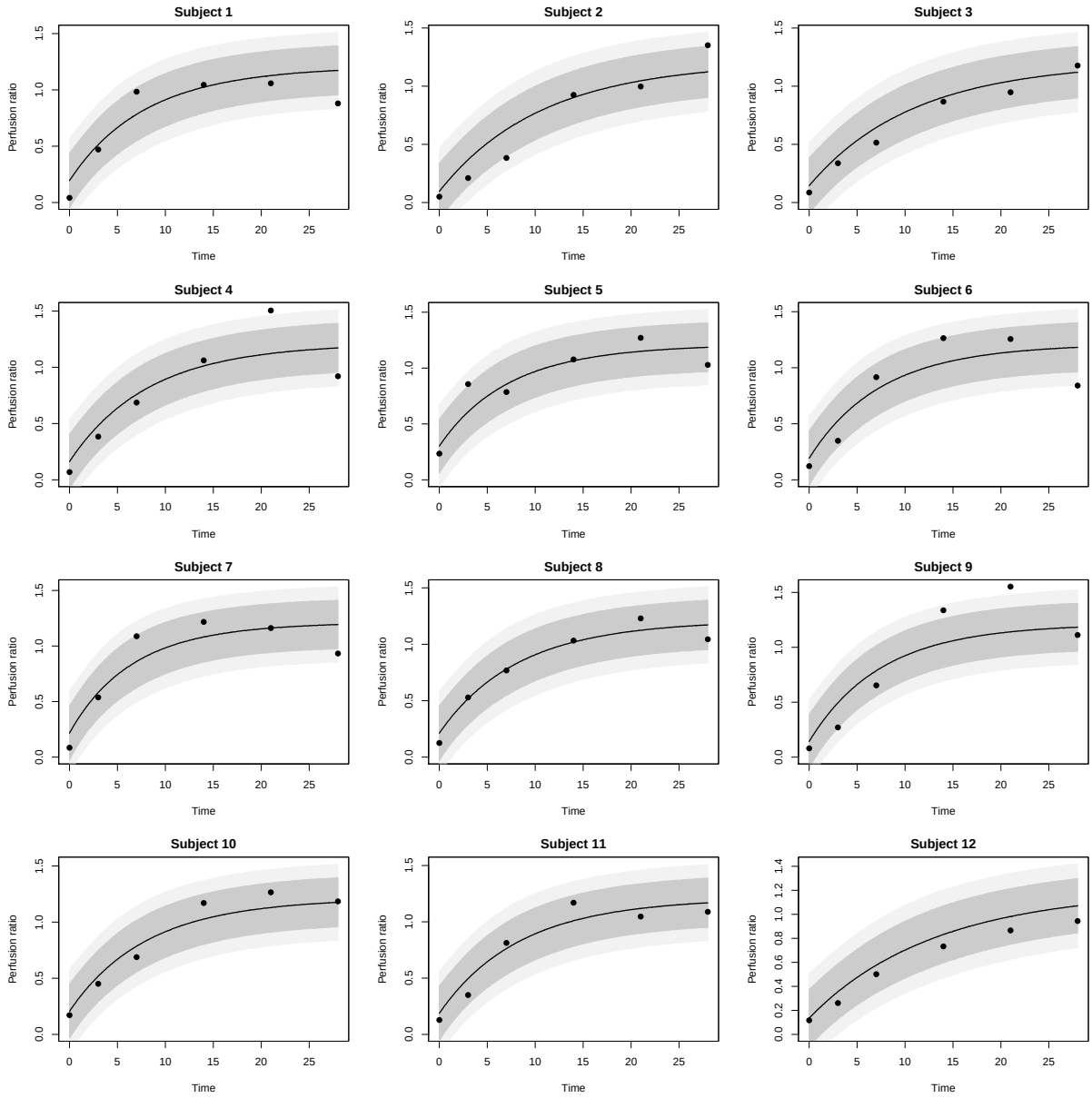


Figure II.11 – Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the perfusion ratio with pointwise posterior predictive median (black solid curve) and observed perfusion ratio (solid circles) versus time for the twelve mice.

Table II.9 – Estimated parameters with credibility sets

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

II.7 Discussion

In this chapter, we have presented a fully Bayesian model based on ODE-penalized B-spline approach to jointly estimate the state functions and parameters of systems of affine differential equations in a hierarchical setting when the error distribution is Gaussian.
This approach involves a linear combination of B-splines for the approximation of the state functions with flexibility counter-balanced by a penalty term related to the system of differential equations. There, we underlined the absolute necessity to include the constant of normalization in the prior distribution for the spline coefficients for the derivation of the joint posterior distribution. It keeps also the advantage of linearity whereas approaches that used the explicit solution of the affine ODE have to manage its nonlinearity.
Simulations suggest that the proposed Bayesian ODE-penalized B-spline approach is superior to the frequentist one in term of estimation of the ODE parameters, coverage probabilities and prediction. When an explicit nonlinear solution to the ODE system is available and when frequentist (FES method) or Bayesian (BES method) estimation can be handled (cf. nonlinearity), the simulation study suggests that the B-ODE-P-spline approach is equally performing.
We have also seen that the basic Bayesian ODE-penalized B-spline approach can be extended to a hierarchical setting by specifying the inter-individual variability of the ODE parameters. This hierarchical model offers the advantage to fit individual state functions for each subject. This was not available in Ramsay et al. (2007)
The possible use of prior information on the ODE parameters and on the precision of measurement is a definite advantage. The proposed model also enables to include prior 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.
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 intrinsically included in the model. We have also shown that the posterior distribution for the ODE parameter and for the precisions of measurement are not altered by the choice of the rate of the gamma prior distribution of the ODE-adhesion parameters if this rate is small (e.g. smaller than 10^{-6}).

Appendix

II.A Penalty matrix in a general system of affine differential equations

In this appendix, 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:

$$\begin{cases} \dfrac{dx_1}{dt}(t) &= \displaystyle\sum_{k=1}^{d} a_{1k} x_k(t) + b_1 \\ \quad\vdots & \\ \dfrac{dx_d}{dt}(t) &= \displaystyle\sum_{k=1}^{d} a_{dk} x_k(t) + b_d, \end{cases}$$

where a_{jk} is the kth coefficient used in the jth differential equation. If the state function $x_k(t)$ is not involved in the jth 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 jth differential equation of the system.
The penalty term for the jth differential equation is:

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

where the integration is over an interval of length $\mathcal{L}$. 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:

$$\boldsymbol{P_{jk}^{ab}} = \int \boldsymbol{B_j}^{(a)}(t) \left(\boldsymbol{B_k}^{(b)}(t) \right)^T dt$$

and

$$\boldsymbol{p}_{\boldsymbol{j}}^{\boldsymbol{a}} \;=\; \int \boldsymbol{B_j}^{(a)}(t)\,dt.$$

Then the penalty term for the jth differential equation is equal to:

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

The overall penalty term $PEN = \sum_{j=1}^{d} \gamma_j PEN_j$ is therefore 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 symmetric block matrix of dimension $K \times K$ with (k,l)th block of dimension $K_k \times K_l$ defined by:

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

where $k, l \in \{1, \ldots, d\}$ and $\delta_{kl} = 1$ 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 $\mathcal{L} \sum_{j=1}^{d} \gamma_j b_j^2$.

II.B Prior distribution for the spline parameters

In order to understand the limit case in the prior distribution for the spline parameter, we will consider the simplest dynamic system, namely the exponential decay problem:

$$\begin{cases} \dfrac{dx}{dt}(t) &= -\theta x(t) \\ x(0) &= x_0. \end{cases}$$

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:

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

and

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

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

$$\begin{aligned} \boldsymbol{V_{c,prior}} &= \begin{pmatrix} R_1^1 + \tau_{x_0} & \boldsymbol{R_1^{2:K}} \\ \boldsymbol{R_{2:K}^1} & \boldsymbol{R_{2:K}^{2:K}} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} s_D^{-1} & -s_D^{-1}\boldsymbol{R_1^{2:K}}\left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1} \\ -\left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1}\boldsymbol{R_{2:K}^1}s_D^{-1} & \left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1} + \left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1}\boldsymbol{R_{2:K}^1}s_D^{-1}\boldsymbol{R_1^{2:K}}\left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1} \end{pmatrix}, \end{aligned}$$

where $s_D = R_1^1 + \tau_{x_0} - \boldsymbol{R_1^{2:K}}\left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1}\boldsymbol{R_{2:K}^1}$.
Note that, when τ_{x_0} tends to infinity then $\boldsymbol{V_{c,prior}}$ tends to $\begin{pmatrix} 0 & 0 \\ 0 & \left(\boldsymbol{R_{2:K}^{2:K}}\right)^{-1} \end{pmatrix}$.

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 c_1 of the spline parameters is fixed (variance is equal to zero and the correlation with the other spline parameters $\boldsymbol{c_{2:K}}$ is null).
Using the same block matrix decomposition, the expression of the mean of the prior distribution of the spline parameters is:

$$\boldsymbol{\mu_{c,prior}} = \begin{pmatrix} s_D^{-1}\tau_{x_0}x_0 \\ -\left(\boldsymbol{R^{2:K}_{2:K}}\right)^{-1}\boldsymbol{R^{1}_{2:K}}s_D^{-1}\tau_{x_0}x_0 \end{pmatrix}.$$

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

II.C Impact of the normalizing constant for the prior distribution of the spline coefficients

The aim of this appendix is to confirm the major role of the normalizing constant for the prior distribution of the spline coefficients. To this end, we consider the study case of Section 4.1.1 of Campbell's thesis. It consists on a simple linear ODE model for unbounded exponential growth:

$$\frac{dx}{dt}(t) = \theta x(t)$$

with initial condition $x(0) = 1$. The value of θ is set to 1.2 and $n = 50$ spaced observations are simulated between $t = 1$ and $t = 3$ with an additive Gaussian noise with standard deviation $\sigma = 0.75$.
The solution of the differential equation is approximated by $\widetilde{x}(t) = (\boldsymbol{B}(t))^T \boldsymbol{c}$ where $\boldsymbol{B}(t)$ is the vector of B-spline basis evaluated at time t and $\boldsymbol{c} \in \mathbb{R}^K$ is the vector of spline coefficients. Note that no prior information about the initial condition of the state function is used in the model specification.

II.C Impact of the normalizing constant for the prior distribution of the spline coefficients

The model proposed by Campbell is as followed:

$$\left\{\begin{array}{rcll} y_k|\boldsymbol{c},\tau & \sim & \mathcal{N}\left(\left(\boldsymbol{B}\left(t_k\right)\right)^T\boldsymbol{c};\tau^{-1}\right) & \forall k\in\{1,\ldots,n\} \\ p\left(\boldsymbol{c}|\gamma,\theta\right) & \propto & \gamma^{K/2}\exp\left(-\frac{\gamma}{2}PEN\right) & \\ \gamma & \sim & \mathcal{G}\left(a_\gamma;b_\gamma\right) & \\ \tau & \sim & \mathcal{G}\left(a_\tau;b_\tau\right) & \\ \theta & \propto & \mathcal{N}\left(\mu_\theta,p_\theta^{-1}\right), & \end{array}\right.$$

where $PEN=\boldsymbol{c}^T\left(\boldsymbol{P^{11}}+\theta\left(\boldsymbol{P^{10}}+\boldsymbol{P^{01}}\right)+\theta^2\boldsymbol{P^{00}}\right)\boldsymbol{c}=\boldsymbol{c}^T\boldsymbol{S}\left(\theta\right)\boldsymbol{c}$ is the penalty term with $\boldsymbol{P^{ij}}=\int\boldsymbol{B}\left(t\right)\left(\boldsymbol{B}\left(t\right)\right)^T dt$. In this proposal, $\gamma^{K/2}$ is the normalizing constant of the prior distribution for the spline coefficients retained by Campbell. With this model specification, all the conditional posterior distribution are from familiar type:

$$\left\{\begin{array}{rcl} \boldsymbol{c}|\tau,\gamma,\theta,\boldsymbol{y} & \sim & \mathcal{N}\left(\left(\tau\boldsymbol{B}^T\boldsymbol{B}+\gamma\boldsymbol{S}\left(\theta\right)\right)^{-1}\tau\boldsymbol{B}^T\boldsymbol{y};\left(\tau\boldsymbol{B}^T\boldsymbol{B}+\gamma\boldsymbol{S}\left(\theta\right)\right)^{-1}\right) \\ \gamma|\boldsymbol{c},\theta & \sim & \mathcal{G}\left(\frac{K}{2}+a_\gamma;\frac{\boldsymbol{c}^T\boldsymbol{S}\left(\theta\right)\boldsymbol{c}}{2}+b_\gamma\right) \\ \tau|\boldsymbol{c},\boldsymbol{y} & \sim & \mathcal{G}\left(\frac{n}{2}+a_\tau;\frac{\left\|\boldsymbol{y}-\boldsymbol{B}\boldsymbol{c}\right\|^2}{2}+b_\tau\right) \\ \theta|\boldsymbol{c},\gamma,\tau,\boldsymbol{y} & \sim & \mathcal{N}\left(\left(\gamma\boldsymbol{c}^T\boldsymbol{P^{00}}\boldsymbol{c}+p_\theta\right)^{-1}\left(\frac{\gamma}{2}\boldsymbol{c}^T\left(\boldsymbol{P^{10}}+\boldsymbol{P^{01}}\right)\boldsymbol{c}+p_\theta\mu_\theta\right);\left(\gamma\boldsymbol{c}^T\boldsymbol{P^{00}}\boldsymbol{c}+p_\theta\right)^{-1}\right), \end{array}\right.$$

where $\boldsymbol{y}$ is the vector of observations and $\boldsymbol{B}$ the $n\times K$-matrix with kth line equal to $\boldsymbol{B}\left(t_k\right)$. Gibbs sampling can be used to obtain a sample from the joint posterior distribution.
It appears that a part of the normalizing constant has been forgotten in this model specification. In order to adequately specify the model, one has to note that the prior distribution for the spline coefficients, $p\left(\boldsymbol{c}|\gamma,\theta\right)\propto\exp\left(-\frac{\gamma}{2}PEN\right)$, is proportional to the density of a multivariate Gaussian density with mean $\mathbf{0}$ and variance-covariance matrix $\left(\gamma\boldsymbol{S}\left(\theta\right)\right)^{-1}$. The revised model is as followed:

$$\left\{\begin{array}{rcll} y_k|\boldsymbol{c},\tau & \sim & \mathcal{N}\left(\left(\boldsymbol{B}\left(t_k\right)\right)^T\boldsymbol{c};\tau^{-1}\right) & \forall k\in\{1,\ldots,n\} \\ p\left(\boldsymbol{c}|\gamma,\theta\right) & \propto & \gamma^{K/2}\left(\det\left(\boldsymbol{S}\left(\theta\right)\right)\right)^{1/2}\exp\left(-\frac{\gamma}{2}PEN\right) & \\ \gamma & \sim & \mathcal{G}\left(a_\gamma;b_\gamma\right) & \\ \tau & \sim & \mathcal{G}\left(a_\tau;b_\tau\right) & \\ \theta & \propto & \mathcal{N}\left(\mu_\theta,p_\theta^{-1}\right). & \end{array}\right.$$

Now, only the conditional posterior distribution for the spline coefficients, the ODE-adhesion parameter and the precision of measurement are from familiar type:

$$\begin{cases} \boldsymbol{c}|\tau,\gamma,\theta,\boldsymbol{y} & \sim & \mathcal{N}\left(\left(\tau\boldsymbol{B}^T\boldsymbol{B}+\gamma\boldsymbol{S}(\theta)\right)^{-1}\tau\boldsymbol{B}^T\boldsymbol{y};\left(\tau\boldsymbol{B}^T\boldsymbol{B}+\gamma\boldsymbol{S}(\theta)\right)^{-1}\right) \\ \gamma|\boldsymbol{c},\theta & \sim & \mathcal{G}\left(\frac{K}{2}+a_\gamma;\frac{\boldsymbol{c}^T\boldsymbol{S}(\theta)\,\boldsymbol{c}}{2}+b_\gamma\right) \\ \tau|\boldsymbol{c},\boldsymbol{y} & \sim & \mathcal{G}\left(\frac{n}{2}+a_\tau;\frac{\|\boldsymbol{y}-\boldsymbol{B}\boldsymbol{c}\|^2}{2}+b_\tau\right) \\ p(\theta|\boldsymbol{c},\gamma,\tau,\boldsymbol{y}) & \propto & (\det(\boldsymbol{S}(\theta)))^{1/2}\exp\left(-\frac{\gamma}{2}PEN\right)\exp\left(-\frac{p_\theta}{2}(\theta-\mu_\theta)^2\right). \end{cases}$$

Therefore, a Metropolis-within-Gibbs algorithm has to be used in order to obtain a sample from the joint posterior distribution.
Figure II.12 shows the evolution of the determinant of the penalty matrix $\boldsymbol{S}(\theta)$ versus θ (in the log and original scale). It is clear that the conditional posterior distribution is influenced by the complete normalizing constant: one should not consider this determinant as a constant independent of θ.

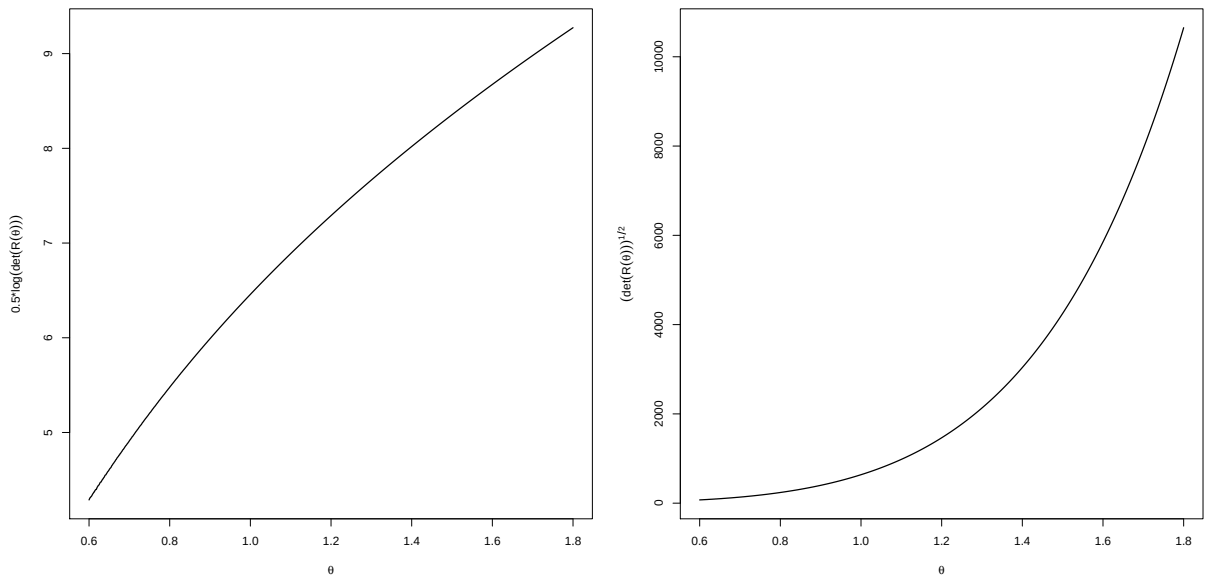


Figure II.12 – Square root of the determinant of the penalty matrix versus the ODE parameter θ (log and original scale).

II.D Markov chains Monte Carlo methods

In Bayesian Analysis, Markov chain Monte Carlo (MCMC) methods are widely used in order to generate sample from the joint posterior distribution $p(\boldsymbol{\psi}|\boldsymbol{y})$ where $\boldsymbol{\psi}$ is the vector of parameters. From a given starting value $\boldsymbol{\psi}^0$, MCMC methods built a sequence $\left\{\boldsymbol{\psi}^t; t = 1, \ldots\right\}$ where the parameter $\boldsymbol{\psi}^t$ at time t is generated only based on the parameter $\boldsymbol{\psi}^{t-1}$ at time $t-1$. After a sufficiently long time (called burn-in period), the distribution of the generated sample coincides with the posterior distribution.
In this appendix, we will focus on the Metropolis algorithm (Metropolis et al., 1953), the Metropolis-Hastings algorithm (Hastings, 1970) and the Gibbs algorithm (Geman and Geman, 1984) that are used in this thesis.

II.D.1 Metropolis algorithm

The Metropolis algorithm is one of the most popular and simple algorithm. From the arbitrary initial value $\boldsymbol{\psi}^0$, the sequence $\left\{\boldsymbol{\psi}^t; t = 1, \ldots\right\}$ is obtained iteratively. Knowing $\boldsymbol{\psi}^{t-1}$ at time $t-1$, $\boldsymbol{\psi}^t$ at time t is obtained by:

1. propose $\boldsymbol{\psi}^{prop}$ using a symmetric distribution $q\left(\cdot|\boldsymbol{\psi}^{t-1}\right)$,

2. calculate the acceptance probability

$$\pi\left(\boldsymbol{\psi}^{prop}, \boldsymbol{\psi}^{t-1}\right) = \min\left(1; \frac{p\left(\boldsymbol{\psi}^{prop}|\boldsymbol{y}\right)}{p\left(\boldsymbol{\psi}^{t-1}|\boldsymbol{y}\right)}\right),$$

3. accept $\boldsymbol{\psi}^{prop}$ with probability $\pi\left(\boldsymbol{\psi}^{prop}, \boldsymbol{\psi}^{t-1}\right)$, i.e. set $\boldsymbol{\psi}^t = \boldsymbol{\psi}^{prop}$ with probability $\pi\left(\boldsymbol{\psi}^{prop}, \boldsymbol{\psi}^{t-1}\right)$ and $\boldsymbol{\psi}^t = \boldsymbol{\psi}^{t-1}$ otherwise.

The most frequent symmetric proposal distributions are the (multivariate) Gaussian distribution and Student distribution. The proposal is computed by: $\boldsymbol{\psi}^{prop} = \boldsymbol{\psi}^{t-1} + \sigma_{prop}U$ where U is distributed according a Gaussian distribution and Student distribution.
An acceptance rate of 40% is targeted when the component in $\boldsymbol{\psi}$ are updated one at time or 20% when all component are updated at the same time. The tuning parameter σ_{prop} can be automatically updated to achieve these level of acceptance (Haario et al., 2001; Lambert and Eilers, 2009).

II.D.2 Metropolis-Hastings algorithm

The Metropolis-Hasting algorithm is a generalization of the Metropolis algorithm where the proposal distribution is no more symmetric. The procedure used to generate a sample from the joint posterior distribution $p(\boldsymbol{\psi}|\boldsymbol{y})$ is similar to the one used in Section II.D.1. Knowing $\boldsymbol{\psi}^{t-1}$ at time $t-1$, $\boldsymbol{\psi}^t$ at time t is obtained by:

1. propose $\boldsymbol{\psi}^{prop}$ using a proposal distribution $q\left(\cdot|\boldsymbol{\psi}^{t-1}\right)$,
2. calculate the acceptance probability

$$\pi\left(\boldsymbol{\psi}^{prop}, \boldsymbol{\psi}^{t-1}\right) = \min\left(1; \frac{p\left(\boldsymbol{\psi}^{prop}|\boldsymbol{y}\right) q\left(\boldsymbol{\psi}^{t-1}|\boldsymbol{\psi}^{prop}\right)}{p\left(\boldsymbol{\psi}^{t-1}|\boldsymbol{y}\right) q\left(\boldsymbol{\psi}^{prop}|\boldsymbol{\psi}^{t-1}\right)}\right),$$

3. set $\boldsymbol{\psi}^t = \boldsymbol{\psi}^{prop}$ with probability $\pi\left(\boldsymbol{\psi}^{prop}, \boldsymbol{\psi}^{t-1}\right)$ and $\boldsymbol{\psi}^t = \boldsymbol{\psi}^{t-1}$ otherwise.

II.D.3 Gibbs algorithm

The Gibbs sampler is used when the conditional posterior distributions can be identified. This algorithm is a special case of the Metropolis algorithm where the acceptance probability is always equal to one.
Let $\boldsymbol{\psi} = \left(\boldsymbol{\psi}_1^T, \ldots, \boldsymbol{\psi}_K^T\right)^T$ be the subdivision of $\boldsymbol{\psi}$ into K subvector (with $\boldsymbol{\psi}_k$ that can be either a scalar or a vector). Each component $\boldsymbol{\psi}_k$ is updated conditionally to the last other updated component. At iteration t, for $k = 1, \ldots, K$, the kth subvector $\boldsymbol{\psi}_k^t$ is sampled from the conditional posterior distribution $p\left(\boldsymbol{\psi}_k^t|\boldsymbol{\psi}_1^t, \ldots, \boldsymbol{\psi}_{k-1}^t, \boldsymbol{\psi}_{k+1}^{t-1}, \ldots, \boldsymbol{\psi}_K^{t-1}\right)$.

Chapter III

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

Abstract

A full Bayesian approach based on ODE-penalized B-spline approach and penalized mixture of Gaussian or Student distributions is proposed to jointly estimate ODE-parameters, state functions and error distributions from the observation of some state functions involved in systems of affine differential equations. Simulations inspired by pharmacokinetic studies show that the proposed method provides comparable results to the method based on the standard ODE-penalized B-spline approach (i.e. with the Gaussian error distribution assumption) when the distribution is Gaussian and outperforms the standard ODE-penalized B-spline approach when the distribution is not Gaussian. The practical utility of a flexible form for the error distribution is illustrated by the analysis of a pharmacokinetic study, while the robustness of the proposed approach is illustrated on a simulated dataset.

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

III.1 Introduction

Chapter II has adapted the ODE-penalized B-spline approach proposed by Ramsay et al. (2007) to a full Bayesian setting when the ODEs are affine and the conditional distribution of the response is assumed Gaussian. The two major drawbacks of the frequentist ODE-

penalized smoothing approach are overcome in the Bayesian framework: the selection of the ODE-adhesion parameters is now automatic and uncertainty measures about parameters can simply be quantified using MCMC. In addition, the Gaussian error density assumption enables to marginalize the joint posterior distribution with respect to the spline coefficients and therefore to get rid of the inconvenient posterior correlation between the spline coefficients and the ODE parameters.

The assumption of a Gaussian distribution for the response is often inappropriate in practice. Lindsey et al. (2000) has shown in the simultaneous modelling of flosequinan concentration and its metabolite concentration that the choice of a parametric error distribution is at least as important as the choice of the pharmacokinetic (PK) compartment model. However, these parametric distributions are restrictive by their definition.

In this chapter, we assume that changes in state functions $\boldsymbol{x}(t) \in \mathbb{R}^d$ are governed by a set of differential equations:

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{f}(\boldsymbol{x}, t, \boldsymbol{\theta}), t \in [0; T] \tag{III.1}$$

where $\boldsymbol{f}$ is a known affine function with respect to $\boldsymbol{x}$ and $\boldsymbol{\theta} \in \mathbb{R}^q$ a vector of unknown parameters. It is assumed that only a subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state functions are measured at time point t_{jk}, $j \in \mathcal{J}$, $k = 1, \ldots, n_j$ with additive measurement error ϵ_{jk}. We denote by

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

the corresponding measurement ($j \in \mathcal{J}$). One further imposes that $\mathbb{E}(\epsilon_j) = 0$ and $\mathbb{V}(\epsilon_j) = 1$ such that $x_j(t)$ and τ_j are respectively the conditional mean and the inverse-variance of y_j.

The goal of this chapter is to generalize the standard Bayesian ODE-penalized B-spline approach by assuming penalized Gaussian mixtures or penalized Student mixtures for the error distributions. This corresponds to jointly estimate the ODE parameters $\boldsymbol{\theta}$, the state functions $\boldsymbol{x}(t)$, the vector of precision $\boldsymbol{\tau} = \text{vec}(\tau_j; j \in \mathcal{J})$ and the standardized error distributions from $\{(t_{jk}, y_{jk}), j \in \mathcal{J}, k = 1, \ldots, n_j\}$.

The plan of the chapter is as follows. Section III.2 describes how this standard Bayesian ODE-penalized B-spline model can be extended to a non Gaussian setting using penalized mixture of Gaussian distributions. Section III.3 proposes to extend the B-ODE-P-spline approach using mixture of Student distributions in order to be at the same time flexible in the specification of the error distribution and robust to outliers. Both extensions only focus on the single-subject modelling but hierarchical modelling can easily adapted (see Section II.4). Section III.4 studies the robustness of the mixture of Student distributions compared to the mixture of Gaussian

distributions on a simulated dataset. The results of a simulation study are also presented in this section in order to asses and to compare the performances of the proposed approach with those of the traditional method. After an application on a pharmacokinetic dataset in Section III.5, we conclude the chapter by a discussion in Section III.6.

III.2 Bayesian ODE-penalized B-spline approach and penalized mixture of Gaussian distributions

III.2.1 Model

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

III.2.1.1 Assumption on the error distribution

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

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

with unknown positive weights π_{jl} summing to one for all $j \in \mathcal{J}$. In order to be able to identify the state functions and the precisions of measurements, we suppose that for all $j \in \mathcal{J}$ and $k = 1, \ldots, n_j$, :

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

and

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

Komárek and Lesaffre (2008) propose to work with $\mu_{j,-L_j} = -5$ and $\mu_{jL_j} = 5$ and place the other μ_{jl} at equidistant positions between -5 and 5 (with $L_j \geq 10$). In practice, we experienced robustness issues with density estimates sensitive to extreme values (see Section III.4.1 for motivation on a simulated dataset). Therefore, we recommend to work with a wider range of value (e.g $\mu_{j,-L_j} = -10$ and $\mu_{jL_j} = 10$ with $L_j \geq 20$). For σ_j^2, the value should be less than

one (see Section III.2.1.4 for details): in practice, one can take $\sigma_j = \frac{2}{3}\delta_j$ where δ_j is the distance between two adjacent means. With this choice, at most five means are within the 0.5% and 99.5% quantile of each Gaussian component.
In traditional approaches in Gaussian mixture models (i.e. when estimating means and variances of the Gaussian components), invariance of the likelihood under relabeling of the mixture components induces label switching (Stephens, 2000). Working with means of the Gaussian components fixed on a grid of equidistant means allows not having to manage this label switching.
Note that the sum to one constraint, $\sum_{l=-L_j}^{L_j} \pi_{jl} = 1$ and $\pi_{jl} > 0$ ($\forall j \in \mathcal{J}$, $\forall l \in \{-L_j, \ldots, L_j\}$) is automatically checked with the generalized logit transformation:

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

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

III.2.1.2 Complete likelihood function

We introduce the group label variable h_{jk} for the kth observation of the jth state function (Diebolt and Robert, 1994; Komárek et al., 2007; Richardson and Green, 1997; Tanner and Wong, 1987). The h_{jk} are supposed independently drawn from the distribution:

$$P(h_{jk} = l|\boldsymbol{\pi_j}) = \pi_{jl}$$

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

$$\epsilon_{jk}|h_{jk}, \boldsymbol{\pi_j} \sim \mathcal{N}\left(\mu_{jh_{jk}}; \sigma_j^2\right).$$

Latent variable h_{jk} corresponds to the label of the group from which each random error variable ϵ_{jk} is drawn. Working with this data augmentation allows to come back to the Gaussian case considered in Chapter II. The complete likelihood function is therefore equal to

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

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

III.2.1.3 Prior for the transformed mixture weights

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

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

for all $j \in \mathcal{J}$, where Δ^s denotes the difference operator of order s. We suggest, as in Komárek and Lesaffre (2008), to set s equal to 3. The vector of roughness penalty parameters $\boldsymbol{\lambda} = \text{vec}\left(\lambda_j; j \in \mathcal{J}\right)$ controls the smoothness of each density and has to be selected. Small values indicate a large flexibility in the specification of the error distribution and large values for the roughness penalty parameter just forces the error distribution to be the standard Gaussian distribution. Note also that the effective number of parameters in the estimation of the error distribution (for a given state function) is smaller than $2L$ due to the finite third-order difference penalization on the transformed mixture weights. In Bayesian terms, these penalties can be translated into a prior distribution for each $\boldsymbol{a_j}$:

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

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

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

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

III.2.1.4 Identification penalty

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

$$\epsilon_{jk}|\boldsymbol{d_j} \sim \sum_{l=-L_j}^{L_j} \pi_{jl}\left(\boldsymbol{d_j}\right) \mathcal{N}\left(\mu_{jl} + \alpha_j; \frac{\sigma_j^2}{\beta_j^2}\right) \tag{III.3}$$

where

$$\alpha_j = -\sum_{l=-L_j}^{L_j} \pi_{jl}\mu_{jl}$$

and

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

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

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

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

III.2.1.5 Other prior distributions

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

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

The vector $\boldsymbol{\mu_c}$ and the 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. The prior density corresponds to a multivariate normal distribution with mean vector $\boldsymbol{V_1}^{-1}\boldsymbol{v_1}$ and covariance matrix $\boldsymbol{V_1}^{-1}$ where $\boldsymbol{v_1} = -\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right) + \boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}$ and $\boldsymbol{V_1} = \boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma}\right) + \boldsymbol{\Sigma_c}^{-1}$. Therefore, the normalizing constant for the conditional prior density of the spline coefficients $\boldsymbol{c}$ is:

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

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

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

where $\mathcal{G}(a; b)$ denotes the gamma distribution with mean a/b and variance a/b^2. As explained in Chapter II, for the ODE-adhesion parameters, we will opt for a shape $a_{\gamma_j} = 1$ and a rate b_{γ_j} smaller than 10^{-6}. This choice for the prior distributions of the ODE-adhesion parameters translates our prior confidence in the specification of the system of differential equations as descriptor of the dynamics of the state function. For the roughness penalty parameters $\boldsymbol{\lambda}$, we suggested to consider a gamma prior distribution with shape $a_{\lambda_j} = 1$ and a rate b_{λ_j} smaller than 10^{-6}. This prior distribution indicates our slight prior preference for the Gaussian distribution for the error distribution in the absence of counter-indications in the data.
For the vector $\boldsymbol{\theta}$ of differential equation parameters, the chosen prior will depends on the context. Let us denote this prior distribution by $p(\boldsymbol{\theta})$.

III.2.2 Posterior distribution and parameter estimation

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

III.2.2.1 Joint posterior distribution

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

$$\begin{aligned}
& \log\left(p\left(\boldsymbol{c}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{d}, \boldsymbol{\lambda}, \boldsymbol{\theta} | \boldsymbol{y}, \boldsymbol{h}\right)\right) \\
= & \sum_{j \in \mathcal{J}} \left\{ \frac{n_j}{2} \log\left(\tau_j\right) + \sum_{l=-L_j}^{L_j} r_{jl} \log\left(\pi_{jl}\right) - \frac{\beta_j^2}{2\sigma_j^2} \left(\tau_j \sum_{k=1}^{n_j} y_{jk}^2 - 2\tau_j^{1/2} \sum_{k=l}^{n_j} \left(\mu_{jh_{jk}} + \alpha_j\right) y_{jk} \right) \right\} \\
& - \frac{1}{2} \boldsymbol{c}^T \operatorname{diag}\left(\boldsymbol{Z_j}; j = 1, \ldots, d\right) \boldsymbol{c} + \boldsymbol{c}^T \operatorname{vec}\left(\boldsymbol{z_j}; j = 1, \ldots, d\right) \\
& + \sum_{j \in \mathcal{J}} \left\{ \frac{2L_j + 1 - s}{2} \log\left(\lambda_j\right) - \frac{\lambda_j}{2} \boldsymbol{d_j}^T \boldsymbol{P_{d,j}} \boldsymbol{d_j} \right\} \\
& - \sum_{j \in \mathcal{J}} \kappa \left(\alpha_j^2 + \left(\beta_j^2 - 1\right)^2 \right)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j\in\mathcal{J}}\left\{\left(a_{\lambda_j}-1\right)\log\left(\lambda_j\right)-b_{\lambda_j}\lambda_j\right\}\\
&+\sum_{j\in\mathcal{J}}\left\{\left(a_{\tau_j}-1\right)\log\left(\tau_j\right)-b_{\tau_j}\tau_j\right\}\\
&-\frac{1}{2}\left\{\boldsymbol{c}^T\boldsymbol{V_1}\boldsymbol{c}-2\boldsymbol{c}^T\boldsymbol{v_1}+\boldsymbol{v_1}^T\boldsymbol{V_1}^{-1}\boldsymbol{v_1}-\log\left(\det\left(\boldsymbol{V_1}\right)\right)\right\}\\
&+\sum_{j=1}^{d}\left\{\left(a_{\gamma_j}-1\right)\log\left(\gamma_j\right)-b_{\gamma_j}\gamma_j\right\}\\
&+\log\left(p\left(\boldsymbol{\theta}\right)\right)
\end{aligned}
$$

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

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

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

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

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

III.2.2.2 Conditional posterior distributions

The conditional posterior distribution for each roughness penalty parameter λ_j corresponds to a gamma distribution:

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

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

$$\boldsymbol{v_2}=\boldsymbol{v_1}+\text{vec}\left(\boldsymbol{z_j}; j=1,\ldots,d\right)$$

and

$$\boldsymbol{V_2}=\boldsymbol{V_1}+\text{diag}\left(\boldsymbol{Z_j}; j=1,\ldots,d\right).$$

The other conditional posteriors are unfortunately not of a familiar type. Therefore, the Metropolis-within-Gibbs algorithm can be used to sample from the joint posterior. As in Chapter II, the joint posterior distribution can be marginalized with respect to the spline coefficients in order to avoid being forced to deal with the strong posterior conditional correlation between the spline coefficients and the differential equation parameters. This is feasible as we are working with mixture of Gaussian distributions. The joint posterior distribution can also be marginalized with respect to $\boldsymbol{\lambda}$. Finally, the conditional posterior probability for the group label variable is proportional to:

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

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

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

III.2.2.3 MCMC algorithm

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

III.2.2.3.1 MCMC algorithm We use a Metropolis-within-Gibbs algorithm. Each component of the transformed weights $\boldsymbol{d_j}$, $j\in\mathcal{J}$ is updated using univariate Metropolis steps with a Gaussian proposal distribution. Gibbs steps are used to update the group label variables $\boldsymbol{h}$, the penalty parameters $\boldsymbol{\lambda}$ and the spline coefficients $\boldsymbol{c}$. Last of all, univariate Metropolis steps are used for all the component of the precision parameters $\boldsymbol{\tau}$, the ODE-adhesion parameters $\boldsymbol{\gamma}$ and the ODE parameters $\boldsymbol{\theta}$. The standard deviation of each Gaussian proposal distribution is tuned during the burn-in phase to ensure a 40% acceptance rate (Haario et al., 2001; Lambert and Eilers, 2009).

III.2.2.3.2 Starting value The starting value for the ODE parameters, the spline coefficients and the precision parameters are obtained using a standard approach (with a Gaussian distribution for the error term distribution). All components of the transformed weights $\boldsymbol{d_j}$, $j \in \mathcal{J}$ can either be set to zero (inefficient starting values) or set to the transformed mixture weights corresponding to a standard Gaussian density.

III.2.2.3.3 Re-parameterization using rotation and translation To accelerate the inference procedure, we also used the strategy proposed by Lambert (2007).
First run the Metropolis-within-Gibbs algorithm for a few thousand iterations. Then, reparameterize the problem by applying a translation and a rotation to the parameter vector $\boldsymbol{\theta}$ and for all vector $\boldsymbol{d_j}$, $j \in \mathcal{J}$.
More precisely, denote by $\boldsymbol{S_\theta}$ (resp. $\boldsymbol{S_{d_j}}$) the empirical variance-covariance matrix of the parameter $\boldsymbol{\theta}$ (resp. $\boldsymbol{d_j}$) evaluated using a first set of iterations of the generated chains and $\bar{\boldsymbol{\theta}}$ (resp. $\bar{\boldsymbol{d_j}}$) the corresponding empirical mean vector. We reparameterize the posterior distribution using $\boldsymbol{\beta_\theta}$ (resp. $\boldsymbol{\beta_{d_j}}$),

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

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

III.3 Bayesian ODE-penalized B-spline approach and penalized mixture of Student distributions

We will now consider that the distribution of error terms is defined by a penalized mixture of Student distributions. With that specification, the error distribution will be at the same time flexible and robust. Note that the Student distribution case is considered in Appendix III.A.

III.3.1 Model

In this section, we assume that the standardized error distribution is given by a mixture of Student distributions. Similarly to Section III.2.1, data augmentation will be introduced in the modelling to reduce the mixture of Student distributions to a conditional Gaussian case.

III.3.1.1 Assumption on the error distribution

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

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

where the dynamic model of the state function is given by Eq. (III.1). The distribution of the error terms associated to the jth state function is modeled by

$$\epsilon_{jk}|\boldsymbol{\pi_j} \sim \sum_{l=-L_j}^{L_j} \pi_{jl} t\left(\nu_j, \mu_{jl}, \sigma_j^2\right). \tag{III.5}$$

If $T_j \sim t\left(\nu_j, \mu_{jl}, \sigma_j^2\right)$ then $T_j = \sigma_j U_j + \mu_{jl}$ where U_j is distributed according a standardized Student distribution with ν_j degrees of freedom. For small value of the degree of freedom ν_j, these standardized Student distributions are more picked than the standard Gaussian distribution but have heavier tails. When the degree of freedom is large (e.g. ν_j larger than 30), this mixture of Student distributions is similar to the mixture of Gaussian distributions presented in Section III.2.1.1. Note that the means are again selected on fixed pre-defined grid $\boldsymbol{\mu_j} = \text{vec}\left(\mu_{jl}; l = -L_j, \ldots, L_j\right)$ centered around $\mu_{j0} = 0$.

III.3.1.2 Data augmentation

In order to manage this mixture of Student distributions, we introduce two latent variables: the group label variable h_{jk} and the latent variable ω_{jk} for the kth observation of the jth state function. The group label variables are supposed independently drawn from distribution:

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

and latent variables ω_{jk} are supposed drawn from distribution:

$$\omega_{jk} \sim \mathcal{G}\left(\frac{\nu_j}{2}; \frac{\nu_j - 2}{2}\right).$$

Given these latent variables h_{jk} and ω_{jk}, one has:

$$\epsilon_{jk}|h_{jk},\omega_{jk} \sim \mathcal{N}\left(\mu_{jh_{jk}}; \frac{\sigma_j^2}{\omega_{jk}}\right).$$

Latent variable h_{jk} corresponds to the label of the group from which each random error variable ϵ_{jk} is drawn. Latent variable ω_{jk} allows to view the Student distribution as a conditional Gaussian distribution.

III.3.1.3 Bayesian model

The Bayesian model is a adaptation of the model used in Section III.2.1:

$$\begin{cases} y_{jk} &= \widetilde{x_j}\left(t_{jk}\right) + \tau_j^{1/2}\epsilon_{jk} \\ \epsilon_{jk}|h_{jk},\omega_{jk} &\sim \mathcal{N}\left(\mu_{jh_{jk}} + \alpha_j; \frac{\sigma_j^2}{\beta_j^2\omega_{jk}}\right) \\ P\left(h_{jk}=l|\boldsymbol{\pi_j}\right) &= \pi_{jl} \\ \omega_{jk} &\sim \mathcal{G}\left(\frac{\nu_j}{2}; \frac{\nu_j-2}{2}\right) \\ p\left(\boldsymbol{c}|\boldsymbol{\theta},\boldsymbol{\gamma}\right) &\propto \left(\det\left(\boldsymbol{V_1}\right)\right)^{1/2}\exp\left(-\frac{1}{2}\left\{\boldsymbol{c}^T\boldsymbol{V_1}\boldsymbol{c} - 2\boldsymbol{c}^T\boldsymbol{v_1} + \boldsymbol{v_1}^T\boldsymbol{V_1}^{-1}\boldsymbol{v_1}\right\}\right) \\ \gamma_j &\sim \mathcal{G}\left(a_{\gamma_j}; b_{\gamma_j}\right) \\ \tau_j &\sim \mathcal{G}\left(a_{\tau_j}; b_{\tau_j}\right) \\ p\left(\boldsymbol{d_j}|\lambda_j\right) &\propto \lambda_j^{L_j-1}\exp\left(-\frac{\lambda_j}{2}\boldsymbol{d_j}^T\boldsymbol{P_{d,j}}\boldsymbol{d_j}\right) \\ \lambda_j &\sim \mathcal{G}\left(a_{\lambda_j}; b_{\lambda_j}\right) \\ \boldsymbol{\theta} &\sim p\left(\boldsymbol{\theta}\right). \end{cases}$$

where the vector $\boldsymbol{v_1}$ and the matrix $\boldsymbol{V_1}$ is defined in Section III.2.1.5. The vector $\boldsymbol{d_j}$, $j \in \mathcal{J}$ is the vector of transformed mixture weights and the matrix $\boldsymbol{P_{d,j}}$ is the finite third order difference penalty matrix (defined in Section III.2.1.3). Similarly to Section III.2.1.4, we introduce correction term α_j to the mean and correction term β_j to the variance of the mixture components to ensure the zero mean and unit variance for the error term.

III.3.2 Posterior distribution and parameter estimation

In this section, we give the conditional posterior distribution and we explain how to to generate sample from the joint posterior distribution using MCMC.

III.3.2.1 Posterior distribution

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

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

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

$$\boldsymbol{v_2} = \boldsymbol{v_1} + \text{vec}\left(\boldsymbol{z_j}; j = 1, \ldots, d\right)$$

and

$$\boldsymbol{V_2} = \boldsymbol{V_1} + \text{diag}\left(\boldsymbol{Z_j}; j = 1, \ldots, d\right),$$

where $\text{diag}\left(\boldsymbol{Z_j}; j = 1, \ldots, d\right)$ corresponds to a block-diagonal matrix where $\boldsymbol{Z_j}$ is equal to

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

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

$$\tau_j \frac{\beta_j^2}{\sigma_j^2} \boldsymbol{B_j}^T \boldsymbol{\Omega_j} \boldsymbol{y_j} - \tau_j^{1/2} \frac{\beta_j^2}{\sigma_j^2} \boldsymbol{B_j}^T \boldsymbol{\Omega_j} \text{vec}\left(\mu_{jh_{jk}} + \alpha_j; k = 1, \ldots, n_j\right)$$

if the jth state function is observed and the null K_j-vector otherwise. The matrix $\boldsymbol{\Omega_j}$ is the diagonal matrix with ω_{jk} on the (k,k)th entry ($j \in \mathcal{J}$, $k = 1, \ldots, n_j$).
The conditional posterior distribution for the latent variable ω_{jk} is a gamma distribution:

$$\omega_{jk}|y_{jk}, h_{jk}, \boldsymbol{c_j}, \tau_j \sim \mathcal{G}\left(\frac{\nu_j + 1}{2}; \frac{\beta_j^2 \left(\tau_j^{1/2}\left(y_{jk} - \boldsymbol{B_j}\left(t_{jk}\right)\boldsymbol{c_j}\right) - \left(\mu_{jh_{jk}} + \alpha_j\right)\right)^2}{2\sigma_j^2} + \frac{\nu_j - 2}{2}\right).$$

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

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

Finally, the joint posterior distribution is marginalized with respect to the spline coefficients. The log of it can be shown to be:

$$
\begin{aligned}
& \log\left(p\left(\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{d}, |\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{h}\right)\right) \\
\doteq\ & \sum_{j \in \mathcal{J}} \left\{ \frac{n_j}{2} \log\left(\tau_j\right) - \frac{\beta_j^2}{2\sigma_j^2} \left(\tau_j \boldsymbol{y_j}^T \boldsymbol{\Omega_j} \boldsymbol{y_j} - 2\tau_j^{1/2} \boldsymbol{y_j}^T \boldsymbol{\Omega_j} \text{vec}\left(\mu_{jh_{jk}} + \alpha_j; j = 1, \ldots, n_j\right)\right)\right\} \\
& + \sum_{l=-L_j}^{L_j} r_{jl} \log\left(\pi_{jl}\right) \\
& - \frac{1}{2}\left\{ \boldsymbol{v_1}^T \boldsymbol{V_1}^{-1} \boldsymbol{v_1} - \log\left(\det\left(\boldsymbol{V_1}\right)\right) - \boldsymbol{v_2}^T \boldsymbol{V_2}^{-1} \boldsymbol{v_2} + \log\left(\det\left(\boldsymbol{V_2}\right)\right)\right\} \\
& + \sum_{j \in \mathcal{J}} \left\{ \left(L_j - 1\right) \log\left(\lambda_j\right) - \frac{\lambda_j}{2} \boldsymbol{d_j}^T \boldsymbol{P_{d,j}} \boldsymbol{d_j} \right\} \\
& - \sum_{j \in \mathcal{J}} \kappa \left(\alpha_j^2 + \left(\beta_j^2 - 1\right)^2 \right) \\
& + \sum_{j \in \mathcal{J}} \left\{ \left(a_{\lambda_j} - 1\right) \log\left(\lambda_j\right) - b_{\lambda_j} \lambda_j \right\} \\
& + \sum_{j \in \mathcal{J}} \left\{ \left(a_{\tau_j} - 1\right) \log\left(\tau_j\right) - b_{\tau_j} \tau_j \right\} \\
& + \sum_{j=1}^{d} \left\{ \left(a_{\gamma_j} - 1\right) \log\left(\gamma_j\right) - b_{\gamma_j} \gamma_j \right\} \\
& + \log\left(p\left(\boldsymbol{\theta}\right)\right).
\end{aligned}
$$

The conditional posterior distribution cannot be identified. Therefore, Metropolis-Hastings steps are used to generate sample from $p\left(\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{d}, |\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{h}\right)$. Again, the marginalization of the joint posterior distribution with respect to the spline coefficients was feasible due to the use of the latent variables $\boldsymbol{\omega}$ and $\boldsymbol{h}$.

III.3.2.2 MCMC algorithm

We use a similar strategy to the one presented in Section III.2.2.3. Each component of the transformed weights $\boldsymbol{d_j}$, $j \in \mathcal{J}$ is updated using univariate Metropolis steps with a Gaussian proposal distribution. Gibbs steps are used to update the group label variables $\boldsymbol{h}$, the latent variables $\boldsymbol{\omega}$, the penalty parameters $\boldsymbol{\lambda}$ and the spline coefficients $\boldsymbol{c}$. Last of all, univariate Metropolis steps are used for all the component of the precision parameters $\boldsymbol{\tau}$, the ODE-adhesion parameters $\boldsymbol{\gamma}$ and the ODE parameters $\boldsymbol{\theta}$. The standard deviation of each Gaussian proposal distribution is tuned during the burn-in phase to ensure a 40% acceptance rate.

III.4 Simulation

In this section, we first point the advantages and the limitations of the penalized mixture of Student distributions on a simulated dataset. Then, we perform a simulation study in order to validate and to compare the performances of the proposed Bayesian ODE-penalized B-spline approach (assuming a penalized Gaussian mixture as error distribution) with the existing one (assuming a Gaussian distribution as error distribution) in the estimation of the ODE parameters and of the error precision.

III.4.1 Robustness of the mixture of Student distributions

The aim of this section is to show the influence of the chosen sequence of means in the mixture of Gaussian distributions and the robustness of the mixture of Student distributions.

III.4.1.1 Data and model specification

We simulate 50 measurements at equidistant time points between 0 and 1. The observations are given by:

$$y_k = x\left(t_k\right) + \tau^{-1/2}\epsilon_k,$$

where the state function is:

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

with $D = 1mg$, $k_e = 5h^{-1}$ and $V = 1L$. The precision of measurement is set to 100. For the error terms $\{\epsilon_k; k = 1, \ldots, 50\}$, we generate realization of a standard Gaussian distribution. The generated error terms with an absolute value larger than 2 have been arbitrarily multiplied by 2 to simulate outliers. The simulated data are shown in Figure III.1.
For the analysis of this simulated dataset, the constant of elimination k_e, the precision of measurement τ, the spline coefficients and the error distribution are estimated. The state function is approximated by $\widetilde{x}\left(t\right) = \left(\boldsymbol{B}\left(t\right)\right)^T \boldsymbol{c}$ using a 5-order B-spline expansion with inner knots at each tenth between 0 and 1. The error density is modeled either by a mixture of Gaussian distributions (Eq. (III.2)) or by a mixture of Student distributions (Eq. (III.5)). For each distributional assumption, 41 components are considered with means selected on an equidistant grid between -4 and 4, centered around 0. Note that four of the generated error terms (corresponding to the four observations represented by a star in Figure III.1) are not within the bounds of this grid of means.

III.4.1.2 Results

Figure III.1 shows the fitted error density using the mixture of Gaussian distributions (bottom left) and the fitted error density using the mixture of Student distributions (bottom right). The fitted error density in the Gaussian mixture case is trimodal: one major mode centered on zero and two minor modes on -4 and 4. These minor modes are due to the grid of means that is not wide enough to capture the simulated outliers and also to the fast decay in the tails of the Gaussian components in the mixture. At the opposite, the fitted error density using the mixture of Student distributions is unimodal, almost centered on 0 with the target Gaussian distribution within its pointwise 95% credibility interval. The influence of the extreme values is reduced thanks to the heavy tails of the Student components in the mixture.
For the analysis of this dataset, 231 seconds were necessary in the Gaussian mixture case and 285 seconds in the Student one (+20%, average time for 10 runs on an Intel Core i7-860 Processor (2.80GHz)). But the computational time increases faster with sample size with a mixture of Student distributions: in the same configuration but with 200 observations, 549 seconds are needed in the Gaussian mixture case whereas 963 seconds (+75%) are necessary in the Student mixture case. This can be explained by the additional computational steps (matrix products, second latent variable) in the Student mixture case.

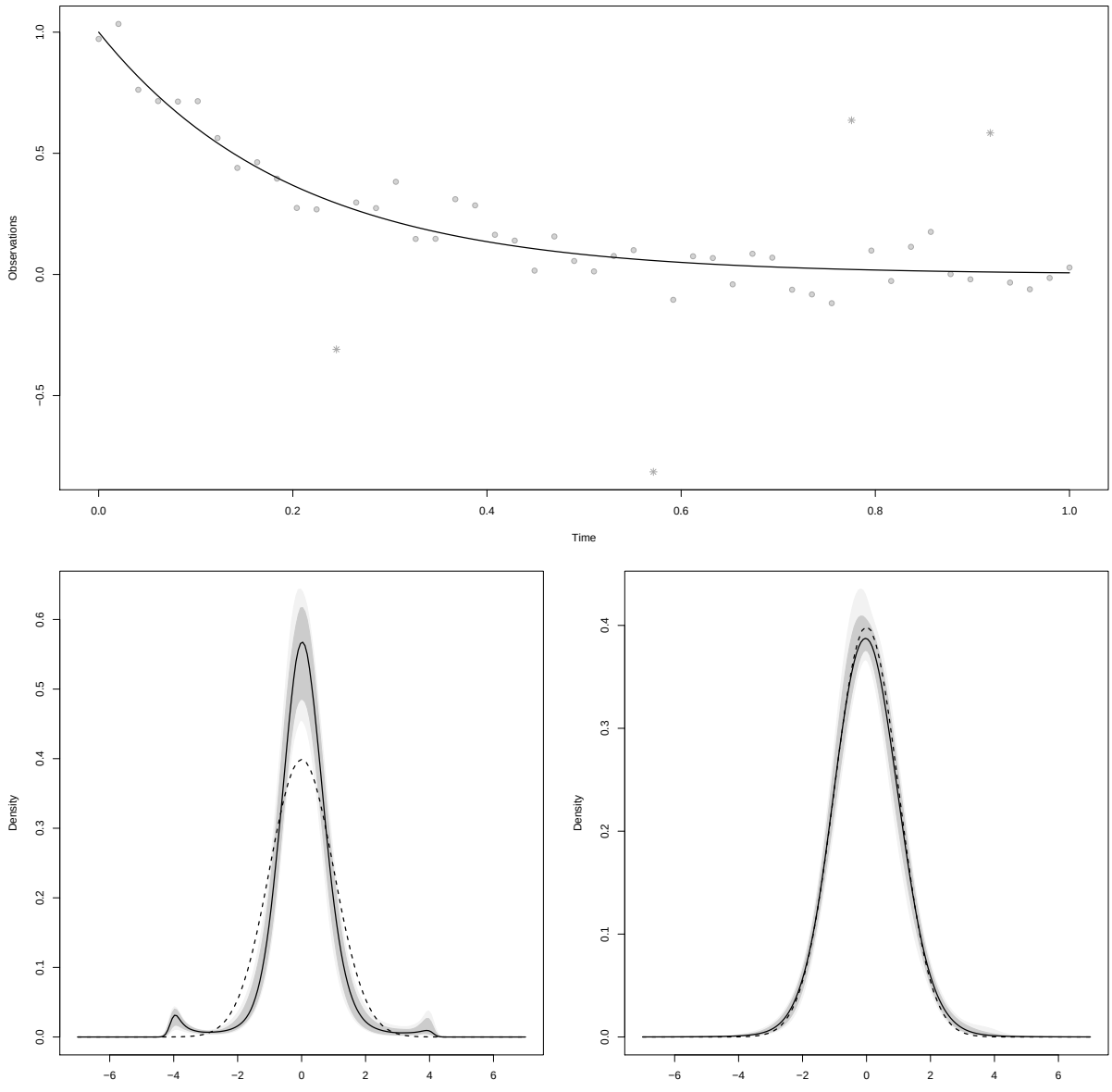


Figure III.1 – Top: Simulated data from an exponential decay model (stars correspond to outliers). Bottom left : fitted error density using the mixture of Gaussian distributions. Bottom right: fitted error density using the mixture of Student distribution. On bottom figures, solid curves are pointwise posterior median for the fitted error density, grey regions delimit pointwise 80% and 95% credibility intervals and the dashed curve corresponds to the target standard normal distribution.

III.4.2 Simulation study

In this simulation study, we compare the performances of the Bayesian ODE-penalized B-spline approach combined with a penalized mixture of Gaussian distributions and the standard Bayesian ODE-penalized B-spline approach in the estimation of the ODE parameters and of the error precision.

III.4.2.1 Scenario

Let us consider the oral 2-compartment model presented in Section I.2.2.3. For this model, we are interested by the estimation of (k_a, k_e, V, τ) from a set of observations $\{(t_k, y_k) : k = 1, \ldots, n\}$ where $y_k = x_2(t_k) + \tau^{-1/2}\epsilon_k$. The sampling times correspond to n equidistant time points between 0 and 2 with four sample sizes: $n = 50, 100, 200, 500$. The error terms were generated from four distributions: a standard Gaussian, a Student t_6, a standardized extreme value and a Gaussian mixture $0.4\mathcal{N}(-1.5; 0.9^2) + 0.6\mathcal{N}(1; 0.6^2)$. Each distribution is standardized to have zero mean and unit variance. For this simulation study, only the state function $x_2(t)$ is observed. For each configuration, 100 datasets were generated with (k_a, k_e, V, τ) equal to $(8, 2, 1, 5)$. For the analysis, we impose k_a to be larger than k_e.

III.4.2.2 Results

Tables III.1, III.2, III.3 and III.4 summarize the results of the simulation study. It gives estimation based on MCMC for the relative bias (in percent, based on the estimated posterior mean), the coverage probabilities at level 80% and 95% (based on 80% and 95%-HPD intervals), the root mean squared error and the mean posterior standard deviation for parameter k_a, k_e, V and τ for the Bayesian ODE-penalized B-spline approach with penalized Gaussian mixtures as error distributions (PGM) and Bayesian ODE-penalized B-spline approach with Gaussian error distribution (G). Figure III.2 shows, for each simulation, the estimated density function for each of the 100 generated datasets (in light grey solid curve) together with the density function used for the data generation (in black solid curve).
When the generated error is Gaussian, the relative bias, estimated coverage probabilities, root mean squared error and mean posterior standard deviation for k_a, k_e, V and τ for G and PGM approaches are similar, meaning that the unnecessary flexibility in the error distribution with PGM is not an handicap (see the first column of Figure III.2 for the estimated error density).
When the generated error term has a Student distribution, the relative bias is small for k_a, k_e and V. All the estimated coverage probabilities are in agreement with their nominal values (see Section II.5.2.1 for details). The RMSE for the ODE parameters is smaller in the PGM

approach than in the G approach when the sample size is larger than 100. For the precision of measurement τ, the relative bias is positive and quite large when $n = 50$, but decreases with sample size. The estimated coverage probabilities for the precision parameter are not in agreement with their nominal value when $n \leq 100$ (except CP80 for the PGM approach). For n larger than 200, only the coverage probabilities for the PGM approach are in agreement with their nominal values. The RMSE for τ is similar in the G and PGM approaches for each sample size. The mean posterior standard deviation of τ is larger in the PGM approach: this could partly explain why the estimated coverage probabilities for that parameter are larger than those estimated in the G approach. Note that the estimated error densities are getting closer to the target error density when the sample size increases.

When the generated error term has an extreme value distribution, the relative bias is small for all the ODE parameters. Their coverage probabilities are almost all in agreement with their nominal value in the PGM approach (except for sample size $n = 50$). The RMSE and the MPSD are smaller in the PGM approach for the ODE parameters and decrease as expected with sample size. The bias of the precision of measurement τ is slightly positive, especially for the PGM approach but decrease with the sample size. Estimated coverage probabilities at level 80% and 95% are larger in the PGM approach than in the G approach. But they are not in agreement with their nominal value, except for the PGM approach when the sample size is equal to 200 (CP95) and when the sample size is equal to 500 (CP80 and CP95). The estimated error densities are again getting closer to the target error density when increasing the sample size (see the third column of Figure III.2).

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

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

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

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

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

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

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

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

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

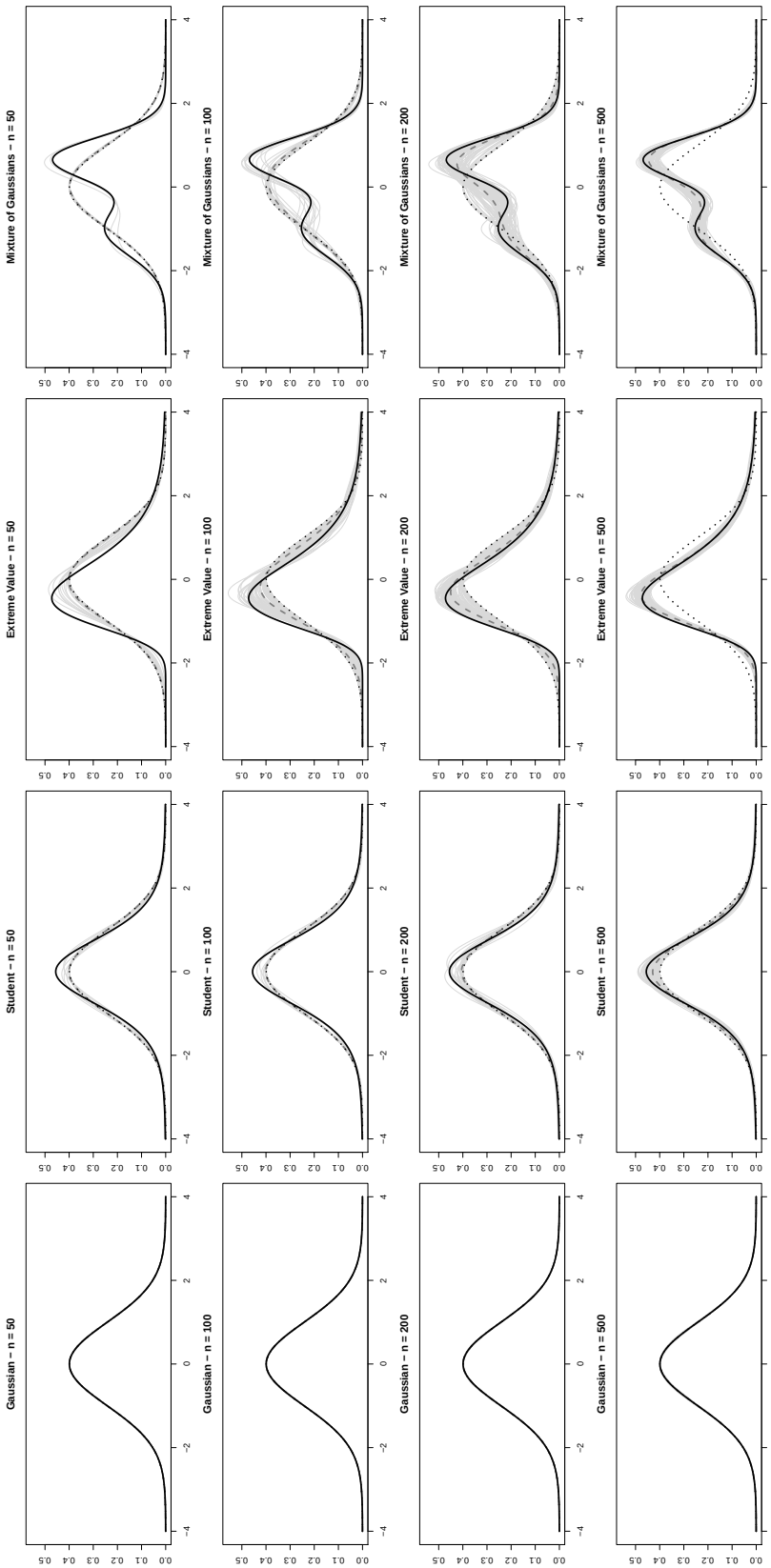


Figure III.2 – Estimated error density functions for sample size (50, 100, 200, 500) (one light grey curve per dataset) with pointwise average of these estimated densities in dark grey dashed curve. The black solid curve is the true density function and the black dotted curve is the standardized Gaussian density.

III.5 Application

We propose one application of the B-ODE-P-spline approach combined with penalized mixture distributions for the modelling of the pharmacokinetics after an oral administration of the drug Theophylline.

III.5.1 Data, objective and model specification

The Theophylline dataset was presented in Section I.2.2.3. The evolution of the concentration $x_2(t)$ of theophylline in blood on 12 subjects is presented on Figure I.5. Using the whole set of observations, the objective is to determine the evolution of the drug in the stomach and in the blood. A common model for the kinetic of theophylline the two compartment model with first order absorption and elimination presented in Section I.2.2.3.

The modelling strategy described in Section III.2.1 was used. For the ith subject, we approximate the amount of drug in the stomach by $x_{1i}(t) = (\boldsymbol{B_1}(t))^T \boldsymbol{c_{1i}}$ and the concentration of the drug in the central compartment by $x_{2i}(t) = (\boldsymbol{B_2}(t))^T \boldsymbol{c_{2i}}$. We parameterize the constant of absorption k_a, the constant of elimination k_e and the volume V of the central compartment in the log-scale with Gaussian random effects at the subject level. The prior information available for the study of this dataset concerns the PK parameters and the initial condition of the state functions. For the PK parameters, one has to force the constant of absorption to be larger than the constant of elimination. We have also a total confidence in the initial condition of the state functions: the initial amount of drug in the stomach is equal to the oral dose and the initial concentration of the drug in the blood is null.

MCMC chains of length 10000 (including a 1000 burn-in) were generated using Metropolis-within-Gibbs algorithm described in Section III.2.2. We use this sample to reparameterize the problem by applying a translation and a rotation to the parameter vector $\theta = (k_a, k_e, V)^T$ and for all the transformed mixture weights $\boldsymbol{d_j}$. A final chain of length 25000 (including an extra burn-in of 5000) is generated for the inference. Convergence for each variable was confirmed by the Geweke's convergence diagnostic.

III.5.2 Results

Figure III.3 shows (in solid black line) the pointwise posterior median of the error density function with the 80% and 95% pointwise credibility intervals (in gray). It suggests that the error density is symmetric and highly peacked. Note that the standard Gaussian distribution (in dotted line) is not within the pointwise credibility intervals, clearly suggesting that the

fitted error density is not Gaussian.

Figure III.4 shows the estimated individual amount of drug in the stomach with pointwise 80% and 95% credibility sets. Figure III.5 shows the fitted individual concentration of Theophylline in blood with pointwise 80% and 95% credibility sets. These figures are obtained by estimating at each MCMC generation of the spline coefficients, the amount of drug in the stomach and the concentration of drug in the central compartment over a detailed grid of times. Then the 2.5%, 10%, 50%, 90% and 97.5% quantiles of the MCMC chain of the estimated curves are taken at each time point. Figure III.6 shows the pointwise credibility interval for the posterior predictive distribution of the concentration of theophylline in blood. It corresponds to a 80% and 95% credibility set of the predictive distribution of the concentration of the theophylline drug. Six observed concentration (subject 2, 5, 8, 9 and 9) are outside the 95% posterior credibility interval which is compatible with the expected coverage.

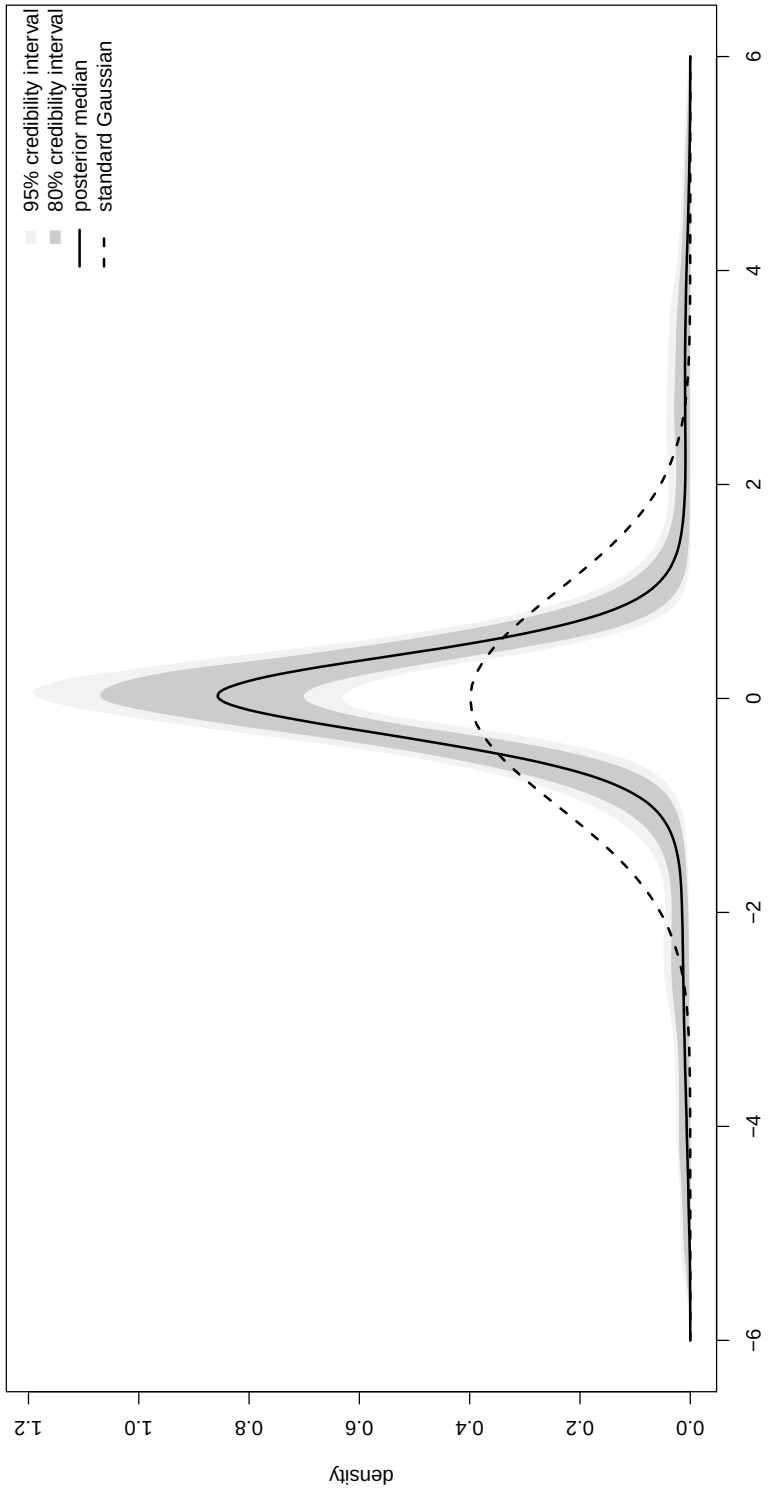


Figure III.3 – Fitted error density (solid curve is the pointwise posterior median and grey regions delimit pointwise 80% and 95% credible intervals), dotted curve is the Gaussian density.

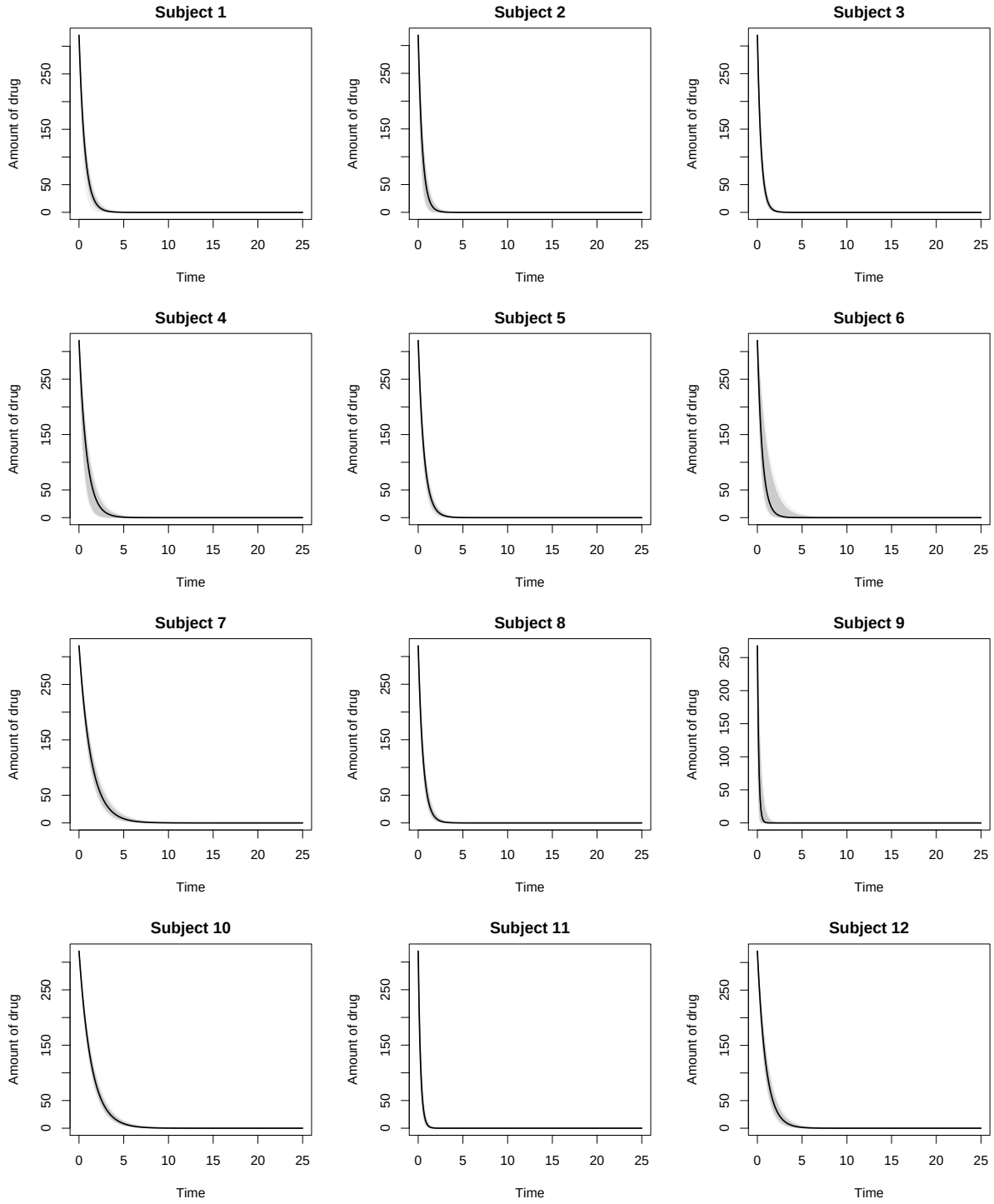


Figure III.4 – Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted amount of Theophylline in stomach with pointwise median (black solid curve) versus time for the twelve subjects.

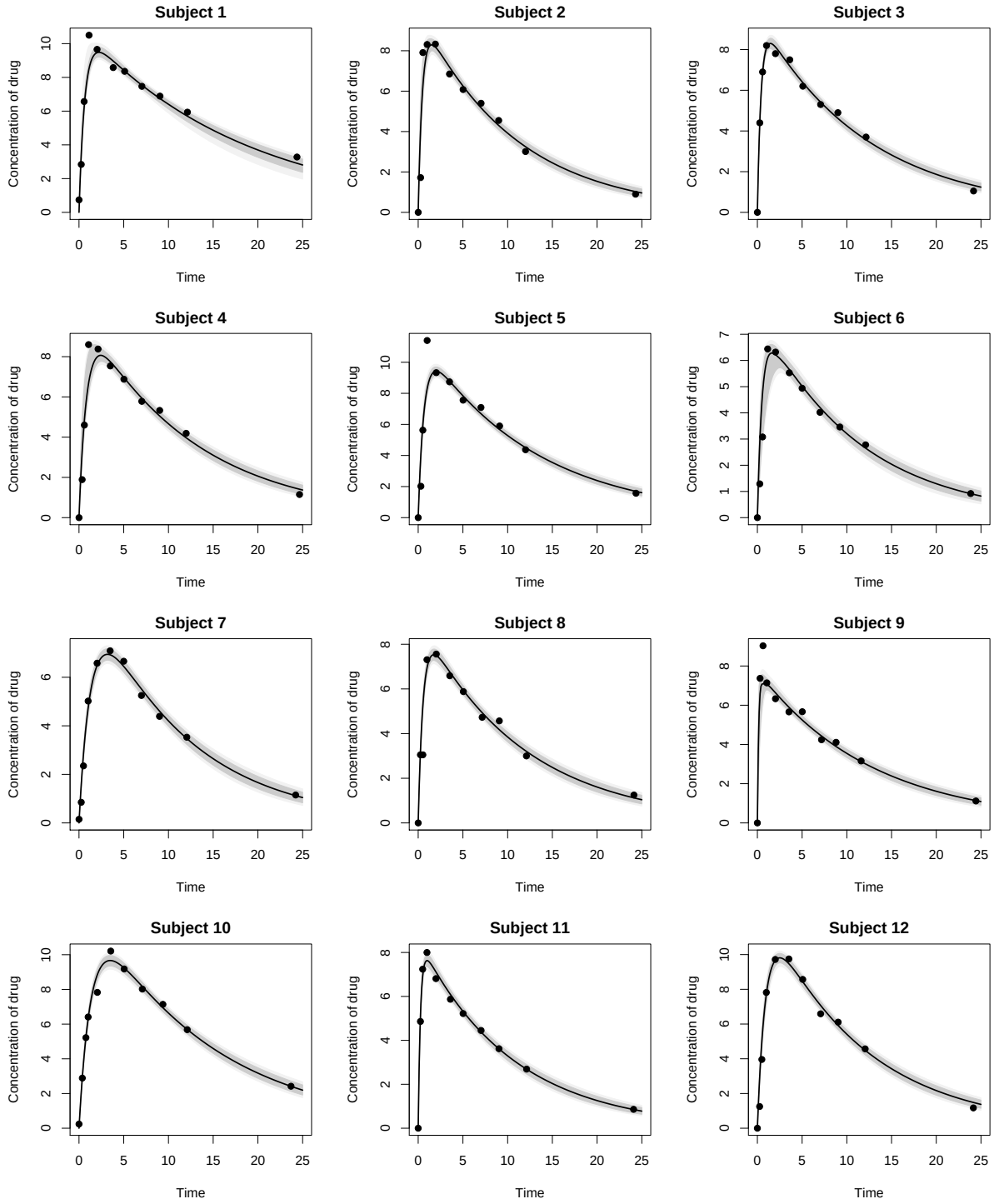


Figure III.5 – Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted concentration of Theophylline with pointwise median (black solid curve) and observed concentration of Theophylline (solid circles) versus time for the twelve subjects.

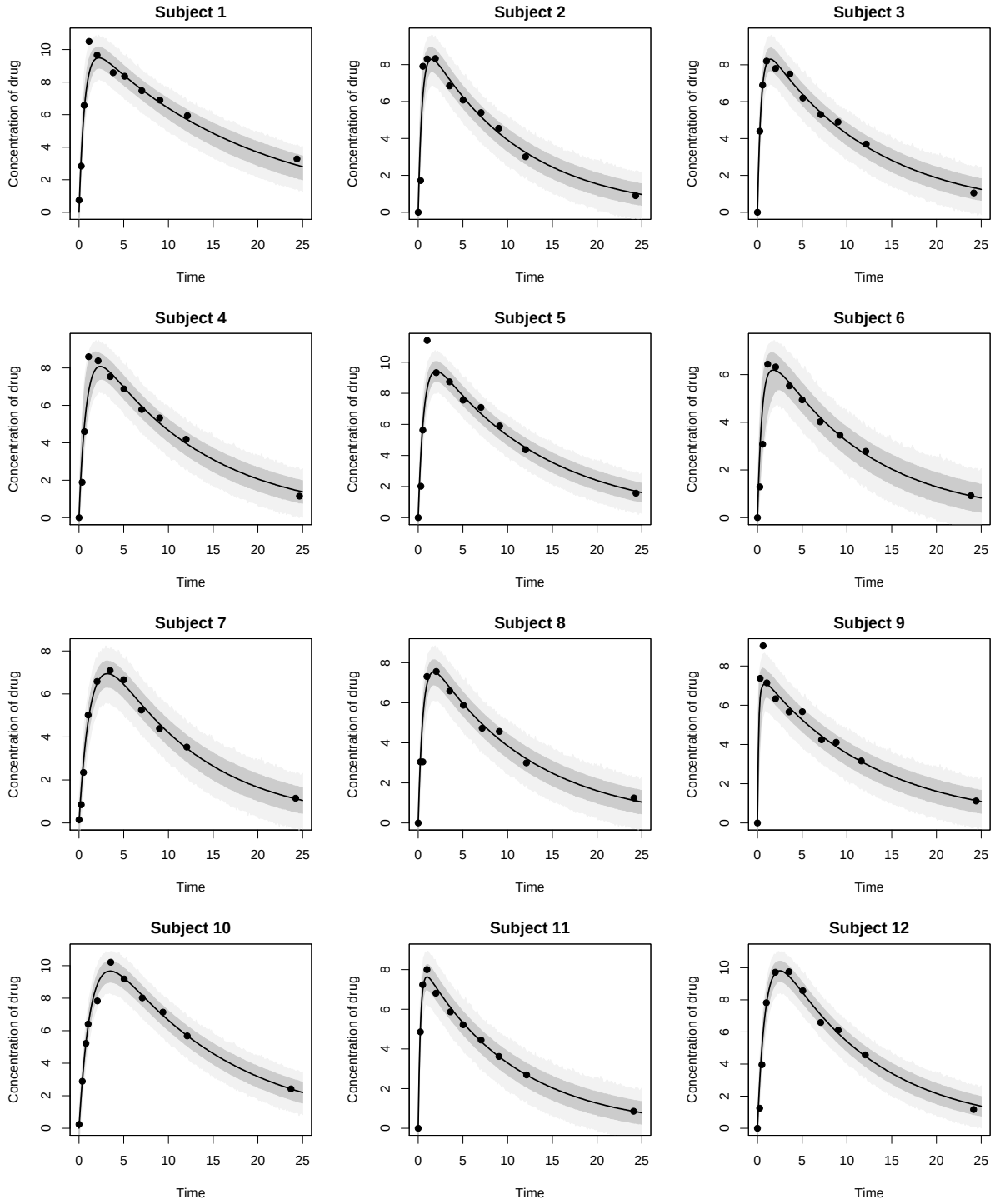


Figure III.6 – Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the concentration of Theophylline with pointwise posterior predictive median (black solid curve) versus time for the twelve subjects.

III.6 Discussion

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

The penalized mixture approaches avoid restrictive parametric assumptions on the error distribution. The Gaussian mixture assumption is convenient as, conditionally on the mixture component (i.e. on the group label variable), one is back to the standard B-ODE-P-spline setting assumed in Chapter II. The Student mixture assumption requires the use of one extra latent variable to go back to the standard B-ODE-P-spline setting. Furthermore, label switching that traditionally occurs is Gaussian/Student mixture model is here avoided by using a fixed grid of means and penalizing the transformed mixture weights.

Simulations suggest that the B-ODE-P-spline approach combined with penalized mixture of Gaussian distributions is as efficient as the standard B-ODE-P-spline approach in terms of relative bias, coverage probabilities, root mean squared errors and mean posterior standard deviation when the distribution of the error terms are Gaussian and outperforms the standard approach for small and medium sample sizes when the error distribution is homogeneous non-Gaussian.

The influence of the grid of means on the mixture of Gaussian distributions and the robustness of the mixture of Student distributions in the modelling of the error distribution was also presented. It was shown that working with the mixture of Gaussian distributions requires a sufficient wide grid of means (i.e. that includes all the standardized errors within its bounds). The mixture of Student distributions is less sensitive to the grid of means and more robust to outliers but is computationally more intensive.

The Bayesian ODE-penalized B-spline approach combined with penalized mixture of Gaussian distribution was used for the analysis of the Theophylline dataset (see Section III.5). Working with a penalized mixture of Gaussian distribution enabled to highlight the non-gaussianity of the standardized error distribution.

Appendix

III.A Bayesian ODE-penalized B-splines with Student error distribution

We assume that the dynamics of a state function $\boldsymbol{x}(t) \in \mathbb{R}^d$ is driven by the system of differential equations (III.1). Only the subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state functions involved in the ODE model are measured at time t_{jk}, $k = 1, \ldots, n_j$ with an additive measurement error ϵ_{jk}. We denote by $y_{jk} = x_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$ the corresponding measurement. For the observed jth state function ($j \in \mathcal{J}$), the error distribution is modeled using a standardized Student distribution with ν_j degrees of freedom. This ensure the zero mean and the unit variance of the error terms. The jth component of the state function $\boldsymbol{x}(t)$ is approximated using a B-spline basis function expansion:

$$\widetilde{x_j}(t) = (\boldsymbol{B_j}(t))^T \boldsymbol{c_j}.$$

where $\boldsymbol{B_j}(t)$ is the K_j-vector of B-spline basis functions of order p evaluated at time t and $\boldsymbol{c_j}$ the corresponding K_j-vector of B-spline coefficients.

III.A.1 The Bayesian model

We recall that the distribution of the error terms for the observation of the jth state function ($j \in \mathcal{J}$) is a standardized Student distribution with ν_j degrees of freedom. In order to manage the strong (posterior) correlation between the ODE parameter and the spline coefficients, we introduce the latent variable ω_{jk} for the kth observation of the jth state function. The latent variables ω_{jk}, $j \in \mathcal{J}$, $k = 1, \ldots, n_j$ are supposed independently drawn form a gamma distribution:

$$\omega_{jk} \sim \mathcal{G}\left(\frac{\nu_j}{2}; \frac{\nu_j - 2}{2}\right).$$

Given this latent variable ω_{jk}, one has:

$$\epsilon_{jk}|\omega_{jk} \sim \mathcal{N}\left(0; \omega_{jk}^{-1}\right).$$

This means, that conditionally on the latent variable ω_{jk}, one is back to the standard B-ODE-P-spline setting considered in Chapter II. The Bayesian model is similar to the one used in the

standard approach, but with an additional specification:

$$\begin{cases} y_{jk}|\omega_{jk}, \boldsymbol{c_j}, \tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B_j}\left(t_{jk}\right)\right)^T \boldsymbol{c_j}; \left(\omega_{jk}\tau_j\right)^{-1}\right) \\ \omega_{jk} & \sim & \mathcal{G}\left(\frac{\nu_j}{2}; \frac{\nu_j - 2}{2}\right) \\ p\left(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\gamma}\right) & \propto & \left(\det\left(\boldsymbol{V_1}\right)\right)^{1/2} \exp\left(-\frac{1}{2}\left\{\boldsymbol{c}^T\boldsymbol{V_1}\boldsymbol{c} - 2\boldsymbol{c}^T\boldsymbol{v_1} + \boldsymbol{v_1}^T\boldsymbol{V_1}^{-1}\boldsymbol{v_1}\right\}\right) \\ \gamma_j & \sim & \mathcal{G}\left(a_{\gamma_j}; b_{\gamma_j}\right) \\ \tau_j & \sim & \mathcal{G}\left(a_{\tau_j}; b_{\tau_j}\right) \\ \boldsymbol{\theta} & \sim & p\left(\boldsymbol{\theta}\right). \end{cases}$$

where the vector $\boldsymbol{v_1}$ and the matrix $\boldsymbol{V_1}$ are defined in Section III.2.1.5.
The strategy used to generate a sample form the joint posterior distribution is similar to the one presented in Sections III.2.2.3 and III.3.2.2. First, one can show that the conditional posterior distribution for the spline coefficients $\boldsymbol{c} = \left(\boldsymbol{c_1}^T, \ldots, \boldsymbol{c_d}^T\right)^T$ is a multivariate Gaussian distribution with mean $\boldsymbol{V_2}^{-1}\boldsymbol{v_2}$ and covariance matrix $\boldsymbol{V_2}^{-1}$ where:

$$\boldsymbol{v_2} = \boldsymbol{v_1} + \text{vec}\left(\boldsymbol{z_j}; j = 1, \ldots, d\right).$$

and

$$\boldsymbol{V_2} = \boldsymbol{V_1} + \text{diag}\left(\boldsymbol{Z_j}; j = 1, \ldots, d\right).$$

The vector $\boldsymbol{z_j}$ is either equal to $\tau_j \boldsymbol{B_j}^T \boldsymbol{\Omega_j} \boldsymbol{y_j}$ (if the jth state function is observed) or to the null vector and the matrix $\boldsymbol{Z_j}$ is either equal to $\tau_j \boldsymbol{B_j}^T \boldsymbol{\Omega_j} \boldsymbol{B_j}$ (if the jth state function is observed) or to the null matrix. The matrix $\boldsymbol{\Omega_j}$ is a diagonal with ω_{jk} on the (k,k)th entry. Let also $\boldsymbol{\omega}$ denote the concatenation of all the latent variables (i.e. $\boldsymbol{\omega} = \text{vec}\left(\text{vec}\left(\omega_{jk}; k = 1, \ldots, n_j\right); j \in \mathcal{J}\right)$).
The conditional posterior distribution for the latent variable ω_{jk} can be show to be:

$$\omega_{jk}|y_{jk}, \boldsymbol{c_j}, \tau_j \sim \mathcal{G}\left(\frac{\nu_j + 1}{2}; \frac{\tau_j\left(y_{jk} - \widetilde{x_j}\left(t_{jk}\right)\right)^2}{2} + \frac{\nu_j - 2}{2}\right).$$

Then the joint posterior distribution is marginalized with respect to the spline coefficients. The log of it is:

$$\begin{aligned}\log\left(p\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\omega}\right)\right) \doteq & \sum_{j\in\mathcal{J}}\left\{\frac{n_j}{2}\log\left(\tau_j\right)-\frac{\tau_j}{2}\boldsymbol{y_j}^T\boldsymbol{\Omega_j}\boldsymbol{y_j}\right\}\\ & -\frac{1}{2}\left\{\boldsymbol{v_1}^T\boldsymbol{V_1}^{-1}\boldsymbol{v_1}-\log\left(\det\left(\boldsymbol{V_1}\right)\right)-\boldsymbol{v_2}^T\boldsymbol{V_2}^{-1}\boldsymbol{v_2}+\log\left(\det\left(\boldsymbol{V_2}\right)\right)\right\}\\ & +\sum_{j\in\mathcal{J}}\left\{\left(a_{\tau_j}-1\right)\log\left(\tau_j\right)-b_{\tau_j}\tau_j\right\}+\sum_{j=1}^{d}\left\{\left(a_{\gamma_j}-1\right)\log\left(\gamma_j\right)-b_{\gamma_j}\gamma_j\right\}\\ & +\log\left(p\left(\boldsymbol{\theta}\right)\right).\end{aligned}$$

The conditional posterior distribution cannot be identified. Therefore, Metropolis-Hastings steps are used to generate sample from $p\left(\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\omega}\right)$. Note that the use of the latent variables $\boldsymbol{\omega}$ was crucial to enable the marginalization of the joint posterior distribution with respect to the spline coefficients.

III.A.2 Simulation study

Let us consider the simple exponential decay model:

$$\begin{cases}\dfrac{dx}{dt}\left(t\right) &= -\theta x\left(t\right),\\ x\left(0\right) &= 1,\end{cases}$$

We are interested by the estimation of (θ,τ) from a set of observations $\{(t_k,y_k):k=1,\ldots,n\}$ where $y_k=x\left(t_k\right)+\tau^{-1/2}\epsilon_k$. The sampling times correspond to n equidistant time points between 0 and 1 (with $n=20,50,100,200,500$). We consider a standardized Student error distribution with 6 degrees of freedom for the distribution of the error terms. For each configuration, 1000 datasets were generated with (θ,τ) equal to $(5,100)$.

Table III.5 gives the estimation based on MCMC for the relative bias (in percent, based on the estimated posterior mean), the coverage probabilities at level 80% and 95% (based on 80% and 95%-HPD intervals), the root mean squared error and the mean posterior standard deviation for parameter θ and τ. All the coverage probabilities are in agreement with their nominal values. The bias is small for ODE-parameter θ. For small sample sizes ($n=20$ and $n=50$), the bias for the precision of measurement is slightly positive. As expected, RMSE and MPSD decrease with sample size. Based on the results obtained on this very simple exponential decay example, we can consider that the proposed approach is able to manage the Student distribution case.

Table III.5 – Simulation - Student distribution with 6 degrees of freedom - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for θ and τ for five sample sizes ($n = 20, 50, 100, 200, 500$) using B-ODE-P-spline approach with Student distribution for the error term. Estimated coverage probabilities in bold are in agreement with their nominal values.

		$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
θ	RB	3.3	1.0	0.4	0.1	0.2
	CP80	**80.2**	**81.6**	**80.4**	**81.1**	**78.2**
	CP95	**95.2**	**95.2**	**94.7**	**95.1**	**94.0**
	RMSE	5.55e-1	3.08e-1	2.12e-1	1.46e-1	9.67e-2
	MPSD	5.80e-1	3.16e-1	2.13e-1	1.49e-1	9.32e-2
τ	RB	17.4	5.6	3.8	1.3	1.1
	CP80	**78.7**	**81.1**	**80.7**	**78.5**	**82.4**
	CP95	**94.8**	**95.2**	**96.0**	**94.4**	**95.5**
	RMSE	5.49e+1	2.81e+1	1.84e+1	1.30e+1	7.69
	MPSD	5.17e+1	2.71e+1	1.84e+1	1.25e+1	7.87

Chapter IV

On the use of adhesion parameters to validate models specified using systems of affine differential equations

Abstract

A strategy for the selection of systems of differential equations is proposed based on Bayesian ODE-penalized B-spline approach. It estimates the ODE parameters, approximates the solution of the ODE model and quantifies the suitability of the proposed differential equations to model the dynamics of the observed state functions. Simulation study confirms that these ODE-adhesion parameters are able to question a system of differential equations as a descriptor of the dynamics in the state functions. This methodology is illustrated on a pharmacokinetic study.

Keywords: Bayesian ODE-Penalized-B-spline; ODE-adhesion parameter; ODE-model selection.

IV.1 Introduction

The frequentist ODE-penalized B-spline approach proposed by Ramsay et al. (2007) was translated into a full Bayesian ODE-penalized B-spline setting when the ODEs are affine and the conditional distribution of the response is either Gaussian (Chapter II) or suitable described by a mixture of Gaussian/Student distributions (Chapter III). In these ODE-penalized B-spline approaches, the unknown state functions involved in the system of differential equations is approximated using linear combinations of a large number of B-splines with flexibility counter-

balanced by a penalty term related to the set of differential equations. This ODE-penalized smoothing approach is a trade off between fitting the data and solving the ODE. Therefore it appears to be robust to model misspecification and to disturbances of the dynamical system. Compared to the frequentist ODE-penalized B-spline approach, the Bayesian one automatically selects the ODE-adhesion parameters with uncertainty measures about model parameters quantified using MCMC.

Model selection is an unavoidable topic. In some cases, several competing models can be proposed for the modeling of a state function. In pharmacokinetics, different models can be suggested to describe the dynamics of a drug concentration (Lindsey et al., 2000). In other cases, even if the dynamical system is perfectly known, it may be too difficult to collect enough data to reliably estimate all the parameters involved in the system of differential equations. Then, one needs either to determine which parameters can be estimated and which should remain to their reference value (Woloszyn and McAuley, 2011), or to select a simplified model (Wu et al., 2011) that gives better predictions and/or more accurate parameter estimates. Generally, in traditional frequentist approach, either F tests are used to select among nested models or information criterion (AIC, BIC and all the sub-criteria that depend on them, Mallows C_p, ...) when the competing models are non-nested. With the frequentist ODE-penalized B-spline approach, ODE-adhesion parameters can be used to select competing ODE models but the value of the ODE-adhesion parameter does not statistically confirm that the dynamics of the state function is driven by the selected ODE model. Moreover, as the ODE-adhesion parameters are selected, uncertainty measure about the selection of a specific ODE model in comparison to a competitor is unavailable.

The goal of this chapter is to propose a Bayesian model selection strategy based on the study of ODE-adhesion parameters. Section IV.2 reminds the key points of the Bayesian ODE-penalized B-spline approach and proposes to use the posterior distribution for the ODE-adhesion parameters for the ODE-model selection. An intensive simulation study is presented in Section IV.3 to confirm the relevance of the strategy. An application to pharmacokinetics is presented in Section IV.4 to illustrate the usefulness of the ODE-adhesion parameter in the selection of two competing models. We conclude the chapter by a discussion in Section IV.5.

IV.2 Bayesian ODE-penalized B-spline approach and ODE-model selection

In this section, we first recall some key points of the ODE-penalized B-spline approach. Then we explain how to use the ODE-adhesion parameters in order to question a proposed system of differential equations as descriptor of the dynamics in the state function or how to use these ODE-adhesion parameters for the selection of competing ODE models.

IV.2.1 The Bayesian ODE-penalized B-spline approach

For simplicity, we consider a single subject study with a standard Gaussian for the error distribution. We refer to Chapter II for a hierarchical setting and Chapter III for flexible distributions of the error terms. It is assumed that only a subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state functions $\boldsymbol{x}(t) = (x_1(t), \ldots, x_d(t))^T$ are measured at time point t_{jk}, $j \in \mathcal{J}$, $k = 1, \ldots, n_j$ with additive standardized measurement error ϵ_{jk}:

$$y_{jk} = \boldsymbol{x}_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$$

where the dynamic of the state function is given by a set of affine differential equations:

$$\frac{d\boldsymbol{x}}{dt}(t) = f(\boldsymbol{x}, t, \boldsymbol{\theta}), t \in [0; T] \tag{IV.1}$$

where f is a known affine function with respect to $\boldsymbol{x}$ and $\boldsymbol{\theta} \in \mathbb{R}^q$ a vector of unknown parameters. The objective is fourfold: estimate the ODE parameters $\boldsymbol{\theta}$, the state function $\boldsymbol{x}(t)$, the precision of measurement $\boldsymbol{\tau} = \text{vec}(\tau_j; j \in \mathcal{J})$ and confirm the suitability of the system of differential equations given in Eq. (IV.1) to describe the dynamic of the observed state functions using the set of observations $\{(t_{jk}, y_{jk}), j \in \mathcal{J}, k = 1, \ldots, n_j\}$.
The Bayesian ODE-penalized B-spline approach rests on three pillars. First of all, each state function $x_j(t), j \in \{1, \ldots, d\}$ involved in the system of differential equations is approximated using a B-spline basis function expansion:

$$\widetilde{x}_j(t) = (\boldsymbol{B_j}(t))^T \boldsymbol{c_j},$$

where $\boldsymbol{B_j}(t)$ is the K_j-vector of B-spline basis function evaluated at time t and $\boldsymbol{c_j}$ the K_j-vector of spline coefficients. Then, a penalty term that is related to the ODE model is used to constrain the variation of the spline coefficients. This penalty term involves some ODE-

adhesion parameters. Finally, this ODE-model based penalty term is combined to other prior informations to construct the full Bayesian model:

$$\left\{\begin{array}{rcl} y_{jk}|\boldsymbol{c_j},\tau_j & \sim & \mathcal{N}\left(\left(\boldsymbol{B_j}\left(t_{jk}\right)\right)^T\boldsymbol{c_j};\tau_j^{-1}\right) \\ p\left(\boldsymbol{c}|\boldsymbol{\theta},\boldsymbol{\gamma}\right) & \propto & \exp\left(-\frac{1}{2}\left\{PEN+\boldsymbol{c}^T{\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{c}-2\boldsymbol{c}^T{\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}\right\}\right) \\ \gamma_j & \sim & \mathcal{G}\left(a_{\gamma_j};b_{\gamma_j}\right) \\ \tau_j & \sim & \mathcal{G}\left(a_{\tau_j};b_{\tau_j}\right) \\ \boldsymbol{\theta} & \sim & p\left(\boldsymbol{\theta}\right). \end{array}\right.$$

The full fidelity-to-ODE measure PEN used in the prior distribution for the spline coefficients $\boldsymbol{c}$ is given by:

$$PEN\left(\widetilde{\boldsymbol{x}}|\boldsymbol{\gamma}\right)=\sum_{j=1}^{d}\gamma_j PEN_j\left(\widetilde{\boldsymbol{x}}\right) \tag{IV.2}$$

where $PEN_j\left(\widetilde{\boldsymbol{x}}\right)$ assessed the proximity on $[0;T]$ between the approximation to the state function, $\widetilde{x_j}\left(t\right)$, and the jth component of the solution $\boldsymbol{x}\left(t\right)$ to the set of differential equations in (IV.1):

$$PEN_j\left(\widetilde{\boldsymbol{x}}\right)=\int\left\{\frac{d\widetilde{x_j}}{dt}\left(t\right)-f_j\left(\widetilde{\boldsymbol{x}},t,\boldsymbol{\theta}\right)\right\}^2 dt.$$

For the conditional precision τ_j $(j\in\mathcal{J})$ of the vector of response $\boldsymbol{y_j}$ and each ODE-adhesion parameter γ_j $(j\in\{1,\ldots,d\})$, gamma prior distributions are convenient choices:

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

where $\mathcal{G}\left(a,b\right)$ denotes the gamma distribution with mean a/b and variance a/b^2. As explained in simulation study given in Section II.5.1, for each ODE-adhesion parameters γ_j, we opt for a gamma prior distribution with shape a_{γ_j} equal to 1 and the rate b_{γ_j} equal to a small value (e.g. b_{γ_j} smaller than 10^{-6}). This choice for the prior distributions of the ODE-adhesion parameters translates our prior confidence in the specification of the system of differential equations as descriptor of the dynamics of the state function. Indeed, such a prior for γ_j puts slightly more weight on value of $\log_{10}\left(\gamma_j\right)$ around $-\log_{10}\left(b_{\gamma_j}\right)$.

IV.2.2 Study of the posterior distribution for the ODE-adhesion parameter

For convenience, assume that all the ODE-adhesion parameters γ_j $(j \in \mathcal{J})$ are identical to a common value γ and that no prior information is available for the initial condition of the state function (implying that matrix $\boldsymbol{\Sigma_c}^{-1}$ and vector $\boldsymbol{\mu_c}$ are null). Therefore, in the ODE-model based penalty, the penalty vector and the penalty matrix can be rewritten as $\boldsymbol{r}(\boldsymbol{\theta}, \gamma) = \gamma \boldsymbol{s}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta}, \gamma) = \gamma \boldsymbol{S}(\boldsymbol{\theta})$. Note that the ODE-adhesion parameter γ appears as a multiplicative factor in the precision matrix of the multivariate Gaussian distribution that defines the prior distribution for the spline coefficients. Therefore, working with a gamma prior distribution forces the positiveness of the ODE-adhesion parameters. Note also that this prior is in this setting a conditional conjugate prior distribution.
From Section II.3.1, the log conditional posterior density for the ODE-adhesion parameter (marginalized with respect to the spline coefficients) is equal to:

$$\begin{aligned} \log\left(p\left(\gamma|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{y}\right)\right) &\doteq \frac{K}{2}\log\left(\gamma\right) - \frac{\gamma}{2}\boldsymbol{s}^T\left(\boldsymbol{\theta}\right)\boldsymbol{S}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{s}\left(\boldsymbol{\theta}\right) \\ &- \frac{1}{2}\log\left(\det\left(\boldsymbol{V_2}\right)\right) + \frac{1}{2}\boldsymbol{v_2}^T\boldsymbol{V_2}^{-1}\boldsymbol{v_2} \\ &+ \log\left(p\left(\gamma\right)\right), \end{aligned} \tag{IV.3}$$

where $K = \sum_{i=1}^{d} K_j$, $\boldsymbol{V_2} = \gamma \boldsymbol{S}(\boldsymbol{\theta}) + \operatorname{diag}\left(\boldsymbol{Z_j}; j = 1, \ldots, d\right)$ with matrix $\boldsymbol{Z_j}$ that is equal to $\tau_j \boldsymbol{B_j}^T \boldsymbol{B_j}$ if the jth state function is observed and the null $K_j \times K_j$-matrix otherwise, and vector $\boldsymbol{v_2} = -\gamma \boldsymbol{s}(\boldsymbol{\theta}) + \operatorname{vec}\left(\boldsymbol{z_j}; j = 1, \ldots, d\right)$ with $\boldsymbol{z_j}$ equal to $\tau_j \boldsymbol{B_j}^T \boldsymbol{y_j}$ if the jth state function is observed and to the null K_j-vector otherwise. Note that matrix $\boldsymbol{B_j}$ is the $n_j \times K_j$-matrix with $\left(\boldsymbol{B_j}\left(t_{jk}\right)\right)^T$ on kth line $(k = 1, \ldots, n_j)$.

IV.2.2.1 Asymptotics of the marginalized posterior distribution for ODE-adhesion parameter

We will show that the difference between the conditional posterior distribution for the ODE-adhesion parameter and its prior tends to a constant independent of γ when γ is large. Large values of γ are supported by the data when the differential equations adequately describe the dynamics of the system. In the log posterior density given in Eq. (IV.3), two terms are of particular interest: $\log\left(\det\left(\boldsymbol{V_2}\right)\right)$ and $\boldsymbol{v_2}^T\boldsymbol{V_2}^{-1}\boldsymbol{v_2}$. One can show that:

$$\boldsymbol{V_2} = \gamma\left(\boldsymbol{S}\left(\boldsymbol{\theta}\right) + \operatorname{diag}\left(\frac{1}{\gamma}\boldsymbol{Z_j}; j = 1, \ldots, d\right)\right),$$

with $\frac{1}{\gamma}\boldsymbol{Z_j}$ tending to the null matrix when γ tends to infinity (it is always equal to the null matrix when the jth state function is unobserved) and

$$\boldsymbol{v_2} = \gamma\left(-\boldsymbol{s}\left(\boldsymbol{\theta}\right) + \text{vec}\left(\frac{1}{\gamma}\boldsymbol{z_j}; j = 1, \ldots, d\right)\right),$$

with $\frac{1}{\gamma}\boldsymbol{z_j}$ tending to the null vector when γ tends to infinity (it is always equal to the null vector when the jth state function is unobserved). Thus,

$$\lim_{\gamma\to\infty}\left(\log\left(\det\left(\boldsymbol{V_2}\right)\right)\right) \doteq K\log\left(\gamma\right),$$

and

$$\lim_{\gamma\to\infty}\left(\boldsymbol{v_2}^T\boldsymbol{V_2}^{-1}\boldsymbol{v_2}\right) \doteq \gamma\boldsymbol{s}^T\left(\boldsymbol{\theta}\right)\boldsymbol{S}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{s}\left(\boldsymbol{\theta}\right).$$

Therefore, the difference between the log posterior density and the log prior density for the ODE-adhesion parameter tends to a constant independent of γ when γ tends to infinity:

$$\lim_{\gamma\to\infty}\left(\log\left(p\left(\gamma|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{y}\right)\right) - \log\left(p\left(\gamma\right)\right)\right) \doteq 0.$$

This result confirms our choice of the gamma distribution with shape equal to 1 and rate equal to small value for the prior distribution for the ODE-adhesion parameter. Indeed, the posterior distribution for this parameter will be substantially identical to its prior if this distribution underlines our (prior) confidence in the proposed ODE system to model the dynamics of the observed state functions. Note that such a prior expresses our prior belief that the proposed system of ODEs correctly describes the dynamics of the observed system. However, if the ODE model does not correctly describes the dynamics of state functions and if the number of observations is larger enough, then the posterior distribution for the ODE-adhesion parameter will shift the probability from the prior distribution towards smaller values of the ODE-adhesion parameter and question the model specification.

IV.2.2.2 Divergence measure and ODE-model selection

Asymptotic results from Section IV.2.2.1 confirm the statement of Jaeger and Lambert (2011), namely that the posterior distribution for the ODE-adhesion parameter is similar to its prior if the proposed system of differential equations adequately models the dynamics of the state function.

Let us recall that a sample from the posterior distribution for the ODE-adhesion parameter

can be obtained using a Metropolis-within-Gibbs algorithm (Sections II.3.2 and III.2.2.3). We denote by $\left\{\gamma^{(m)}; m = 1, \ldots, M\right\}$ the ordered sample from the posterior distribution for the ODE-adhesion parameter. The goal is now to measure the divergence between the prior and the posterior distribution for the ODE-adhesion parameter using the MCMC sample of the posterior distribution for the ODE-adhesion parameter.

To this end, we propose to consider several goodness of fit statistics. Let denote $F_M(\cdot)$ the empirical cumulative distribution function based on the sample of the posterior distribution for the ODE-adhesion parameter and $F(\cdot)$ the gamma prior distribution function. The first class of divergence measures are related to the empirical cumulative distribution function. The Kolmogorov-Smirnov (KS) statistic

$$\text{KS} = \sup_{\gamma} |F_M(\gamma) - F(\gamma)|$$

quantifies the maximal distance between the empirical distribution function F_M and the reference distribution function F. This statistics tends to be more sensitive near the center of the distribution than at tails.

The Cramér-von Mises (CVM) statistic integrates the whole squared difference between the empirical cumulative distribution function F_M and the corresponding (target) prior cumulative distribution function F:

$$\text{CVM} = M \int_{-\infty}^{\infty} (F_M(\gamma) - F(\gamma))^2 \, dF(\gamma).$$

Note that this statistic can be rewritten as:

$$\text{CVM} = \sum_{m}^{M} \left(\frac{2m-1}{2} - F\left(\gamma^{(m)}\right) \right)^2 + \frac{1}{12M},$$

and can be therefore simply computed using the sample of the posterior distribution for the ODE-adhesion parameter. The Anderson-Darling (AD) statistic also integrates the whole squared difference but with a weighting function:

$$\text{AD} = M \int_{-\infty}^{\infty} \frac{(F_M(\gamma) - F(\gamma))^2}{F(\gamma)(1 - F(\gamma))} dF(\gamma).$$

Compared to the CVM statistic, the AD statistic puts more weight on observations in the tails of the target prior distribution. Again, the inconvenient integral can be replaced by a sum:

$$\mathrm{AD} = -\frac{1}{M}\sum_{m}^{M}(2m-1)\left(\log\left(F\left(\gamma^{(m)}\right)\right) - \log\left(F\left(\gamma^{(M+1-m)}\right)\right)\right) - M.$$

The second class of divergence measure is the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951):

$$\mathrm{KL}\left(p\left(\cdot|\boldsymbol{y}\right)|p\left(\cdot\right)\right) = \int p\left(\gamma|\boldsymbol{y}\right)\log\left(\frac{p\left(\gamma|\boldsymbol{y}\right)}{p\left(\gamma\right)}\right)d\gamma.$$

As the normalizing constant of the posterior density function for the ODE-adhesion parameter is unknown, this KL divergence is estimated by:

$$\mathrm{KL}\left(p\left(\cdot|\boldsymbol{y}\right)|p\left(\cdot\right)\right) \approx \frac{1}{M}\sum_{m=1}^{M}\log\left(\widetilde{q}\left(\gamma_m\right)\right) - \frac{1}{M}\sum_{m=1}^{M}\log\left(p\left(\gamma_m\right)\right)$$

where $\{\gamma^m; m = 1, \ldots, M\}$ is a sample of the posterior distribution for the ODE-adhesion parameter and $\widetilde{q}\left(\cdot\right)$ denotes a kernel density estimate of the posterior distribution for the ODE-adhesion parameter based on that sample. Note that in order to stabilize the kernel density estimate, we recommend considering the ODE-adhesion parameter on the log10 scale.

IV.3 Simulation

Intensive simulations were put in place to confirm the ability of the ODE-adhesion parameter to select or validate ODE models.
All the scenarios are based either on the exponential decay case

$$\left\{\begin{array}{rcl} \dfrac{dx}{dt}\left(t\right) & = & -k_e x\left(t\right) \\ x\left(0\right) & = & 1, \end{array}\right. \tag{IV.4}$$

or on the bi-exponential decay case

$$\left\{\begin{array}{rcl} \dfrac{dx_1}{dt}\left(t\right) & = & -\left(k_{cp}+k_e\right)x_1\left(t\right) + k_{pc}x_2\left(t\right) \\ \dfrac{dx_2}{dt}\left(t\right) & = & k_{cp}x_1\left(t\right) - k_{pc}x_2\left(t\right) \\ x_1\left(0\right) & = & 1 \\ x_2\left(0\right) & = & 0. \end{array}\right. \tag{IV.5}$$

If Eq. (IV.4) (resp. Eq. (IV.5)) is considered, then the simulated observations are given by $y_k = x(t_k) + \tau^{-1/2}\epsilon_k$ (resp. $y_k = x_1(t_k) + \tau^{-1/2}\epsilon_k$) where $\{t_k; k = 1, \ldots, n\}$ is a set of n equidistant time points between 0 and 1. Unless it is specified, we consider the standardized Gaussian distribution for the error term ϵ_k. For each simulation, 1000 datasets were generated. The prior distribution of the ODE-adhesion parameter supposed for the complete simulation is a gamma distribution $\mathcal{G}(1; 10^{-8})$. The results provided in the following subsections describe the simulation results.

IV.3.1 Sensitivity to the sample size

The aim of this simulation is to see if the posterior distribution for the ODE-adhesion parameter is sensitive to the sample size. To do that, we generate data using Eq. (IV.4) with seven sample sizes (namely $n = 10, 20, 50, 100, 200, 500$ and 1000) and $(k_e, \tau) = (5, 100)$. All the datasets were analyzed with the Bayesian model specified in Section IV.2.1 with ODE parameter k_e and precision of measurement set to their simulation values. Figure IV.1 shows the that posterior median of $\log(\gamma)$ is not impacted by the sample size.

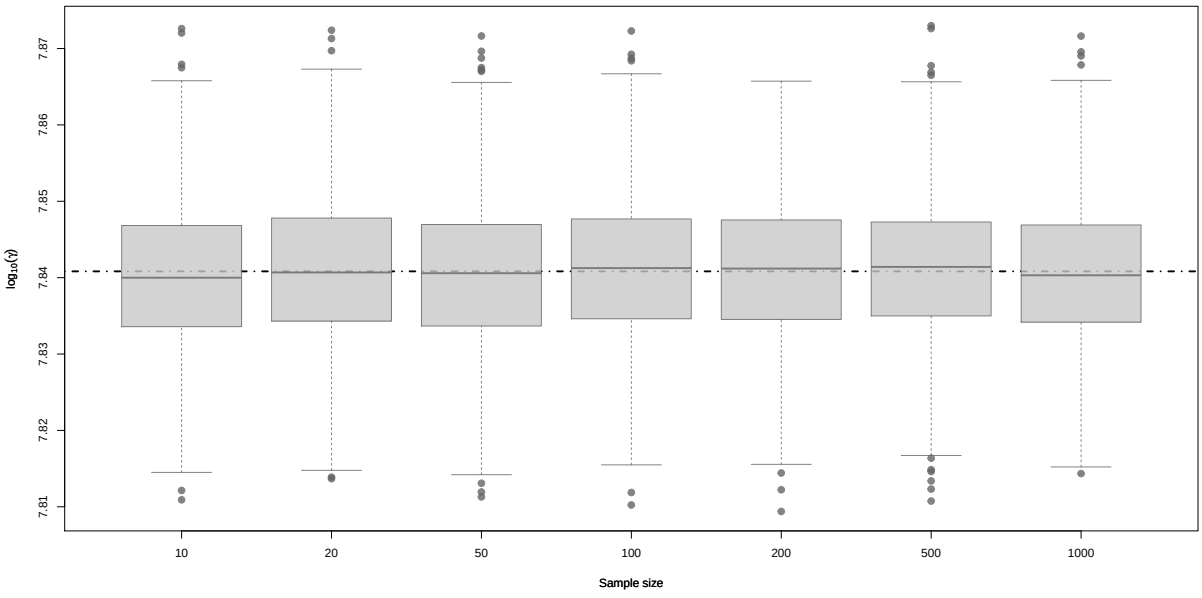


Figure IV.1 – Sensitivity to the sample size - Boxplot of posterior empirical median of $\log_{10}(\gamma)$ for the seven sample size with prior median (horizontal dashed line).

This is also visually confirmed for all the divergence measure (e.g. the Kullback-Leibler divergence in Figure IV.2).

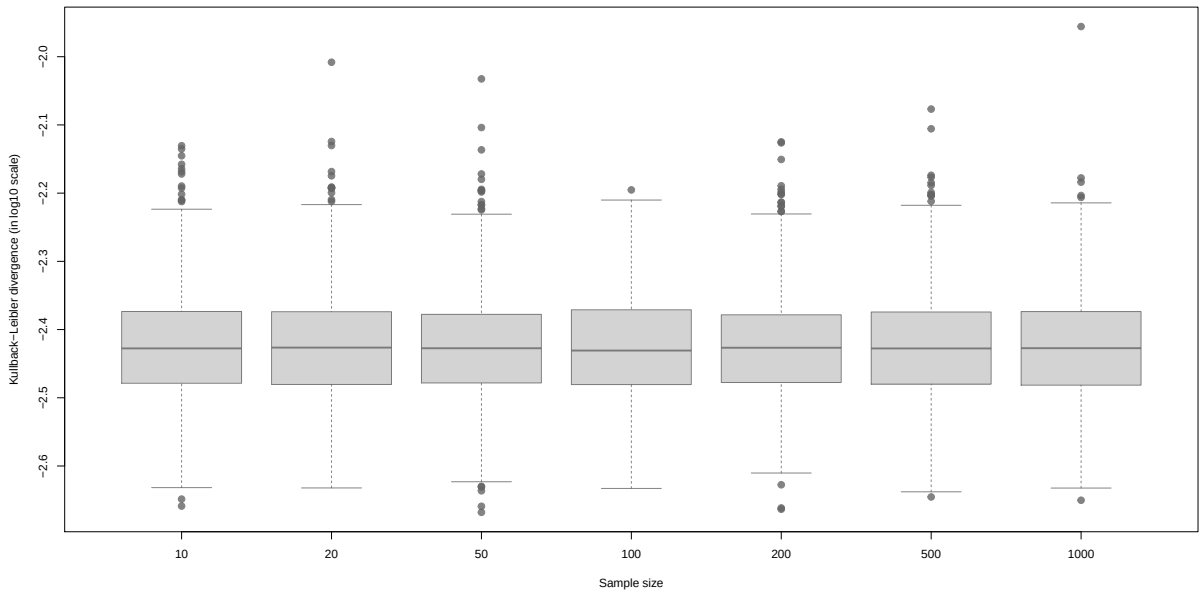


Figure IV.2 – Sensitivity to the sample size - Boxplot of estimated Kullback-Leibler divergence (on the log10 scale) for the seven sample size.

IV.3.2 Sensitivity to a wrong specification of the ODE parameter or of the precision of measurement

Wrong specification of the ODE parameter

We consider Eq. (IV.4) with seven sample sizes (namely $n = 10$, 20, 50, 100, 200, 500 and 1000) and simulate 1000 datasets with $(k_e, \tau) = (5, 100)$. All the datasets were analyzed with the Bayesian model specified in Section IV.2.1 with the precision of measurement τ set to its simulation value. A grid of 33 equidistant values between 3 and 7 for the ODE parameter k_e (centered on the simulation value $k_e = 5$) was considered for the analysis of each simulated dataset. The posterior distribution for the ODE-adhesion parameter γ was sensitive to a wrong specification of the ODE parameter k_e especially when the sample size increases. Figure IV.3 and Table IV.1 show that the AD statistic and the KL divergence were more likely to detect differences between the prior and the posterior distribution for the ODE-adhesion parameter. This could be explained by the importance of the distribution tails in the definition of these two quantities. Figure IV.4 shows the histogram of a sample of the posterior distribution for the ODE-adhesion parameter when the ODE parameter k_e was set either to the value used for the data generation ($k_e = 5$, dark grey histogram) or set to a wrong value ($k_e = 4$, light grey histogram). It visually confirms that the posterior distribution for the ODE-adhesion parameter is similar to its prior when the model is adequately specified whereas the posterior

distribution for the ODE-adhesion parameter differs from its prior when the ODE model is wrongly specified.

Table IV.1 – Sensitivity to a wrong specification of the ODE parameter - Proportion of divergence measures (Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence) larger than the one estimated with $k_e = 5$ when $n = 100$ with k_e equals to 3, 4, 6 or 7 in the MCMC procedure.

		Divergence measure			
		KS	CVM	AD	KL
k_e	3	100	100	100	100
	4	74.2	73.9	80.4	85.2
	6	54.3	55.4	56.6	59.8
	7	87.7	89.1	91.4	91.6

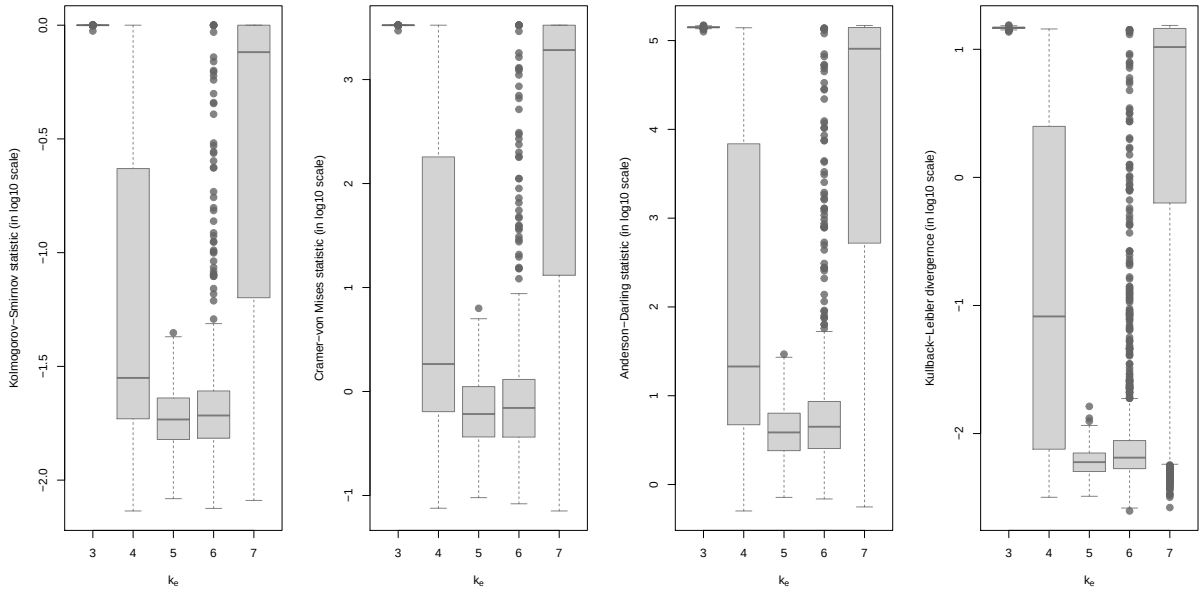


Figure IV.3 – Sensitivity to a wrong specification of the ODE parameter - Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence (on the log10 scale) when $n = 100$ and k_e is set equal to 3, 4, 5, 6 or 7 in the MCMC procedure (while $k_e = 5$ was used to generate datasets).

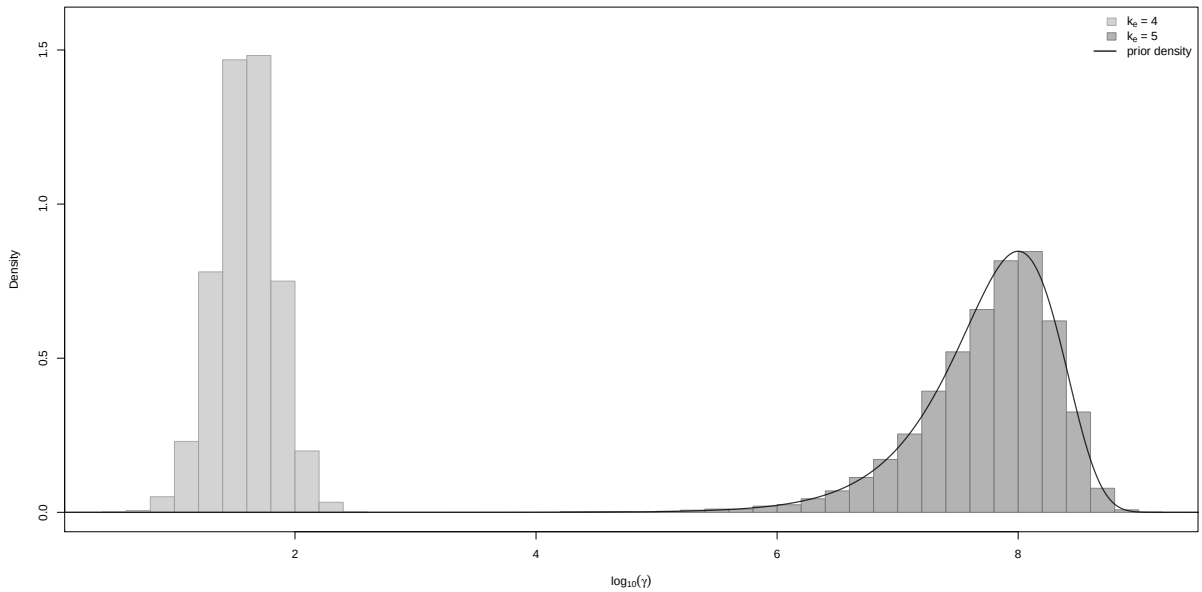


Figure IV.4 – Sensitivity to a wrong specification of the ODE parameter - Histograms based on a sample of the posterior distribution for the ODE-adhesion parameter when $n = 100$ and k_e is equal to 4 (light gray histogram) or 5 (dark grey histogram) while $k_e = 5$ was used to generate datasets.

Wrong specification of the precision of measurement

Again, we consider Eq. (IV.4) with seven sample sizes (namely $n = 10$, 20, 50, 100, 200, 500 and 1000) and simulate 1000 datasets with $(k_e, \tau) = (5, 100)$. All the datasets were analyzed with the Bayesian model specified in Section IV.2.1 with ODE parameter k_e equals to its simulation value. A grid of 25 equidistant values between -1 and 5 for $\log_{10}(\tau)$ (centered on the simulation value $\log_{10}(\tau) = 2$) was considered for the analysis of each simulated dataset. The posterior distribution for the ODE-adhesion parameter is only affected by values of the precision larger than the one used for the data generation (see Figure IV.5). This result was expected as the B-spline approximation should get closer to the data when the precision is set to a larger value than the one used for the data generation. This implies that the fitted state function is too wiggly to be driven by the proposed differential equation. Table IV.2 seems to confirm that the AD statistic and the KL divergence are more likely to detect differences between the prior and the posterior distribution for the ODE-adhesion parameter.

Table IV.2 – Sensitivity to a wrong specification of the precision of measurement - Proportion of divergence measures (Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence) larger than the one estimated with $\tau = 100$ when $n = 100$ with $\log_{10}(\tau)$ equals to 1, 1.5, 2, 2.5, 2.75 or 3 in the MCMC procedure.

		Divergence measure			
		KS	CVM	AD	KL
$\log_{10}(\tau)$	1	51.9	50.5	51.6	50.6
	1.5	47	48.1	47.8	48.2
	2.5	62	63.1	66.1	67.3
	2.75	92.4	92.8	94.2	94.0
	3	99.5	99.5	99.4	99.5

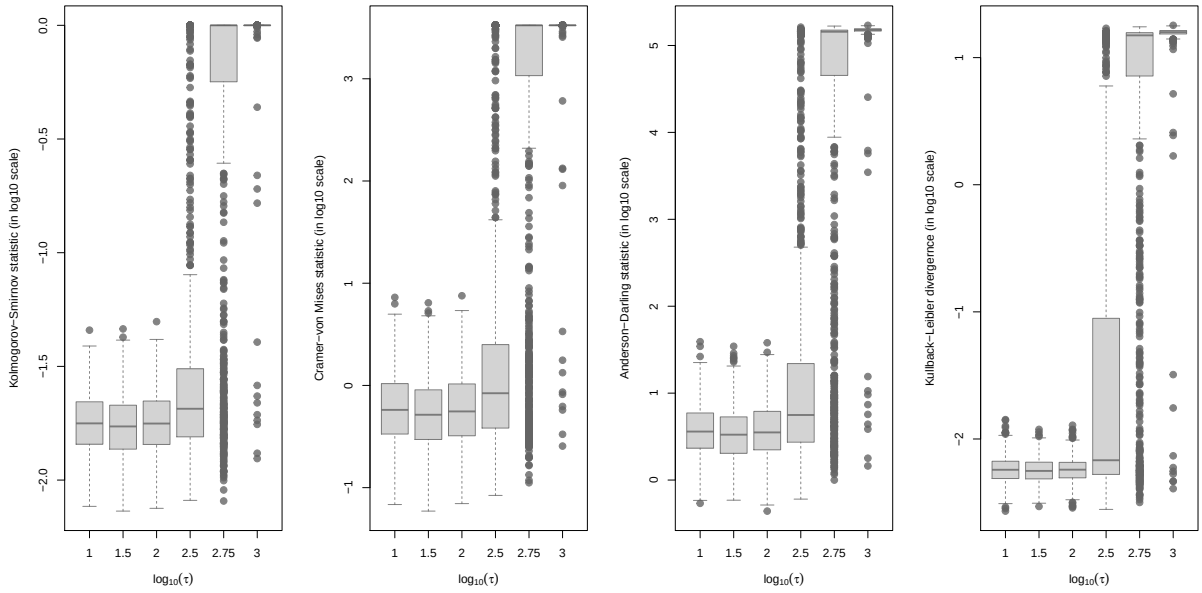


Figure IV.5 – Sensitivity to a wrong specification of the precision of measurement - Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence (on the log10 scale) when $n = 100$ and $\log_{10}(\tau)$ is set equal to 1, 1.5, 2, 2.5, 2.75 or 3 in the MCMC procedure (while $\tau = 100$ was used to generate datasets).

IV.3.3 Sensitivity to a misspecification of the error distribution

The aim of this simulation is to see if the posterior distribution for the ODE-adhesion parameter is sensitive to a misspecification of the error distribution. Again, we consider Eq. (IV.4) with sample size equal to $n = 10$, 20, 50, 100, 200, 500 and 1000 and four possible error distributions (Gaussian, Student t_6, extreme value and the Gaussian mixture $0.4\mathcal{N}(-1.5; 0.9^2) + 0.6\mathcal{N}(1; 0.6^2)$). All the simulated datasets were analyzed with the Bayesian model specified in Section IV.2.1 assuming a Gaussian error distribution with ODE parameter k_e and precision of measurement τ equal to their simulation values. According to the simulation results, the posterior distribution for the ODE-adhesion parameter seems to be insensitive to the misspecification of the error distribution (see Figure IV.6 and Table IV.3 for the Kullback-Leibler divergence when then sample size is equal to 100). This result is consistent with the simulation results obtained in Chapter III.

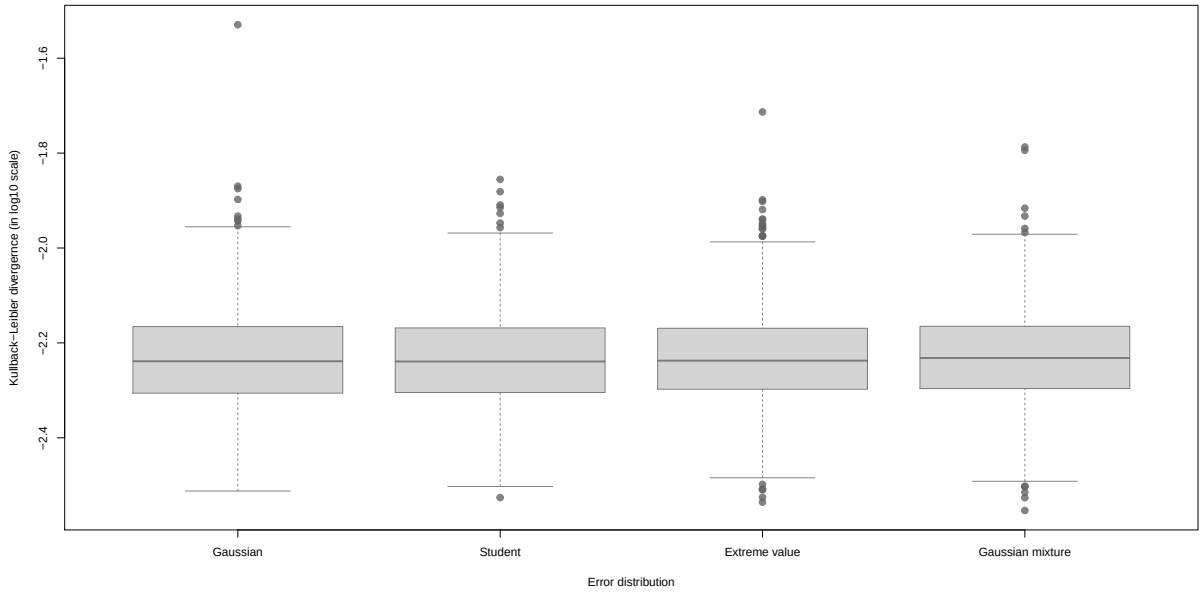


Figure IV.6 – Sensitivity to a misspecification of the error distribution - Boxplot of estimated Kullback-Leibler divergence (on the log10 scale) for sample size equals to 100 with four error distributions (Gaussian, Student, extreme value and mixture of two Gaussians) used for the data generation. Gaussian distribution used for the analysis of each simulated dataset.

Table IV.3 – Sensitivity to a misspecification of the error distribution - Quantile at 5%, 10%, 50%, 90% and 95% of the log10 estimated Kullback-Leibler divergence for sample size equals to 100 with four error distributions (Gaussian, Student, extreme value and mixture of two Gaussians) used for the data generation. A Gaussian distribution was wrongly assumed in the analysis of each simulated dataset.

		Error distribution			
		Gaussian	Student	Extreme value	Gaussian mixture
Quantile	5%	-2.40	-2.40	-2.40	-2.39
	10%	-2.36	-2.36	-2.36	-2.35
	50%	-2.24	-2.24	-2.24	-2.23
	90%	-2.09	-2.10	-2.10	-2.09
	95%	-2.04	-2.06	-2.06	-2.05

IV.3.4 Sensitivity to a ODE-model misspecification

For this simulation, we consider Eq. (IV.5) and simulate 1000 datasets with seven sample sizes (namely $n = 10$, 20, 50, 100, 200, 500 and 1000) and parameters $(k_e, k_{cp}, k_{pc}, \tau) = (5, 3, 2, 1000)$ and a Gaussian error term. Each simulated dataset is analyzed with the Bayesian model specified in Section IV.2.1 wrongly assuming that the dynamic of the observed state function is given by Eq. (IV.4). Figure IV.7 shows that, based on the Kullback-Leibler divervence, the posterior distribution for the ODE-adhesion parameter is not sensitive to the model misspecification for small sample sizes ($n = 10$ and $n = 20$). The ODE-model misspecification is clearly detected when the sample size is at least 100. Note also that the posterior distribution for the ODE-adhesion parameter appears to be sensitive to the sample size when the ODE-model misspecification is detected.

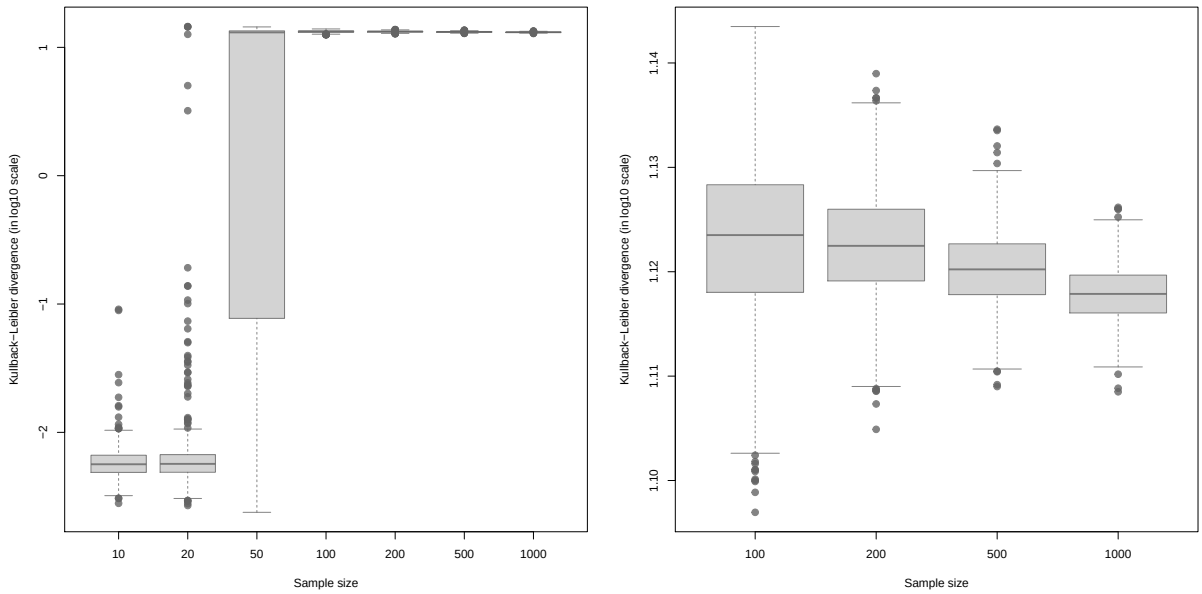


Figure IV.7 – Sensitivity to ODE-model misspecification - Boxplot of the Kullback-Leibler divergence (on the log10 scale) estimated from data generated using a bi-exponential decay model while an exponential decline one was (wrongly) assumed in the analysis. Figure on the right is a zoom for large sample sizes.

IV.4 Application

The selection capabilities of the ODE-adhesion parameter is illustrated on the pharmacokinetic dataset Indometh (Kwan et al., 1976) available in `R`.

IV.4.1 Data and objective

The Indomethacin dataset was presented in Section I.2.2.2. The evolution of the concentration of the drug in blood on the six subjects is given in Figure I.2.
The goal of this study is to confirm that the kinetic of Indomethacin after a bolus intravenous injection is modeled by a 2-compartment intravenous bolus model and that the 1-compartment intravenous bolus model fails to describe the kinetics of the drug in the plasma.
In the 1-compartment intravenous bolus model, the entire body is assimilated to a unique central compartment of volume $V > 0$ where the dynamic of the concentration $x(t)$ is described by this differential equation:

$$\begin{cases} \dfrac{dx}{dt}(t) &= -k_e x(t) \\ x(0) &= \frac{D}{V}, \end{cases}$$

where D is the intravenous dose and k_e the rate of elimination of the drug from the central compartment.

The 2-compartment intravenous bolus model arises when a part of the drug in the central compartment passes through a peripheral compartment. Let $x_1(t)$ denote the concentration of drug in the central compartment of volume V and $x_2(t)$ the amount of drug in the peripheral compartment at time t. A part of the drug is eliminated from the central compartment with rate k_e, another part goes to the peripheral compartment at rate k_{cp} and comes back in the central compartment at rate k_{pc}. In this setting, only the concentration of drug in the central compartment is usually observed. The system of differential equations with the initial conditions is:

$$\begin{cases} \frac{dx_1}{dt}(t) &= -(k_e + k_{cp})\, x_1(t) + \frac{k_{pc}}{V} x_2(t) \\ \frac{dx_2}{dt}(t) &= k_{cp} V x_1(t) - k_{pc} x_2(t) \\ x_1(0) &= \frac{D}{V} \\ x_2(0) &= 0. \end{cases}$$

For this model, it is supposed that k_e, k_{cp}, k_{pc} and V are positive with $k_{cp} > k_{pc}$. An analytic solution of this dynamic system exists and is available in Bonate (2011).

IV.4.2 Bayesian model

For the 1-compartment intravenous bolus model, the concentration of Indomethacin in blood is approximated $\widetilde{x}(t) = V^{-1} (\boldsymbol{B}(t))^T \boldsymbol{c}$ using a 5-order B-spline basis function expansion with knots at each tenth between $0h$ and $8h$. The Bayesian model is:

$$\begin{cases} y_k|\boldsymbol{c}, \tau, V &\sim \mathcal{N}\left(V^{-1} (\boldsymbol{B}(t_k))^T \boldsymbol{c}; \tau^{-1}\right) \\ p(\boldsymbol{c}|\gamma, k_e, V) &\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{V_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) \\ \gamma_{11}, \tau &\sim \mathcal{G}(1; 10^{-8}), \end{cases}$$

where $\boldsymbol{c}$ are the spline coefficients, γ_{11} is the ODE-adhesion parameter, $\boldsymbol{V_1} = \boldsymbol{R}(k_e, V, \gamma_{11}) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1} \boldsymbol{\mu_c}$ (see Appendix II.A for the construction 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^{10}. The other components in $\boldsymbol{\mu_c}$ and ${\boldsymbol{\Sigma_c}}^{-1}$ are null. Note that the ODE-adhesion parameter γ_{11} is related to the variation of drug concentration in the central compartment. We parameterize the PK parameters in the log-scale with Gaussian random effect at the subject level: lke and lV are the log population parameters and τ_{lk_e} and τ_{lV} are the respective precisions of random effects. For the 2-compartment intravenous bolus model, the drug concentration in the central com-

partment is approximated by $\widetilde{x_1}(t) = V^{-1}(\boldsymbol{B}(t))^T \boldsymbol{c_1}$ and the amount of drug in the peripheral compartment is approximated by $\widetilde{x_2}(t) = (\boldsymbol{B}(t))^T \boldsymbol{c_2}$. As before, we use a 5-order B-spline basis with inner knots at each tenth between $0h$ and $8h$. The Bayesian model is:

$$\begin{cases} y_k|\boldsymbol{c_1}, \tau, V & \sim & \mathcal{N}\left(V^{-1}(\boldsymbol{B_1}(t_k))^T \boldsymbol{c_1}; \tau^{-1}\right) \\ p(\boldsymbol{c_1}, \boldsymbol{c_2}|\gamma_{21}, \gamma_{22}, k_e, k_{cp}, k_{pc}, V) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{V_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) \\ \gamma_{21}, \gamma_{22}, \tau & \sim & \mathcal{G}(1; 10^{-8}), \end{cases}$$

where $\boldsymbol{c}^T = \left(\boldsymbol{c_1}^T, \boldsymbol{c_2}^T\right)$, $\boldsymbol{V_1} = \boldsymbol{R}(k_e, k_{cp}, k_{pc}, V, \gamma_{21}, \gamma_{22}) + \boldsymbol{\Sigma_c}^{-1}$ and $\boldsymbol{v_1} = \boldsymbol{\Sigma_c}^{-1}\boldsymbol{\mu_c}$ (see Appendix II.A for the automatic construction of the penalty matrix). The first component of $\boldsymbol{\mu_c}$ is set equal to D, the (K_1+1)th component of $\boldsymbol{\mu_c}$ to 0 and the corresponding precision in $\boldsymbol{\Sigma_c}^{-1}$ to 10^{10}. The other components in $\boldsymbol{\mu_c}$ and $\boldsymbol{\Sigma_c}^{-1}$ are set to zero. Note that in the original ananlysis made in Kwan et al. (1976), the intravenous dose was not taking into account to fit a bi-exponential model. The ODE-adhesion parameter γ_{21} is related to the variation of drug concentration in the central compartment while γ_{22} is related to the variation of drug amount in the peripheral compartment.

We parameterize the PK parameters in the log-scale with Gaussian random effect at subject level: lk_e, lk_{cp}, lk_{pc} and lV are the log population parameters and τ_{lk_e}, $\tau_{lk_{cp}}$, $\tau_{lk_{pc}}$ and τ_{lV} are the respective precisions of random effects. As the drug cannot be eliminated from the periphal compartment more than it is absorbed, we also force k_{cp} to be larger than k_{pc} .

Note that taking the same prior distribution for all ODE-adhesion parameters indicates that one has the same prior confidence in each proposed system of differential equations. Please note also that working with a penalized mixture of Gaussian distributions is here unnecessary.

IV.4.3 Results

For the two models, a chain of length 10000 was run after a burn-in of 1000 iterations for all the parameters using the Metropolis-within-Gibbs algorithm described in Chapter II. Figures IV.8, IV.9 and IV.10 shows the traces and the histograms for all the population parameters. Convergence for each variable was confirmed by the Geweke's convergence diagnostic.

Model selection is made by comparing the prior and the posterior distributions of the ODE-adhesion parameters γ_{11} and γ_{21}. These two parameters are related to the variation of drug concentration in the central compartment. Figure IV.11 shows on the same graph the histogram of the MCMC chain for the ODE-adhesion parameter γ_{11} (light grey), the histogram of the MCMC chain for the ODE-adhesion parameter γ_{21} (dark grey) and their common prior

distribution on the log10 scale (black solid curve). The posterior distribution for the ODE-adhesion parameter γ_{21} is similar to its prior distribution suggesting that the 2-compartment intravenous bolus model reasonably describes the dynamics of the plasma concentration of Indomethacin. The histogram of the ODE-adhesion parameter γ_{11} (on the log10 scale) is bi-modal: one mode is centered around 7 and the other is located close to 2. This suggests that the dynamics of the concentration is only partially modeled by the 1-compartment intravenous bolus model. Table IV.4 gives the Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence for the log10 ODE-adhesion parameters $\log_{10}(\gamma_{11})$ and $\log_{10}(\gamma_{21})$. It confirms the preference for the 2-compartment model. Note also that the posterior probability that γ_{21} is larger than γ_{11} is estimated by 0.9539. It suggests to select the 2-compartment intravenous bolus model to describe the dynamics of the plasma concentration of Indomethacin.

The scatter plot of the posterior predictive residuals versus the mean predictive concentrations does not clearly indicate any trend (Figure IV.12). The two outliers are for subject 2. Figure IV.13 gives the pointwise 80% and 95% quantile of the fitted individual plasma concentration of Indomethacin: the fitted concentrations are in close agreement with the observed ones. Figure IV.14 gives the pointwise 80% and 95% quantile for the fitted individual peripheral concentration of drug: the maximum peripheral concentration is reached after one hour. Figure IV.15 provides the pointwise 80% and 95% quantile of the predictive plasma concentration of Indomethacin for each subject. Only two observed drug concentrations (for subject 2) are outside the 95% predictive credibility interval which is compatible with the expected coverage.

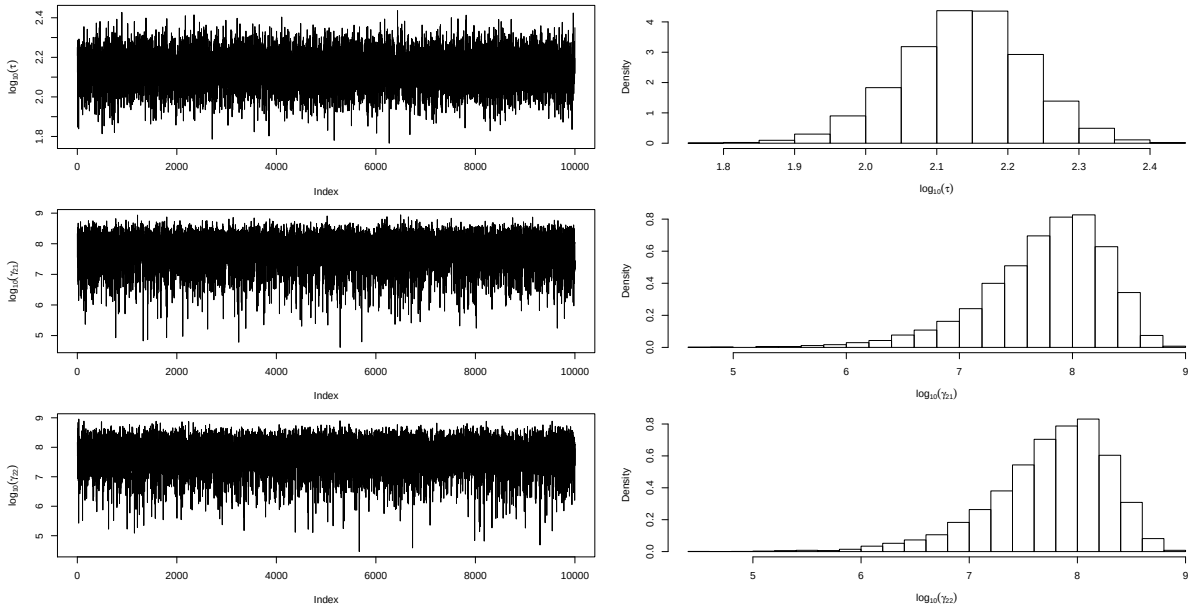


Figure IV.8 – Traces and histograms for $\log_{10}(\tau)$, $\log_{10}(\gamma_{21})$ and $\log_{10}(\gamma_{22})$.

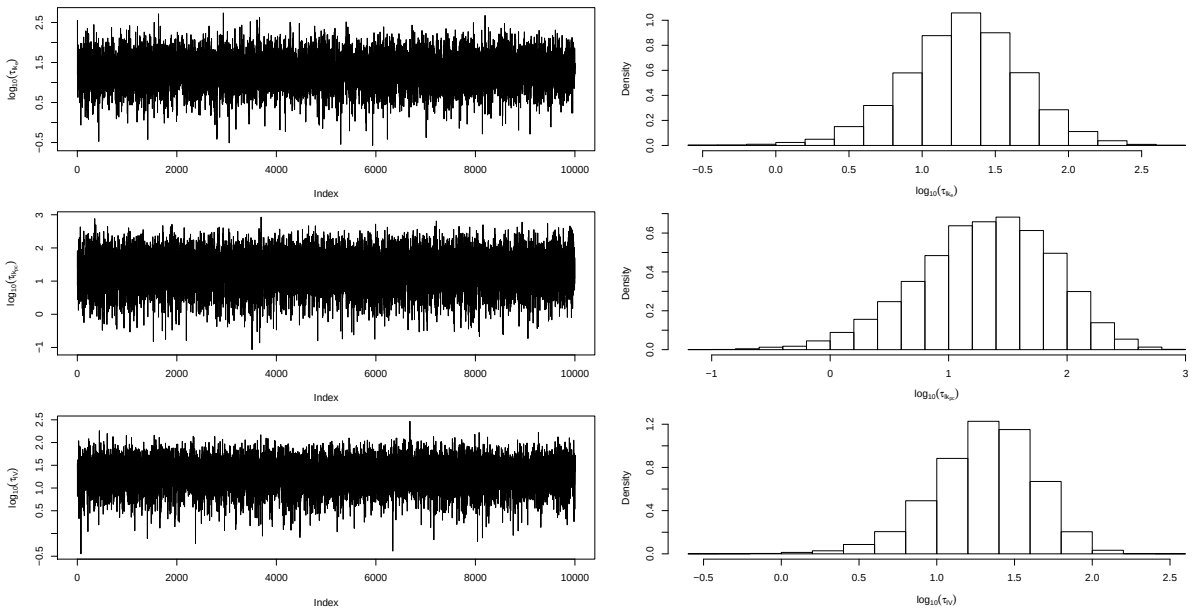


Figure IV.9 – Traces and histograms for $\log_{10}(\tau_{lk_e})$, $\log_{10}(\tau_{lk_{pc}})$ and $\log_{10}(\tau_{lV})$

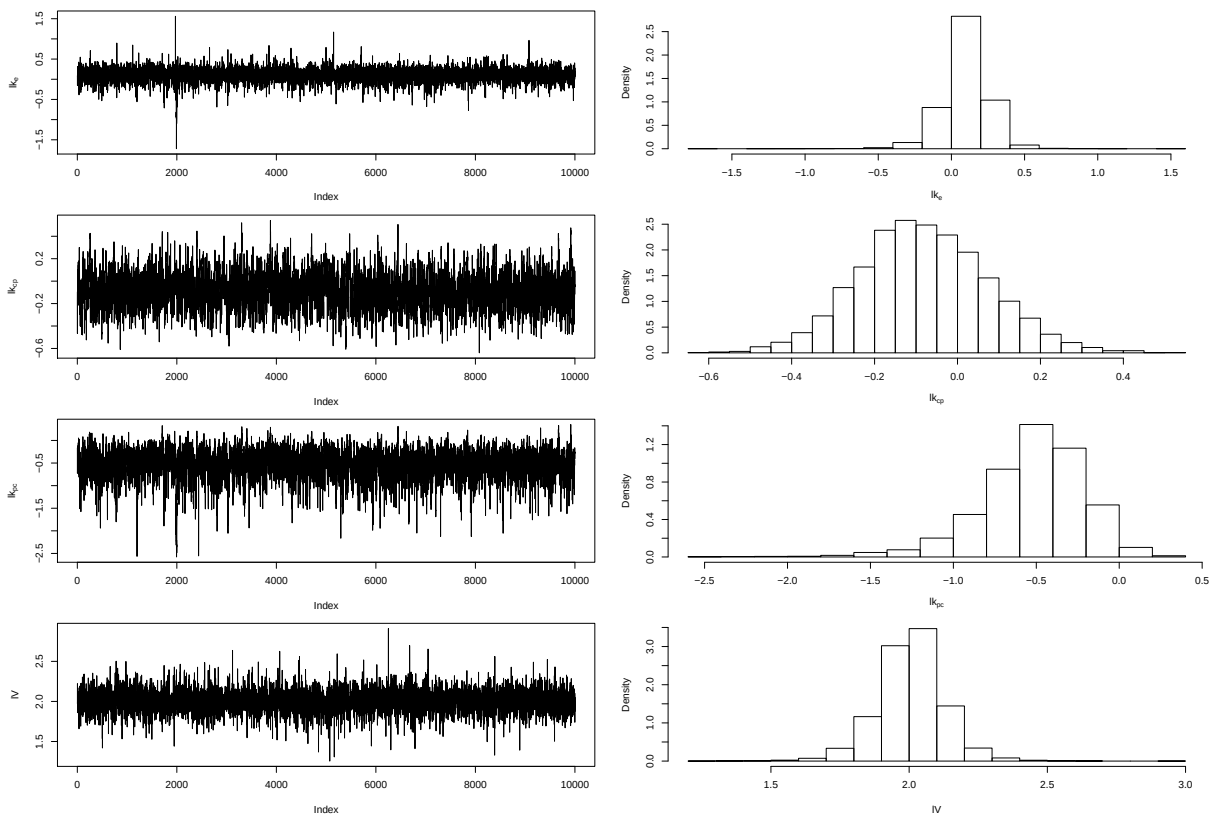


Figure IV.10 – Traces and histograms for $\log(k_e)$, $\log(k_{cp})$, $\log(k_{pc})$ and $\log(V)$.

Table IV.4 – Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence for the log10 ODE-adhesion parameters $\log_{10}(\gamma_{11})$ and $\log_{10}(\gamma_{21})$.

	Model	
Divergence measure	1-compartment IV bolus	2-compartment IV bolus
Kolmogorov-Smirnov	7.62e-1	1.07e-2
Cramér-von Mises	2.56e+3	2.48e-1
Anderson-Darling	4.76e+4	1.46
Kullback-Leibler	6.03	2.98e-3

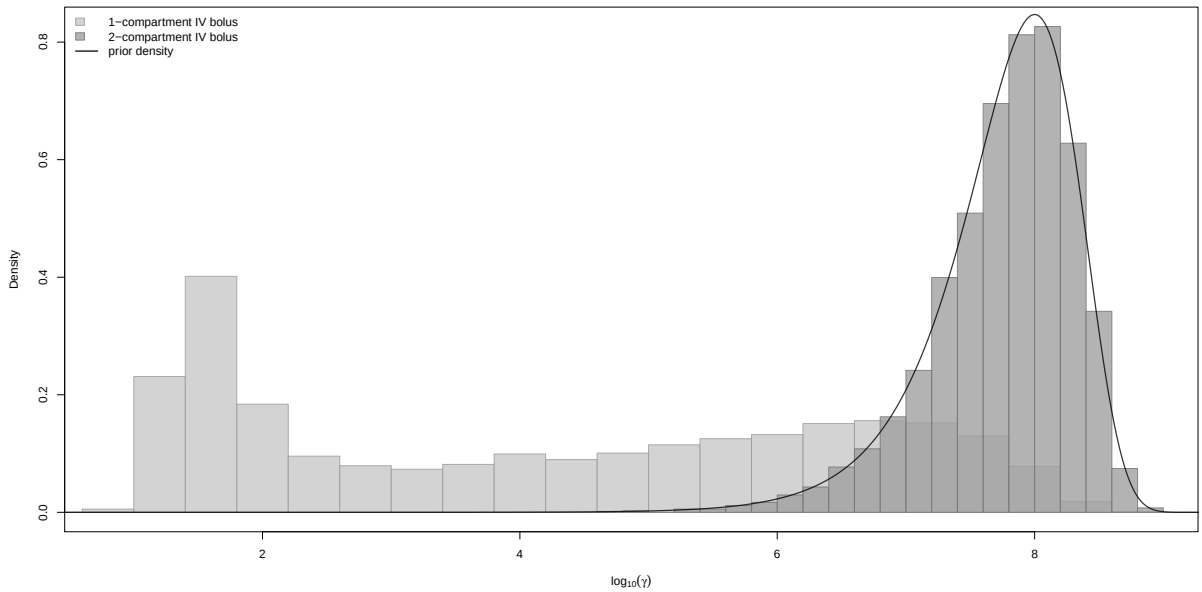


Figure IV.11 – Histogram based on a sample of the posterior distribution for the ODE-adhesion parameter γ_{11} (light gray, log10 scale) in the 1-compartment intravenous bolus model, histogram based on a sample of the posterior distribution for the ODE-adhesion parameter γ_{21} (dark gray, log10 scale) in the 2-compartment intravenous bolus model and the common prior distribution for the corresponding ODE-adhesion parameter (black solid line).

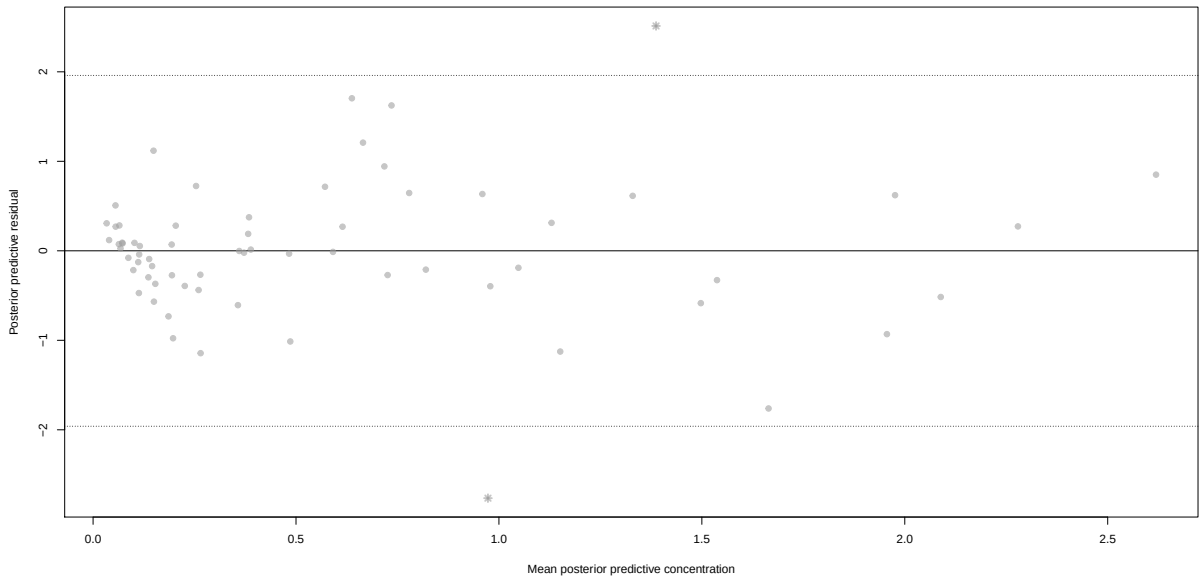


Figure IV.12 – Scatter plot of posterior predictive residuals versus mean predictive concentrations. Outliers are represented by stars.

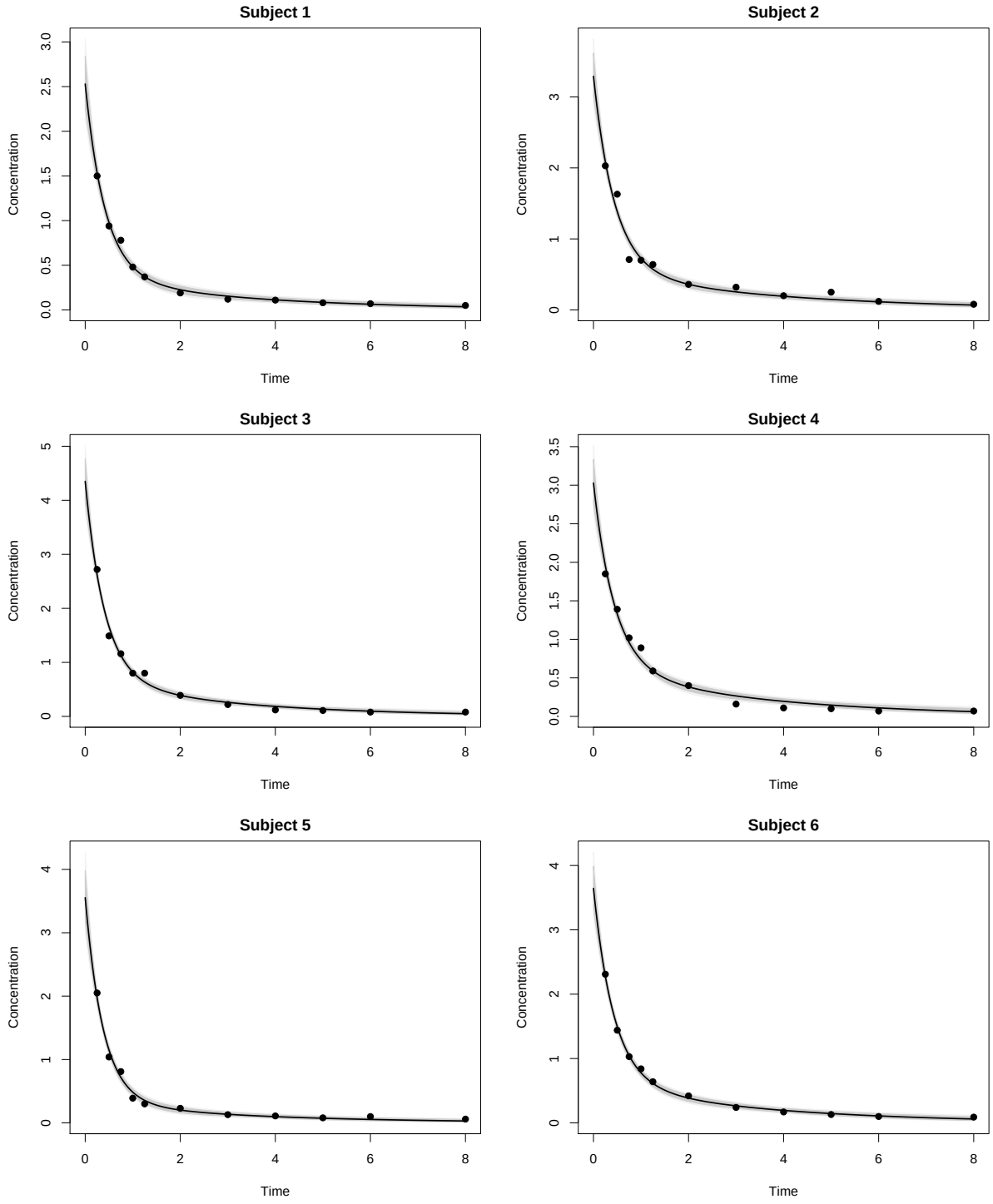


Figure IV.13 – Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted plasma concentration of Indomethacin with pointwise median (black solid curve) and observed concentration of Indomethacin (solid circles) versus time for the six subjects.

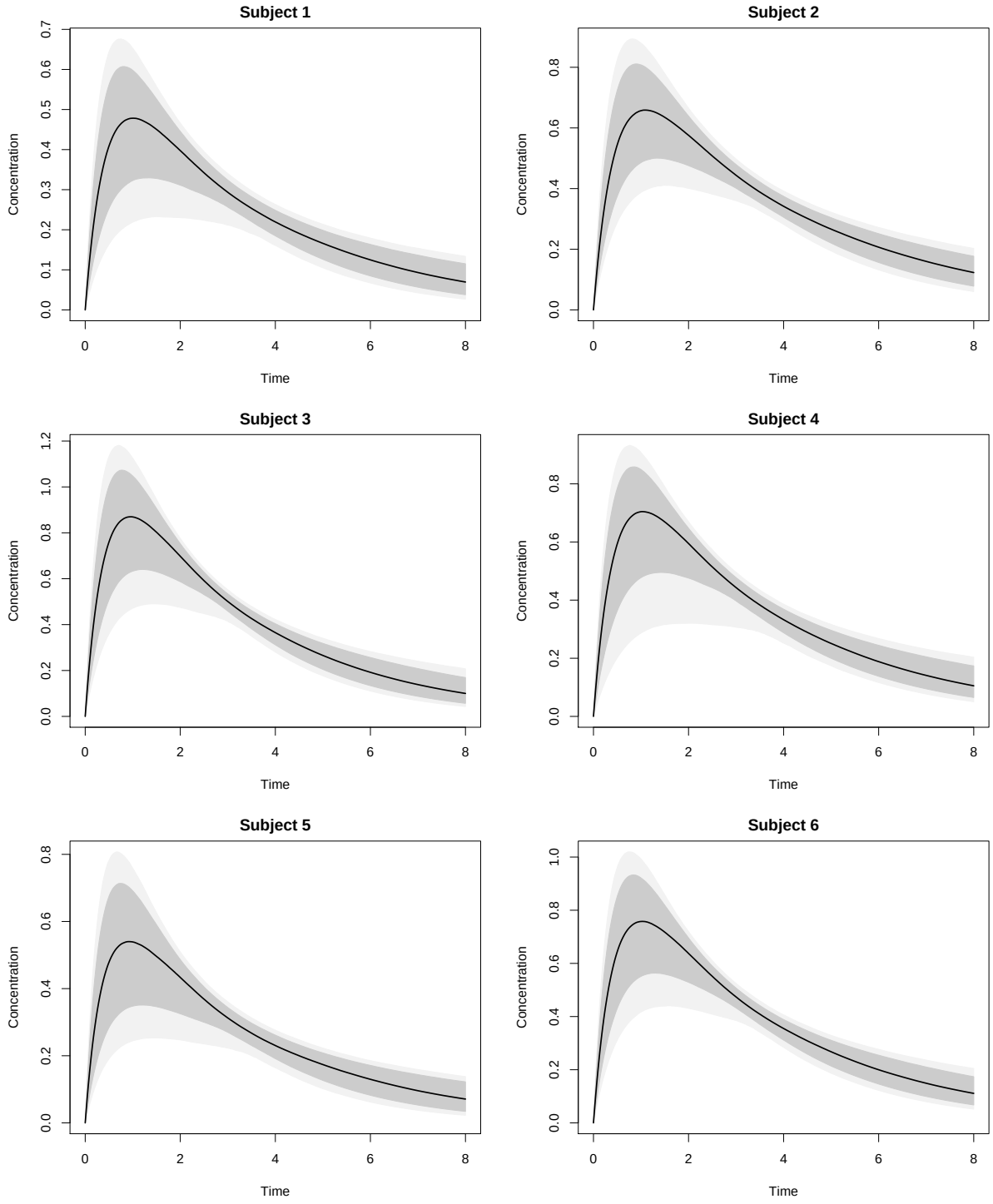


Figure IV.14 – Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted peripheral concentration of Indomethacin with pointwise median (black solid curve) versus time for the six subjects.

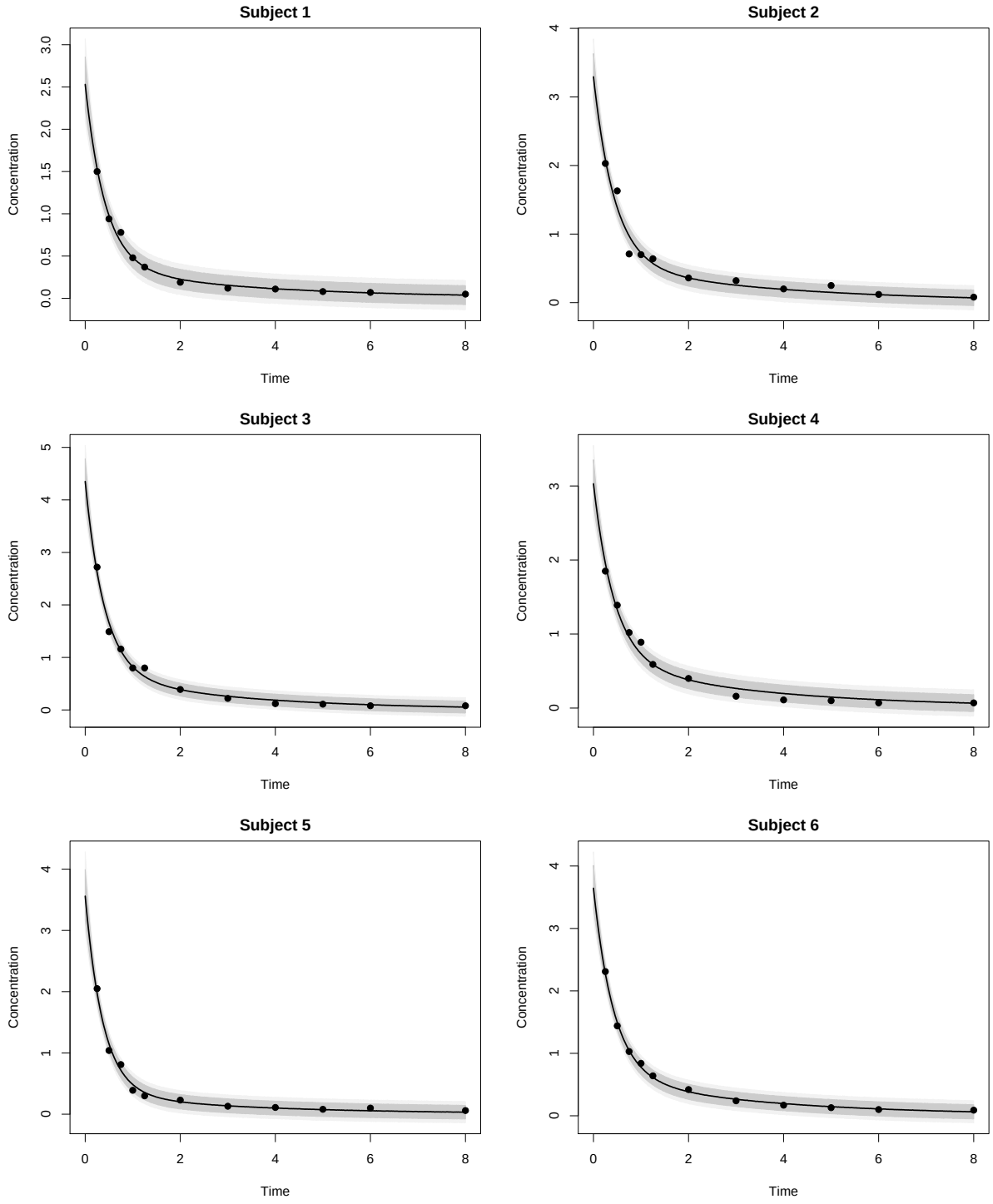


Figure IV.15 – Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the plasma concentration of Indomethacin with pointwise posterior predictive median (black solid curve) and observed concentration of Indomethacin (solid circles) versus time for the six subjects.

IV.5 Discussion

We have presented a strategy based on Bayesian ODE-penalized B-spline approach where the suitability of a system of differential equations to model the dynamics of a state function is assessed and quantified by the comparison of the posterior distribution for the ODE-adhesion parameter with its prior. To our best knowledge, it is the first time that penalty parameters in a smoothing approach are used to assess the suitability of a model.
This Bayesian ODE-penalized B-spline approach also enables to select competing ODE-models by comparing the posterior distributions of the ODE-adhesion parameters. The ODE-model selection was feasible in the frequentist ODE-penalized B-spline approach, but compared to the Bayesian approach, uncertainty measure about the selection is not available and the adjusted/selected value of the ODE-adhesion parameter does not statistically validate or discard the proposed ODE model.
The comparison of the posterior distribution for the ODE-adhesion parameter with its prior is either made through visual comparison between the histogram of the posterior distribution and the prior density function or by the use of divergence measures. Simulations confirm the suitability of ODE-adhesion parameters to select ODE models. It has also highlighted that the Anderson-Darling statistic and the Kullback-Leibler divergence are more likely to detect a difference between the prior and the posterior distribution for the ODE-adhesion parameter.
We recommend to combine that strategy with other established analysis such as residual inspection.

Concluding comments

The aim of the thesis was to introduce a Bayesian smoothing approach for the resolution of systems defined by differential equations. The proposed methodology can either be viewed as a generalization in a Bayesian setting of the approach proposed in Ramsay et al. (2007) or as an adaptation of Bayesian P-splines (Lang and Brezger, 2004).
This approach is based on an approximation of the state function using a linear combination of the elements of a B-spline basis, combined with a prior distribution for the spline coefficients that counter-balances the flexibility of the fitted curves/trajectories using a penalty term related to the system of differential equations under consideration.

Main contributions

This thesis focuses on two main points: the development of a fully Bayesian approach for making inference in models specified using systems of affine differential equations when the error distribution is homogeneous (Gaussian or mixture of Gaussian/Student distributions) and the validation/selection of systems of differential equations. The following contributions have been introduced in this thesis:

- Chapter II proposes a fully Bayesian model based on ODE-penalized B-spline approach to jointly estimate the state functions, the ODE parameters and the precision of measurements when the ODE model is affine and the conditional distributions of the responses are assumed Gaussian.
  The method was also extended to deal with hierarchical studies where, for example, systems are observed on multiple subjects. This hierarchical extension is not available for

the frequentist proposal made by Ramsay et al. (2007).
Our study of the Bayesian ODE-penalized B-spline model highlighted the important role played by the selected prior distribution for the spline coefficients and the necessity to include the complete normalizing constant of that prior in the calculation of the joint posterior.
The proposed model formulation enables to include some prior information about the ODE parameters and the precision of measurements. It is also possible to include prior uncertainty with respect to the initial condition of the state functions.
In contrast to the frequentist approach, the selection of the ODE-adhesion parameters is automatic and is intrinsically included in the model.

- The case where the data distribution is homogeneous but unknown was considered in Chapter III. There, the non-gaussianity was managed either through a penalized mixture of Gaussian distributions or through a penalized mixture of Student distributions.
  The combination of the Bayesian ODE-penalized B-spline approach with penalized mixture of Gaussian or Student distributions has the advantage of not forcing the user to specify a particular distribution for the error terms.
  Such assumptions on the error distributions are convenient as conditionally on latent variables, one is back to the standard setting assumed in Chapter II.
  Of course, such specifications can be used in any problem where an unknown homogeneous distribution is involved, like for example in semi- and nonparametric regression models.

- Chapter IV focuses on the ODE-adhesion parameter and its ability to help in the validation or the selection of ODE models. To our best knowledge, it is the first time that penalty parameters are used in that way.
  Based on the comparison between the prior and the posterior distribution of the ODE-adhesion parameters, it is possible to confirm or deny that the dynamics of the observed state functions is driven by a set of differential equations. This comparison is made using a visual inspection whereas selection is performed either using divergence measures or posterior probabilities. This validation of systems of differential equations was not available in the frequentist ODE-penalized B-spline approach.
  Furthermore, when selecting competing ODE models, uncertainty measure with respect to the selection of one model against competitors, is available in a Bayesian setting.

Further research

Many challenges / difficulties remain to be overcome. They can be pooled into six categories:

- The first challenge focuses on the limitations of the Gaussian or penalized mixture of Gaussians assumption for the distribution of error terms: the conditions of existence of the state functions are not necessarily respected (e.g. positiveness). Alternative methods should be developed while keeping the advantages of the marginalization of the joint posterior distribution with respect to the spline coefficients. Observation models could be investigated. There, the observations $\{y_{jk}; j \in \mathcal{J}; k = 1, \ldots, n_j\}$ are related to the state functions using observation functions $\{h_j(\cdot); j \in \mathcal{J}\}$ and some measurement error $\{\epsilon_{jk}; j \in \mathcal{J}; k = 1, \ldots, n_j\}$:
$$y_{jk} = h_j(\boldsymbol{x}(t_{jk})) + \epsilon_{jk}.$$

- The case of homogeneous distribution for the error terms is also too restrictive. Two strategies can be considered to manage this issue.

  - The first one would be to express the precision of measurement as a smooth function of time. The log of the precision of measurement could be expressed as a linear combination of B-splines with the flexibility counter-balanced by a roughness penalty computed based on the integral of the square of the first, second or third order derivative. When only one state function is observed, the model specification for the precision of measurement could be given by:
$$\begin{cases} \log(\widetilde{\tau}(t)) &= (\boldsymbol{B_\tau}(t))^T \boldsymbol{c_\tau} \\ p(\boldsymbol{c_\tau}|\gamma_\tau) &\propto \exp\left(-\frac{\gamma_\tau}{2} PEN_\tau\right). \end{cases}$$
where is $\boldsymbol{B_\tau}(t)$ the vector of B-spline basis functions evaluated at time t, $\boldsymbol{c_\tau}$ is the vector of spline coefficients, PEN_τ is the roughness penalty term and γ_τ the roughness penalty parameter. This generalization appears quite manageable, in the sense that the conditional posterior distribution for the spline coefficients used for the approximation of the state functions will be again a multivariate Gaussian distribution. Thanks to that, the joint posterior distribution still can be marginalized with respect to these spline coefficients. The order of the penalty must also be selected in the penalized B-spline model for the precision. The first order would be preferred when testing homoscedasticity, while 2nd or 3rd orders would provide more flexibility.

  - A second strategy would directly relate the variance of the error terms to the corresponding state function. In this case, the joint posterior distribution cannot be easily marginalized with respect to these spline coefficients. Therefore, new procedures have to be put in place to manage the posterior correlations between the ODE parameters and the spline coefficients.

- The actual procedure for the validation / selection of systems of differential equations using the ODE-adhesion parameters have revealed some limitations:

  - The actual recommended prior distribution for the ODE-adhesion parameters (i.e. a gamma distribution with shape equal to one and rate smaller than 10^{-6}) seems too restrictive for small sample sizes. In such a setting, the posteriors of these parameters cannot differ for their priors even if the proposed ODE model is not adapted to the dynamic of the state functions. The proposal made by Jullion and Lambert (2007) is a possible starting point. There, the lack of flexibility in a standard Bayesian P-spline model was countered either by the specification of a hyperprior distribution for the rate of the gamma prior distribution for the ODE-adhesion parameter or by using a mixture of gamma distributions with equidistant rates on the log10 scale.

  - The ODE-validation is made by a visual inspection. More sophisticated methods could be developed.
    One possibility could be to consider a mixture of two gamma distributions for the prior distribution for the ODE-adhesion parameter with one component of the mixture that expresses the prior confidence in the proposed ODE model and the other one that expresses no prior belief in the ODE model. Even if no prior belief on the ODE model is introduced on a vague gamma prior distribution, some smoothing is nevertheless performed, which leads to a posterior distribution for that ODE-adhesion parameter that significantly differs from its vague prior.
    One other possibility could be to look after the effective number of parameters. Indeed, the larger the penalty on the spline coefficients, the smaller the effective number of parameters. However, in order to take a decision, one will need to select a critical value for the number of parameters under which the ODE-model will be declared to be a valid descriptor of the dynamics of the state functions.

- Before collecting the data, attention should be given to the experimental design to avoid poor parameter estimation and/or avoid ineffective model-validation (Retout and Mentré, 2003). Optimal design strategies based on the Fisher Information matrix have to

be extended to the (frequentist) ODE-penalized B-spline approach. There, the Fisher Information matrix is

$$\boldsymbol{M_F}(\boldsymbol{\psi}, \boldsymbol{\xi}) = \mathbb{E}\left(-\frac{\partial^2 J(\boldsymbol{\psi}, \boldsymbol{y})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^T}\right)$$

where J is the log-penalized likelihood defined in Eq. (II.3), $\boldsymbol{\psi}^T = \left(\boldsymbol{\theta}^T, \boldsymbol{\tau}^T, \boldsymbol{\gamma}^T, \boldsymbol{c}^T\right)^T$ is the vector of all the parameters and $\boldsymbol{\xi}$ represents the design, i.e. the vector of all the times of measurement. Traditionally, the determinant of the Fisher information matrix (D-optimality criterion, see Wu (1978) for other optimality criteria) is optimized with respect to $\boldsymbol{\xi}$ for a given prior value of the vector of parameters $\boldsymbol{\psi}$. Reasonable prior value are usually easily to select for the ODE parameter $\boldsymbol{\theta}$ and for the precisions of measurement $\boldsymbol{\tau}$. For the ODE-adhesion parameters $\boldsymbol{\gamma}$, large values will indicate our prior confidence in the ODE model to further describe the dynamic of the upcoming data. But for the spline coefficients, prior values are not easy to select. Therefore, we recommend to start from the marginal likelihood after integration of the spline coefficients. Note that in contrast to existing approaches, the ODE-penalized B-spline approach won't require a Taylor expansion.

- The Mathematical models are not exact. Therefore, some modelling discrepancies have to be part of the model (Varziri et al., 2008). There, modelling errors are considered using system noise $\boldsymbol{\eta}(t)$ in the dynamic model:

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{f}(\boldsymbol{x}, t, \boldsymbol{\theta}) + \boldsymbol{\eta}(t),$$

  where the covariance matrix of the system noise is supposed to be known. The Bayesian ODE-penalized B-spline approach should be extended to this stochastic differential equations case: the differential equation operators used for the penalty terms are in fact just the system noise.

- The actual proposed Bayesian approach was focused on the affine ODE case. It enabled us to highlight the fundamental role of the normalization constant

$$K(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \int p(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\gamma})\, d\boldsymbol{c}$$

  in the prior distribution of the spline coefficients. There, this constant has an explicit expression (see Section II.3.1). For nonlinear systems of differential equations, explicit expressions for the ODE-penalty term and for the normalization constant will not be

available anymore. Numerical approximations of these integrals will be required.

List of Symbols

$\mathbb{R}$	Set of real numbers
D^l	l-order differential operator
$\mathcal{N}(\mu; \tau^{-1})$	Gaussian distribution with mean μ and precision (inverse-variance) τ
$\mathcal{N}_p\left(\boldsymbol{\mu}; \boldsymbol{P}^{-1}\right)$	Mulivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and precision matrix (inverse-covariance matrix) $\boldsymbol{P}$
$t(\nu; \mu; \sigma^2)$	Noncentral Student distribution with ν degrees of freedom, mean μ and variance σ^2
$\mathcal{G}(a; b)$	Gamma distribution with shape equal to a and rate equal to b
$\mathcal{W}_q(\boldsymbol{V}; s)$	Wishart distribution with s degrees of freedom and scale matrix V
$\log$	Natural logarithm
$\log_{10}$	Base 10 logarithm
$\text{vec}(\boldsymbol{z_k}; k = 1, \ldots, K)$	Concatenation of vectors $\boldsymbol{z_k}$, $k = 1, \ldots, K$
$\text{diag}(\boldsymbol{Z_k}; k = 1, \ldots, K)$	Block diagonal matrix with $\boldsymbol{Z_k}$ on the kth diagonal block
$\propto$	Proportionality up to a multiplicative constant
$\doteq$	Equality up to an additive constant

List of Figures

I.1 Ischemic/non-ischemic perfution ratio. 7

I.2 Pharmacokinetic profiles for the Indomethacin dataset. 8

I.3 1-compartment intravenous bolus model. 9

I.4 2-compartment intravenous bolus model. 9

I.5 Pharmacokinetic profiles for the Theophylline dataset. 10

I.6 Oral 2-compartment model. 11

I.7 Error propagation between the true concentration (solid grey curve) and numerical approximation (using four order Runge-Kutta algorithm, dashed black curve) in a oral 2-compartment model after periodical administration (dotted light gray vertical lines). 14

II.1 Cubic B-splines (solid curves) with inner knots (dotted grey vertical lines) at each tenth between $[0; T]$. 25

II.2 Influence of the number of functions in the B-spline basis. Left part: knots at each half-integer are placed between 0 and 2. Right part: knots at each twentieth are placed between 0 and 2. Grey dotdashed curve corresponds to the true curve, black solid curve corresponds to the B-spline approximation and the dark grey solid curves correspond to the B-spline basis function times the corresponding spline coefficient. 27

II.3 Influence of the ODE-adhesion parameter on the B-spline approximation. Grey dotdashed curve corresponds to the true curve, black solid curve corresponds to the B-spline approximation and the dark grey solid curves correspond to the B-spline basis function times the corresponding spline coefficient. Value of the ODE-adhesion parameter is given on the top right corner. 28
II.4 Posterior correlation between spline and PK parameters in an oral 2-compartment model: scatterplots of parameter samples generated from the joint posterior using MCMC. 34
II.5 Scatter plot between samples from the posterior distribution for the ODE parameters after the analysis of a simulated dataset (oral 2-compartment model) 35
II.6 1-compartment intravenous bolus model - Mean posterior variance of k_e (top figures) and τ (bottom figures) with respect to $-\log_{10}(b_\gamma)$ with sample size equal to 10, 15, 20, 50, 100 and 200. 43
II.7 1-compartment intravenous bolus model - Average 95% HPD interval width of k_e (top figures) and τ (bottom figures) with respect to $-\log_{10}(b_\gamma)$ with sample size equal to 10, 15, 20, 50, 100 and 200. 44
II.8 Traces and histograms for σ, $\log_{10}(\gamma)$ and $\log_{10}(\tau_{l\beta})$ 62
II.9 Traces and histograms for $l\alpha = \log(\alpha)$ and $l\beta = \log(\beta)$ 62
II.10 Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted perfusion ratio with pointwise median (black solid curve) and observed perfusion ratio (solid circles) versus time for the twelve mice. . . 63
II.11 Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the perfusion ratio with pointwise posterior predictive median (black solid curve) and observed perfusion ratio (solid circles) versus time for the twelve mice. 64
II.12 Square root of the determinant of the penalty matrix versus the ODE parameter θ (log and original scale). 72

III.1 Top: Simulated data from an exponential decay model (stars correspond to outliers). Bottom left : fitted error density using the mixture of Gaussian distributions. Bottom right: fitted error density using the mixture of Student distribution. On bottom figures, solid curves are pointwise posterior median for the fitted error density, grey regions delimit pointwise 80% and 95% credibility intervals and the dashed curve corresponds to the target standard normal distribution. 91

III.2 Estimated error density functions for sample size (50, 100, 200, 500) (one light grey curve per dataset) with pointwise average of these estimated densities in dark grey dashed curve. The black solid curve is the true density function and the black dotted curve is the standardized Gaussian density. 98

III.3 Fitted error density (solid curve is the pointwise posterior median and grey regions delimit pointwise 80% and 95% credible intervals), dotted curve is the Gaussian density. 101

III.4 Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted amount of Theophylline in stomach with pointwise median (black solid curve) versus time for the twelve subjects. 102

III.5 Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted concentration of Theophylline with pointwise median (black solid curve) and observed concentration of Theophylline (solid circles) versus time for the twelve subjects. 103

III.6 Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the concentration of Theophylline with pointwise posterior predictive median (black solid curve) versus time for the twelve subjects. 104

IV.1 Sensitivity to the sample size - Boxplot of posterior empirical median of $\log_{10}(\gamma)$ for the seven sample size with prior median (horizontal dashed line). 119

IV.2 Sensitivity to the sample size - Boxplot of estimated Kullback-Leibler divergence (on the log10 scale) for the seven sample size. 120

IV.3 Sensitivity to a wrong specification of the ODE parameter - Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence (on the log10 scale) when $n = 100$ and k_e is set equal to 3, 4, 5, 6 or 7 in the MCMC procedure (while $k_e = 5$ was used to generate datasets). 121

IV.4 Sensitivity to a wrong specification of the ODE parameter - Histograms based on a sample of the posterior distribution for the ODE-adhesion parameter when $n = 100$ and k_e is equal to 4 (light gray histogram) or 5 (dark grey histogram) while $k_e = 5$ was used to generate datasets. 122

IV.5 Sensitivity to a wrong specification of the precision of measurement - Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence (on the log10 scale) when $n = 100$ and $\log_{10}(\tau)$ is set equal to 1, 1.5, 2, 2.5, 2.75 or 3 in the MCMC procedure (while $\tau = 100$ was used to generate datasets). 123

IV.6 Sensitivity to a misspecification of the error distribution - Boxplot of estimated Kullback-Leibler divergence (on the log10 scale) for sample size equals to 100 with four error distributions (Gaussian, Student, extreme value and mixture of two Gaussians) used for the data generation. Gaussian distribution used for the analysis of each simulated dataset. 124

IV.7 Sensitivity to ODE-model misspecification - Boxplot of the Kullback-Leibler divergence (on the log10 scale) estimated from data generated using a bi-exponential decay model while an exponential decline one was (wrongly) assumed in the analysis. Figure on the right is a zoom for large sample sizes. . . 126

IV.8 Traces and histograms for $\log_{10}(\tau)$, $\log_{10}(\gamma_{21})$ and $\log_{10}(\gamma_{22})$. 130

IV.9 Traces and histograms for $\log_{10}(\tau_{lk_e})$, $\log_{10}\left(\tau_{lk_{pc}}\right)$ and $\log_{10}(\tau_{lV})$ 130

IV.10 Traces and histograms for $\log(k_e)$, $\log(k_{cp})$, $\log(k_{pc})$ and $\log(V)$. 131

IV.11 Histogram based on a sample of the posterior distribution for the ODE-adhesion parameter γ_{11} (light gray, log10 scale) in the 1-compartment intravenous bolus model, histogram based on a sample of the posterior distribution for the ODE-adhesion parameter γ_{21} (dark gray, log10 scale) in the 2-compartment intravenous bolus model and the common prior distribution for the corresponding ODE-adhesion parameter (black solid line). 132

IV.12 Scatter plot of posterior predictive residuals versus mean predictive concentrations. Outliers are represented by stars. 132

IV.13 Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted plasma concentration of Indomethacin with pointwise median (black solid curve) and observed concentration of Indomethacin (solid circles) versus time for the six subjects. 133

IV.14 Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted peripheral concentration of Indomethacin with pointwise median (black solid curve) versus time for the six subjects. 134

IV.15 Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the plasma concentration of Indomethacin with pointwise posterior predictive median (black solid curve) and observed concentration of Indomethacin (solid circles) versus time for the six subjects. . . . 135

List of Tables

II.1 1-compartment intravenous bolus model - $\sigma = 0.1$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for FGPE, FES, BGPE, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values. 49

II.2 1-compartment intravenous bolus model - $\sigma = 0.3$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for FGPE, FES, BGPE, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values. 49

II.3 1-compartment intravenous bolus model - $\sigma = 0.5$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for FGPE, FES, BGPE, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values. 50

II.4 1-compartment intravenous bolus model - Average 80%-HPD interval width (P1), average 95%-HPD interval width (P2) and average posterior variance (P3) for B-ODE-P-spline and BES methods for the three different levels of standard deviation σ. 51

II.5 Oral 2-compartment model - $\sigma = 0.3$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for F-ODE-P-spline, FES, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values. 55

II.6 Oral 2-compartment model - $\sigma = 0.6$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for F-ODE-P-spline, FES, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values. 56

II.7 Oral 2-compartment model - $\sigma = 0.9$ - Relative bias in percent (RB), estimated coverage probability (CP) at level 80% (CP80%), 95% (CP95%) and root mean squared error (RMSE) for F-ODE-P-spline, FES, B-ODE-P-spline and BES methods. Estimated coverage probabilities in bold characters are compatible with their nominal values. 57

II.8 Oral 2-compartment model - Average 80%-HPD interval width (P1), average 95%-HPD interval width (P2) and average posterior variance (P3) for B-ODE-P-spline and BES methods for the three different levels of standard deviation σ. 58

II.9 Estimated parameters with credibility sets 65

III.1 Simulation - $n = 50$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values. 94

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

III.3 Simulation - $n = 200$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values. 96

III.4 Simulation - $n = 500$ - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for k_a, k_e, V and τ with four standardized error distribution (Gaussian, Student, extreme value and mixture of two Gaussians) using B-ODE-P-spline-PGM approach (PGM) (and B-ODE-P-spline-G approach (G) for the Gaussian error distribution case). Estimated coverage probabilities in bold are not in agreement with their nominal values. 97

III.5 Simulation - Student distribution with 6 degrees of freedom - Relative bias in percent (RB), estimated coverage probability at level 80% (CP80), 95% (CP95), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for θ and τ for five sample sizes ($n = 20, 50, 100, 200, 500$) using B-ODE-P-spline approach with Student distribution for the error term. Estimated coverage probabilities in bold are in agreement with their nominal values. 109

IV.1 Sensitivity to a wrong specification of the ODE parameter - Proportion of divergence measures (Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence) larger than the one estimated with $k_e = 5$ when $n = 100$ with k_e equals to 3, 4, 6 or 7 in the MCMC procedure. 121

IV.2 Sensitivity to a wrong specification of the precision of measurement - Proportion of divergence measures (Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence) larger than the one estimated with $\tau = 100$ when $n = 100$ with $\log_{10}(\tau)$ equals to 1, 1.5, 2, 2.5, 2.75 or 3 in the MCMC procedure. 123

IV.3 Sensitivity to a misspecification of the error distribution - Quantile at 5%, 10%, 50%, 90% and 95% of the log10 estimated Kullback-Leibler divergence for sample size equals to 100 with four error distributions (Gaussian, Student, extreme value and mixture of two Gaussians) used for the data generation. A Gaussian distribution was wrongly assumed in the analysis of each simulated dataset. 125

IV.4 Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence for the log10 ODE-adhesion parameters $\log_{10}(\gamma_{11})$ and $\log_{10}(\gamma_{21})$. 131

References

E Baake, M Baake, H Bock, and K Briggs. Fitting ordinary differential equations to chaotic data. *Physical Review A*, 45:5524–5529, 1992. doi: http://dx.doi.org/10.1103/PhysRevA.45.5524. 14

L Biegler. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers & Chemical Engineering*, 8:243–247, 1984. doi: http://dx.doi.org/10.1016/0098-1354(84)87012-X. 16

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. 12

P Bonate. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer US, 2011. doi: http://dx.doi.org/10.1007/978-1-4419-9485-1. 1, 8, 10, 127

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. 34, 47

D Campbell. *Bayesian collocation tempering and generalized profiling for estimation of parameters from differential equation models*. PhD thesis, McGill University, 2007. 19, 20, 29, 30, 45, 46, 48, 52, 70, 71

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. 24

M Davidian and D Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, 1995. 10

C de Boor. *A Practical Guide to Splines*. Springer-Verlag New York Inc, 2001. 26

J Diebolt and C Robert. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2):363–375, 1994. URL `http://www.jstor.org/stable/2345907`. 78

J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. Oxford University Press, 2001. 15

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. 24

A Gelman and D Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992. 36, 60

S Geman and D Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984. doi: http://dx.doi.org/10.1080/02664769300000058. 73

D Griffiths and D Higham. *Numerical Methods for Ordinary Differential Equations: Initial Value Problems*. Springer Undergraduate Mathematics Series. Springer, 2010. doi: http://dx.doi.org/10.1007/978-0-85729-148-6. 13

H Haario, E Saksman, and J Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2): 223–242, 2001. URL `http://www.jstor.org/stable/3318737`. 73, 83

W Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: http://dx.doi.org/10.1093/biomet/57.1.97. 73

J Jaeger and P Lambert. Bayesian generalized profiling estimation in hierarchical linear dynamic systems. Discussion Paper DP2011/01, Institut de Statistique, Biostatistique et Sciences Actuarielles, 2011. 3, 116

J Jaeger and P Lambert. Bayesian penalized smoothing approaches in models specified usingaffine differential equations with unknown error distributions. Discussion Paper DP2012/17, Institut de Statistique, Biostatistique et Sciences Actuarielles, 2012a. 3

J Jaeger and P Lambert. On the use of adhesion parameters to validate models specified using systems of affine differential equations. Discussion Paper DP2012/18, Institut de Statistique, Biostatistique et Sciences Actuarielles, 2012b. 3

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. 30, 140

B.W. Kernighan and D.M. Ritchie. *The C Programming Language*. Prentice-Hall Software Series. Prentice Hall, 1988. ISBN 9780131103627. 36

A Komárek and E Lesaffre. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, 103:523–533, 2008. doi: http://dx.doi.org/10.1198/016214507000000563. 77, 79

A Komárek, E Lesaffre, and C Legrand. Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*, 26(30):5457–5472, 2007. ISSN 1097-0258. doi: 10.1002/sim.3083. URL `http://dx.doi.org/10.1002/sim.3083`. 78

S Kullback and R Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951. URL `http://www.jstor.org/stable/2236703`. 118

K Kwan, G Breault, E Umbenhauer, F McMahon, and D Duggan. Kinetics of indomethacin absorption, elimination, and enterohepatic circulation in man. *Journal of Pharmacokinetics and Pharmacodynamics*, 4(3):255–280, 1976. ISSN 1567-567X. doi: http://dx.doi.org/10.1007/BF01063617. 8, 126, 128

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. 35, 84

P Lambert. Nonparametric additive location-scale models for interval censored data. *Statistics and Computing*, 2011. doi: http://dx.doi.org/10.1007/s11222-011-9292-6. 80

P Lambert and P Eilers. Bayesian density estimation from grouped continuous data. *Computational Statistics & Data Analysis*, 53(4):1388–1399, 2009. doi: http://dx.doi.org/10.1016/j.csda.2008.11.022. 73, 83

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. 19, 24, 29, 137

H Liang and H Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statitistical Association*, 103(484):1570–1583, 2008. 16

J K Lindsey, J Wang, B Jones W D Byrom, and I D Hind. Simultaneous modelling of flosequinan and its metabolite. *European Journal of Clinical Pharmacology*, 55:827–836, 2000. 76, 112

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. 1, 14

N Metropolis, A Rosenbluth, M Rosenbluth, Augusta A Teller, and E Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: http://dx.doi.org/10.1063/1.1699114. 73

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(4):698–708, 2006. doi: http://dx.doi.org/10.1016/j.compchemeng.2005.11.008. 2, 18, 19, 24

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org`. ISBN 3-900051-07-0. 36

J Ramsay. Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3):495–508, 1996. URL `http://www.jstor.org/stable/2345889`. 17

J Ramsay and B Silverman. *Functional Data Analysis*. Springer Series in Statistics, 2005. 2, 26

J 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 (Statistical Methodology)*, 69(5):741–796, 2007. doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00610.x. 2, 18, 19, 20, 24, 28, 45, 46, 66, 75, 111, 137, 138

S Retout and F Mentré. Optimization of individual and population designs using splus. *Journal of Pharmacokinetics and Pharmacodynamics*, 30(6):417–443, 2003. doi: http://dx.doi.org/10.1023/B:JOPA.0000013000.59346.9a. 140

S Richardson and P Green. On Bayesian analysis of mixture with an unknown number of components. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(4):731–792, 1997. URL `http://www.jstor.org/stable/2985194`. 78

G Seber and J Wild. *Nonlinear Regression*. John Wiley & Sons, 1989. 1

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 (Statistical Methodology)*, 64(4):583–639, 2002. doi: http://dx.doi.org/10.1111/1467-9868.00353. 60

M Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62:795–809, 2000. doi: http://dx.doi.org/10.1111/1467-9868.00265. 78

M Tanner and W Wong. The calculation of the posterior distributions through data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1987. doi: http://dx.doi.org/10.1080/01621459.1987.10478458. 78

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(11):2310–2318, 2007. doi: http://dx.doi.org/10.1161/ATVBAHA.107.151407. 6, 7, 59

J Varah. A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal of Computational and Graphical Statistics*, 3(1):28–46, 1982. doi: http://dx.doi.org/10.1137/0903003. 17, 19

M Varziri, A Poyton, K McAuley, P McLellan, and J Ramsay. Selecting optimal weighting factors in ipda for parameter estimation in continuous-time dynamic models. *Computers & Chemical Engineering*, 32(12):3011–3022, 2008. doi: http://dx.doi.org/10.1016/j.compchemeng.2008.04.005. 141

P-F Verhulst. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique*, 10:113–121, 1838. 1

J. Woloszyn and K. McAuley. Application of parameter selection and estimation techniques in a thermal styrene polymerization model. *Macromolecular Reaction Engineering*, 5:453–466, 2011. ISSN 1862-8338. doi: http://dx.doi.org/10.1002/mren.20110002. 112

Chien-Fu Wu. Some algorithmic aspects of the theory of optimal designs. *The Annals of Statistics*, 6(6):1286–1301, 1978. URL `http://www.jstor.org/stable/2958716`. 141

S Wu, K McAuley, and T Harris. Selection of simplified models: I. analysis of model-selection criteria using mean-squared error. *The Canadian Journal of Chemical Engineering*, 89:148–158, 2011. ISSN 1939-019X. doi: http://dx.doi.org/10.1002/cjce.20406. 112

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`. 45