INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN


DISCUSSION
PAPER

2012/18

# ON THE USE OF ADHESION PARAMETERS TO VALIDATE MODELS SPECIFIED USING SYSTEMS OF AFFINE DIFFERENTIAL EQUATIONS

JAEGER, J. and Ph. LAMBERT

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

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

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

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

May 24, 2012

**Abstract**

A strategy for the selection of system 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.

**Acknowledgment:** *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 franaise de Belgique", granted by the "Académie universitaire Louvain", are gratefully acknowledged.*

# 1 Introduction

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). The existing approaches to estimate and solve these systems of differential equations can be classified into two classes. The first one relies on nonlinear least squares (NLS). It uses minimization techniques for the estimation of the ODE parameters and a numerical solver for the approximation of the solution. It can be computationally intensive and sometimes poorly suited for statistical inference. Alternatively, Ramsay et al. (2007) proposed to approximate the unknown state functions involved in the system of differential equations 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 this approach appears to be robust to model misspecification and to disturbances of the dynamical system. This frequentist approach 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 (Jaeger and Lambert, 2011) or suitable described by a mixture of Gaussian distributions (Jaeger and Lambert, 2012). 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 NLS 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 paper is to propose a Bayesian model selection strategy based on the study of ODE-adhesion

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

# 2 Bayesian ODE-penalized B-spline approach and ODE-model selection

In this section, we first recall the 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.

## 2.1 The Bayesian ODE-penalized B-spline approach

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

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

where $f$ is a known 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 $\epsilon_{jk}$. We denote by $y_{jk} = \boldsymbol{x}_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$ the corresponding measurement ($j \in \mathcal{J}$). The objective is to jointly estimate the ODE parameters $\boldsymbol{\theta}$, the state functions $\boldsymbol{x}(t)$ and the precision of measurement $\boldsymbol{\tau} = \text{vec}(\tau_j, j \in \mathcal{J})$ from $\{(t_{jk}, y_{jk}), j \in \mathcal{J}, k = 1, \ldots, n_j\}$. For simplicity, consider a single subject study with a standard Gaussian for the error distribution (see Jaeger and Lambert (2011) for a hierarchical setting and Jaeger and Lambert (2012) for flexible distributions of the error terms).

The solution of the system of differential equations is approximated using a B-spline basis function expansion. Denote by $\boldsymbol{B_j}(t)$ the $K_j$-vector of B-spline basis functions of order $p$ evaluated at time $t$ that is used to approximate the $j$th component of the state function $\boldsymbol{x}(t)$ and $\boldsymbol{c_j}$ the corresponding $K_j$-vector of B-spline coefficients. The approximation $\widetilde{x_j}(t)$ of $x_j(t)$ is expressed as a linear combination of these B-spline basis functions:

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

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

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

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

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

where $\boldsymbol{c} = \left(\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_d^T\right)^T$ is the vector of all spline coefficients of length $K = \sum_{j=1}^{d} K_j$, $\boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a $K \times K$ block-matrix constructed by placing the corresponding penalty matrices involved in each $PEN_j(\tilde{\boldsymbol{x}})$, $\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a vector of length $K$ corresponding to a penalty vector and $l(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is a penalty constant (see Section 2.4 in Jaeger and Lambert (2011) for the automatic construction of these quantities). The vector of ODE-adhesion parameters $\boldsymbol{\gamma} = \text{vec}(\gamma_j, j = 1, \ldots, d)$ permits 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.

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

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

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

$$\boldsymbol{v_1} = -\boldsymbol{r}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + \boldsymbol{\Sigma_c}^{-1} \boldsymbol{\mu_c} \tag{3}$$

and

$$\boldsymbol{V_1} = \boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + \boldsymbol{\Sigma_c}^{-1}. \tag{4}$$

Therefore, the normalizing constant for the conditional prior density of the spline coefficients $\boldsymbol{c}$ is:

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

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

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

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

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

For the vector $\boldsymbol{\theta}$ of differential equation parameters, the chosen prior, $p(\boldsymbol{\theta})$, will depends on the context.

## 2.2 Study of the posterior distribution of the ODE-adhesion parameter

For convenience, assume that all the ODE-adhesion parameters $\gamma_j$ $(j \in \mathcal{J})$ are identical to a common value $\gamma$ 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, the penalty vector (in Eq. (3)) and the penalty matrix (in Eq. (4)) can be rewritten as $\boldsymbol{r}(\boldsymbol{\theta}, \gamma) = \gamma \boldsymbol{s}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta}, \gamma) = \gamma \boldsymbol{S}(\boldsymbol{\theta})$. From Jaeger and Lambert (2011), one can show that the log conditional posterior density for the ODE-adhesion parameter (marginalized with respect to the spline coefficients) is equal to:

$$\begin{aligned} \log\left(p\left(\gamma|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{y}\right)\right) &\doteq \frac{K}{2}\log(\gamma) - \frac{\gamma}{2}\boldsymbol{s}^T(\boldsymbol{\theta})\,\boldsymbol{S}^{-1}(\boldsymbol{\theta})\,\boldsymbol{s}(\boldsymbol{\theta}) \\ &- \frac{1}{2}\log\left(\det(\boldsymbol{H})\right) + \frac{1}{2}\boldsymbol{h}^T\boldsymbol{H}^{-1}\boldsymbol{h} \\ &+ \log\left(p(\gamma)\right), \end{aligned} \tag{5}$$

where $\boldsymbol{H} = \gamma \boldsymbol{S}(\boldsymbol{\theta}) + \operatorname{diag}(\boldsymbol{Z_j}, j = 1, \ldots, d)$ with matrix $\boldsymbol{Z_j}$ that is equal to $\tau_j {\boldsymbol{B_j}}^T \boldsymbol{B_j}$ if the $j$th state function is observed and the null $K_j \times K_j$-matrix otherwise, and vector $\boldsymbol{h} = -\gamma \boldsymbol{s}(\boldsymbol{\theta}) + \operatorname{vec}(\boldsymbol{z_j}, j = 1, \ldots, d)$ with $\boldsymbol{z_j}$ equal to $\tau_j {\boldsymbol{B_j}}^T \boldsymbol{y_j}$ if the $j$th state function is observed and to the null $K_j$-vector otherwise. The symbol $\doteq$ means here "is equal up to an additive constant independent of $\gamma$".

### 2.2.1 Asymptotics of the marginalized posterior distribution of ODE-adhesion parameter

We will show that the difference between the conditional posterior distribution of the ODE-adhesion parameter and its prior tends to a constant independent of $\gamma$ when $\gamma$ is large. Large values of $\gamma$ are supported by the data when the differential equations adequatelly describe the dynamics of the system. In the log posterior density given in Eq. (5), two terms are of particular interest: $\log(\det(\boldsymbol{H}))$ and $\boldsymbol{h}^T\boldsymbol{H}^{-1}\boldsymbol{h}$. One can show that:

$$\boldsymbol{H} = \gamma\left(\boldsymbol{S}(\boldsymbol{\theta}) + \operatorname{diag}\left(\frac{1}{\gamma}\boldsymbol{Z_j}, j = 1, \ldots, d\right)\right), \tag{6}$$

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

$$\boldsymbol{h} = \gamma\left(-\boldsymbol{s}(\boldsymbol{\theta}) + \operatorname{vec}\left(\frac{1}{\gamma}\boldsymbol{z_j}, j = 1, \ldots, d\right)\right), \tag{7}$$

with $\frac{1}{\gamma}\boldsymbol{z_j}$ tending to the null vector when $\gamma$ tends to infinity (it is always equal to the null vector when the $j$th state function is unobserved). Therefore, given Eqs. (6) and (7), the difference between the log posterior density and the log prior density tends to a constant independent of $\gamma$ when $\gamma$ 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 results confirms that our choice for the prior distribution of the ODE-adhesion parameter indicates a prior confidence in the proposed ODE model.

### 2.2.2 Divergence measure and ODE-model selection

Asymptotic results from Section 2.2.1 confirms the statement of Jaeger and Lambert (2011), namely that the posterior distribution of 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 of the ODE-adhesion parameter can be obtained using a Metropolis-within-Gibbs algorithm (Jaeger and Lambert, 2011, 2012). The goal is now to measure the divergence between the prior and the posterior distribution of the ODE-adhesion parameter using the MCMC sample of the posterior distribution of the ODE-adhesion parameter.

To this end, we propose to consider several goodness of fit statistics. The first class of divergence measures are related to the empirical cumulative distribution function based on the sample of the posterior distribution of the ODE-adhesion parameter. These are the Kolmogorov-Smirnov (KS) statistic, the Cramér-von Mises (CVM) statistic and the Anderson-Darling (AD) statistic (Anderson and Darling, 1954). If these statistics are

small (i.e. closed to the reference values), then it suggests that the posterior distribution of the ODE-adhesion parameter can be considered similar to its prior distribution. In contrast to the KS statistic, CVM and AD statistics integrate the whole squared difference between the empirical cumulative distribution function and the corresponding (target) prior cumulative distribution function. Note also that the AD statistic is particularly sensitive to divergence in the tails of the target prior distribution.

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

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

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

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

where $\{\gamma_m, m=1,\ldots,M\}$ is a sample of the posterior distribution of the ODE-adhesion parameter and $\widetilde{q}\left(\cdot\right)$ denotes a kernel density estimate of the posterior distribution of 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.

# 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

$$\begin{cases} \frac{dx(t)}{dt} &= -k_e x\left(t\right) \\ x\left(0\right) &= 1 \\ k_e &\geq 0, \end{cases} \tag{8}$$

or on the bi-exponential decay case

$$\begin{cases} \frac{dx_1(t)}{dt} &= -\left(k_{cp}+k_e\right)x_1\left(t\right)+k_{pc}x_2\left(t\right) \\ \frac{dx_2(t)}{dt} &= k_{cp}x_1\left(t\right)-k_{pc}x_2\left(t\right) \\ x_1\left(0\right) &= 1 \\ x_2\left(0\right) &= 0 \\ k_e, k_{cp}, k_{pc} &\geq 0 \\ k_{cp} &> k_{pc}. \end{cases} \tag{9}$$

If Eq. (8) (resp. Eq. (9)) 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 $\epsilon_k$. For each simulation, 1000 datasets were generated. The prior distribution of the ODE-adhesion parameter supposed for the complete simulation is a gamma distribution $\mathcal{G}\left(1, 10^{-8}\right)$. The results provided in the following subsections describe the simulation results.

### 3.1 Sensitivity to the sample size

The aim of this simulation is to see if the posterior distribution is sensitive to the sample size. To do that, we generate data using Eq. (8) 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 2.1 with ODE parameter $k_e$ and precision of measurement set to their simulation values. Figure 1 shows the that posterior median of $\log(\gamma)$ is not impacted by the sample size. This is also visually confirmed for all the divergence measure (e.g. the Kullback-Leibler divergence in Figure 2).

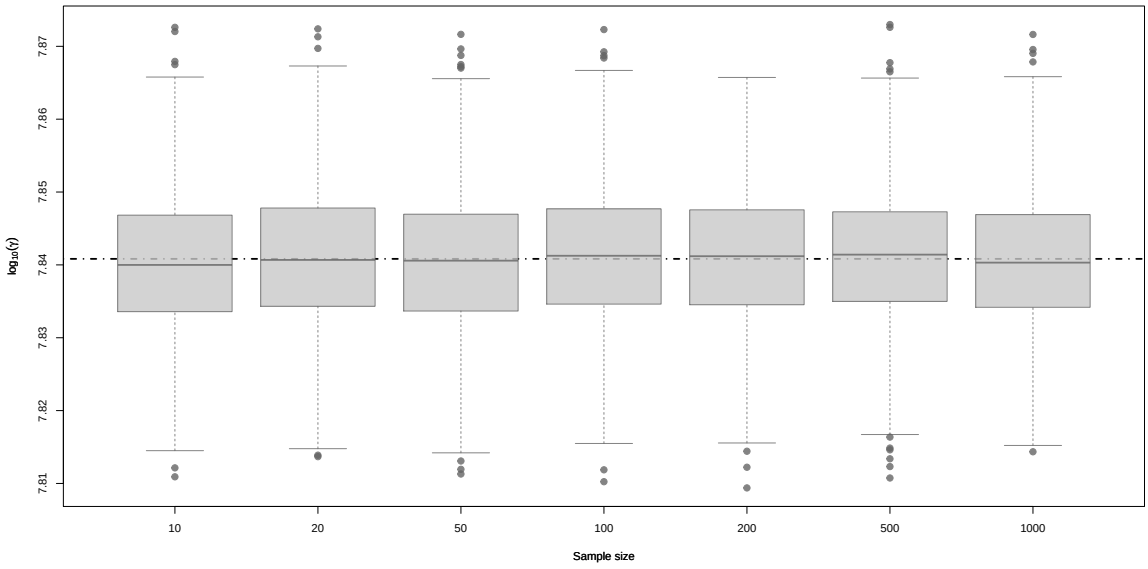


Figure 1: Sensitivity to the sample size - Boxplot of posterior empirical median of $\log_{10}(\gamma)$ for the seven sample size with prior median (horizontal dashed line).

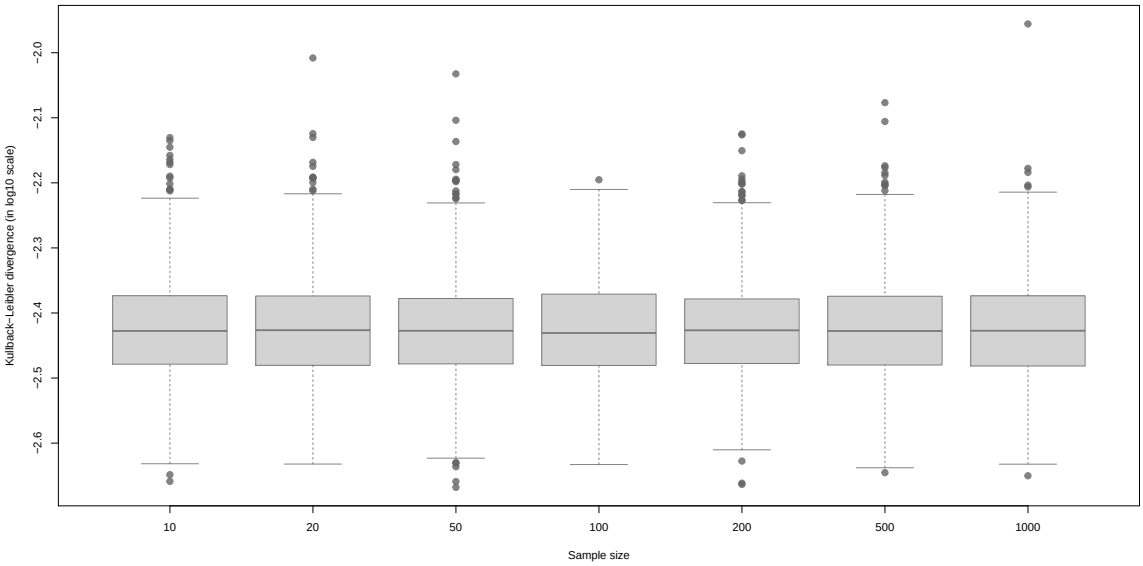


Figure 2: Sensitivity to the sample size - Boxplot of estimated Kullback-Leibler divergence (on the log10 scale) for the seven sample size.

## 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. (8) 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 2.1 with the precision of measurement $\tau$ 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 of the ODE-adhesion parameter $\gamma$ was sensitive to a wrong specification of the ODE parameter $k_e$ especially when the sample size increases. Figure 3 and Table 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 4 shows the histogram of a sample of the posterior distribution of 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 of the ODE-adhesion parameter is similar to its prior when the model is adequately specified whereas the posterior distribution of the ODE-adhesion parameter differs from

its prior when the ODE model is wrongly specified.

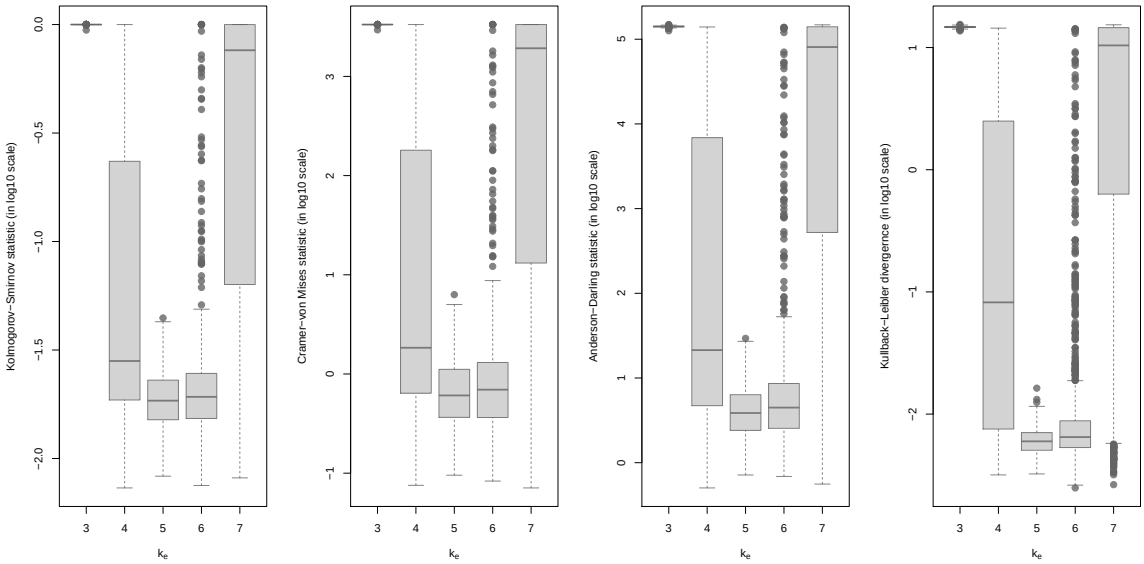


Figure 3: Sensitivity to a wrong specification of the ODE parameter - Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence (on the log10 scale) when $n = 100$ and $k_e$ is set equal to 3, 4, 5, 6 or 7 in the MCMC procedure (while $k_e =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

**Wrong specification of the precision of measurement**

Again, we consider Eq. (8) 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 2.1 with ODE parameter $k_e$ equals to its simulation value. A grid of 25 equidistant value 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 of the ODE-adhesion parameter is only affected by values of the precision larger than the one used for the data generation (see Figure 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 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 of the ODE-adhesion parameter.

| | | 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 |

Table 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.

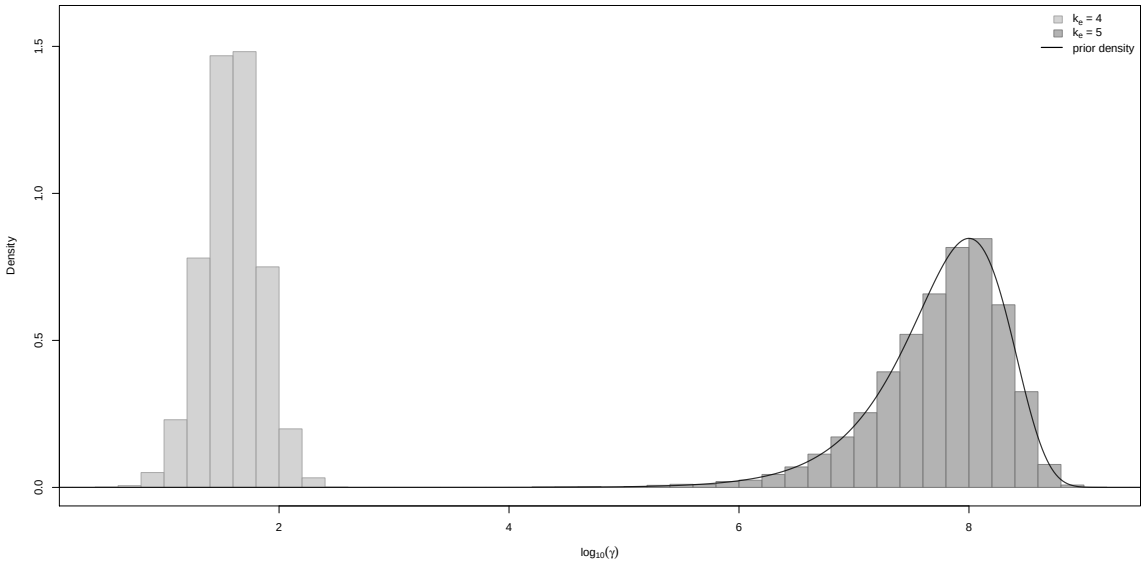


Figure 4: Sensitivity to a wrong specification of the ODE parameter - Histograms based on a sample of the posterior distribution of the ODE-adhesion parameter when $n = 100$ and $k_e$ is equal to 4 (light gray histogram) or 5 (dark grey histogram) while $k_e = 5$ was used to generate datasets.

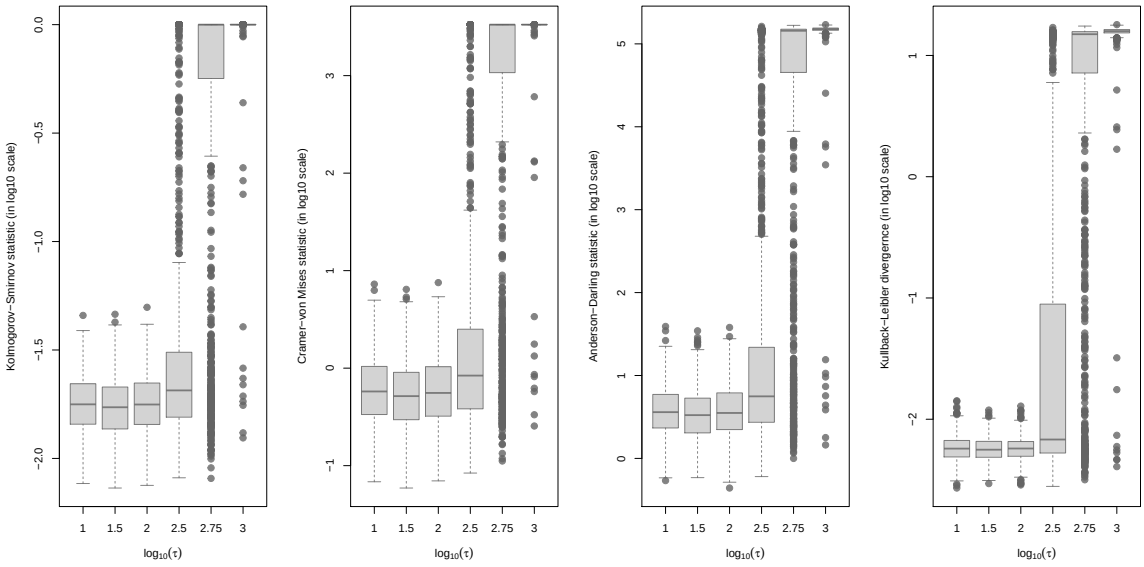


Figure 5: Sensitivity to a wrong specification of the precision of measurement - Kolmogorov-Smirnov statistic, Cramér-von Mises statistic, Anderson-Darling statistic and Kullback-Leibler divergence (on the log10 scale) when $n = 100$ and $\log_{10}(\tau)$ is set equal to 1, 1.5, 2, 2.5, 2.75 or 3 in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

| | | 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 |

Table 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.

### 3.3 Sensitivity to a misspecification of the error distribution

The aim of this simulation is to see if the posterior distribution of the ODE-adhesion parameter is sensitive to a misspecification of the error distribution. Again, we consider Eq. (8) 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 $0.4\mathcal{N}\left(-1.5, 0.9^2\right)+0.6\mathcal{N}\left(1, 0.6^2\right)$). All the simulated datasets were analyzed with the Bayesian model specified in Section 2.1 assuming a Gaussian error distribution with ODE parameter $k_e$ and precision of measurement $\tau$ equal to their simulation values. According to the simulation results, the posterior distribution of the ODE-adhesion parameter seems to be insensitive to the misspecification of the error distribution (see Figure 6 and Table 3 for the Kullback-Leibler divergence when then sample size is equal to 100). This result is consistent with the simulation results obtained in Jaeger and Lambert (2012).

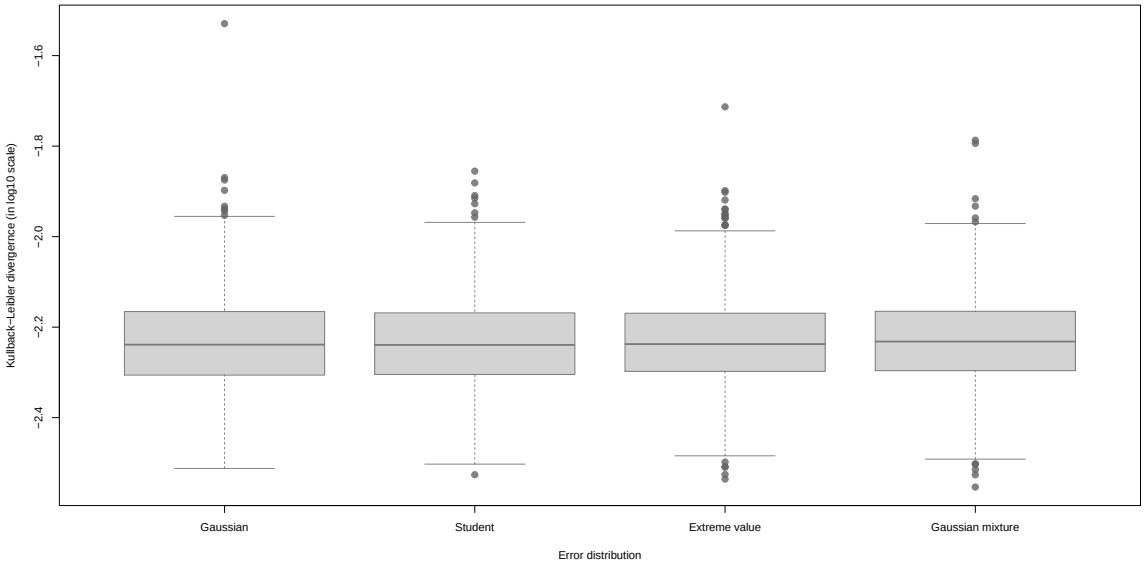


Figure 6: Sensitivity to a misspecification of the error distribution - Boxplot of estimated Kullback-Leibler divergence (on the log10 scale) for sample size equals to 100 with four error distributions (Gaussian, Student, extreme value and mixture of two Gaussians) used for the data generation. Gaussian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

### 3.4 Sensitivity to a ODE-model misspecification

For this simulation, we consider Eq. (9) 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

| | | 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 |

Table 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.

simulated dataset is analyzed with the Bayesian model specified in Section 2.1 wrongly assuming that the dynamic of the observed state function is given by Eq. (8). Figure 7 shows that, based on the Kullback-Leibler divervence, the posterior distribution of 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 of the ODE-adhesion parameter appears to be sensitive to the sample size when the ODE-model misspecification is detected.

# 4 Application

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

## 4.1 Data and pharmacokinetic compartment model

Kwan et al. (1976) studied the pharmacokinetics (PK) 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 $\mu$g/mL) were measured 11 times from 15 minutes to 8 hours after injection. Figure 8 shows the PK profiles. All the concentration curves have a similar shape but seem to differ across individuals. Our goal is to confirm that the kinetic of Indomethacin after a bolus intravenous injection is modeled by a 2-compartment IV bolus model

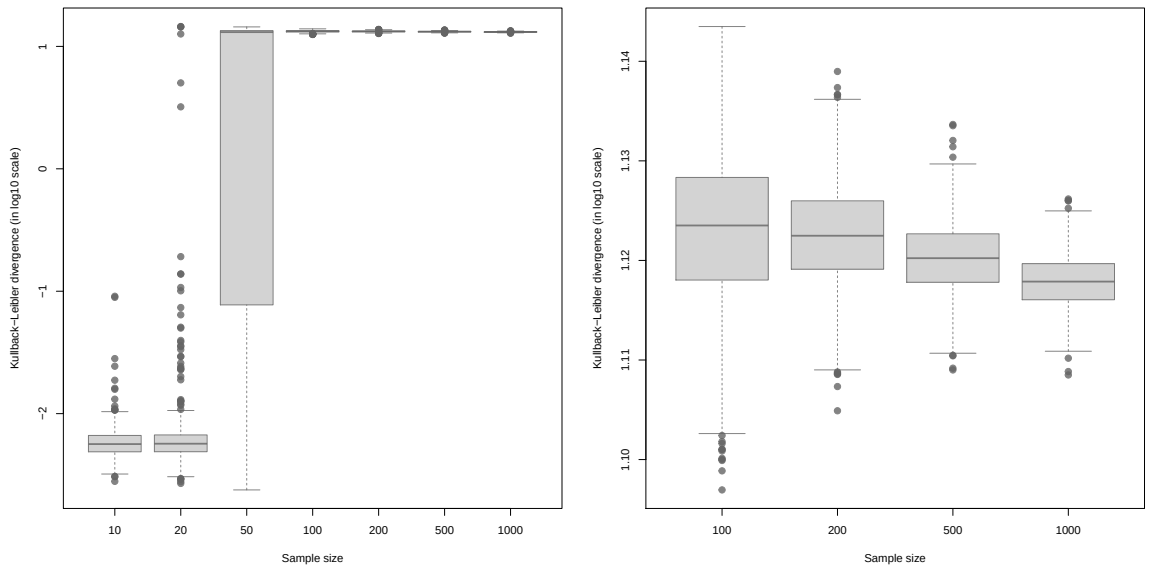


Figure 7: Sensitivity to ODE-model misspecification - Boxplot of the Kullback-Leibler divergence (on the log10 scale) estimated from data generated using a bi-exponential decay model while an exponential decline one was (wrongly) assumed in the analysis. Figure on the right is a zoom for large sample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

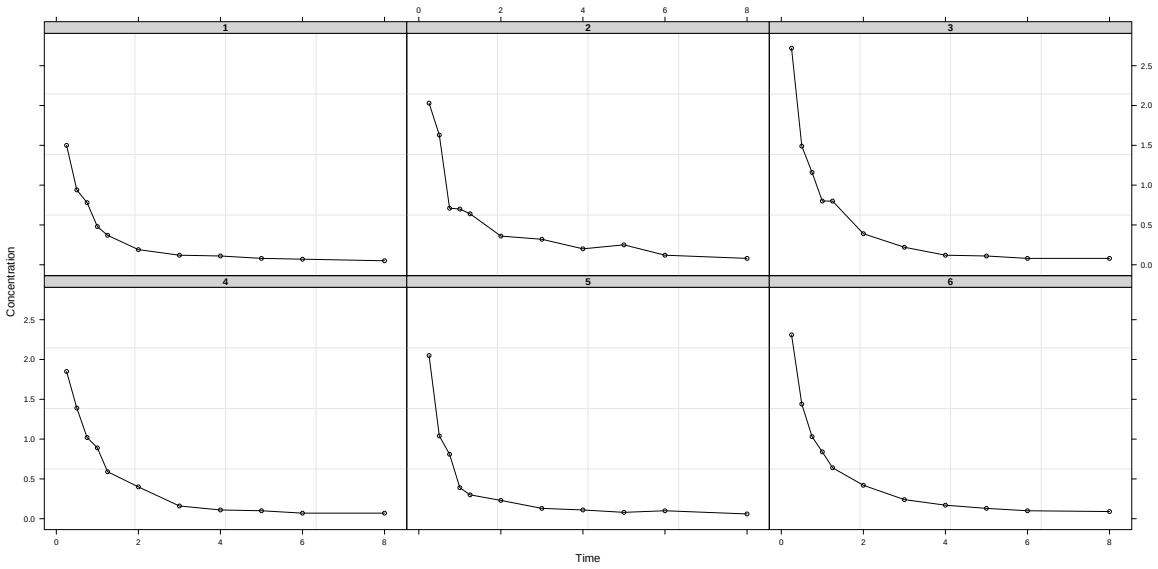


Figure 8: Concentration of Indomethicin over time for six subjects after an IV bolus injection.

and that the 1-compartment IV bolus model fails to describe the kinetics of the drug in the plasma. The basic idea with PK compartment models is to view the body as a system of compartments communicating with each other. Each compartment corresponds to a tissue or a group of tissues with similar blood flow and uniform drug concentration.

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

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

The 2-compartment IV 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:

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

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).

## 4.2 Bayesian model

For the 1-compartment IV bolus model, the plasma concentration of Indomethacin is approximated using a 5-order B-spline basis function expansion with knots at each tenth between $0h$ and $8h$. The Bayesian model is:

$$\begin{cases} y_k|\boldsymbol{c}, \tau, V &\sim \mathcal{N}\left(\frac{1}{V}\left(\boldsymbol{B}(t_k)\right)^T \boldsymbol{c}; \tau^{-1}\right) \\ p(\boldsymbol{c}|\gamma, k_e, V) &\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T \boldsymbol{M_1} \boldsymbol{c} - 2\boldsymbol{c}^T \boldsymbol{v_1}\right)\right) \\ \gamma_{11}, \tau &\sim \mathcal{G}\left(1; 10^{-8}\right) \\ p(k_e, V) &\propto \mathbb{I}\{k_e > 0, V > 0\}, \end{cases}$$

where $\boldsymbol{c}$ are the spline coefficients, $\gamma_{11}$ is the ODE-adhesion parameter, $\boldsymbol{M_1} = \boldsymbol{R}(k_e, V, \gamma_{11}) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Jaeger and Lambert (2011) 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 $\gamma_{11}$ 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.

For the 2-compartment IV bolus model, the drug concentration in the central compartment 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:

$$\begin{cases} y_k|\boldsymbol{c_c}, \tau, V & \sim & \mathcal{N}\left(\frac{1}{V}\left(\boldsymbol{B_c}(t_k)\right)^T \boldsymbol{c_c}; \tau^{-1}\right) \\ p(\boldsymbol{c_1}, \boldsymbol{c_2}|\gamma_{21}, \gamma_{22}, k_e, k_{cp}, k_{pc}, V) & \propto & \exp\left(-\frac{1}{2}\left(\boldsymbol{c}^T\boldsymbol{M_1}\boldsymbol{c} - 2\boldsymbol{c}^T\boldsymbol{v_1}\right)\right) \\ \gamma_{21}, \gamma_{22}, \tau & \sim & \mathcal{G}\left(1; 10^{-8}\right) \\ p(k_e, k_{cp}, k_{pc}, V) & \propto & \mathbb{I}\{k_e > k_{cp} > k_{pc} > 0, V > 0\}, \end{cases}$$

where $\boldsymbol{c}^T = \left(\boldsymbol{c_1}^T, \boldsymbol{c_2}^T\right)$, $\gamma_{21}$ and $\gamma_{22}$ are the ODE-adhesion parameters, $\boldsymbol{M_1} = \boldsymbol{R}(k_e, k_{cp}, k_{pc}, V, \gamma_{21}, \gamma_{22}) + {\boldsymbol{\Sigma_c}}^{-1}$ and $\boldsymbol{v_1} = {\boldsymbol{\Sigma_c}}^{-1}\boldsymbol{\mu_c}$ (see Jaeger and Lambert (2011) 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. The ODE-adhesion parameter $\gamma_{21}$ is related to the variation of drug concentration in the central compartment while $\gamma_{22}$ is related to the variation of drug concentration in the peripheral compartment. We parameterize the PK parameters in the log-scale with Gaussian random effect at subject level and force $k_{cp}$ to be larger than $k_{pc}$.

For the two models, a chain of length 10000 was run after a burnin of 10000 iterations for all the parameters using the Metropolis-within-Gibbs algorithm described in Jaeger and Lambert (2011). Note also 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.

### 4.3 Results

Model selection is made by comparing the prior and the posterior distributions of the ODE-adhesion parameters $\gamma_{11}$ and $\gamma_{21}$. These two parameters are related to the variation of drug concentration in the central compartment. Figure 9 shows on the same graph the histogram of the MCMC chain for the ODE-adhesion parameter $\gamma_{11}$ (light grey), the histogram of the MCMC chain for the ODE-adhesion parameter $\gamma_{21}$ (dark grey) and their common prior

distribution on the log10 scale (black solid curve). The posterior distribution for the ODE-adhesion parameter $\gamma_{21}$ is similar to its prior distribution suggesting that the 2-compartment IV bolus model reasonably describes the dynamics of the plasma concentration of Indomethacin. The histogram of the ODE-adhesion parameter $\gamma_{11}$ (on the log10 scale) is bimodal: 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 IV bolus model. Table 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 $\gamma_{21}$ is larger than $\gamma_{11}$ is estimated by 0.9539. It suggests to select the 2-compartment IV bolus model to describe the dynamics of the drug concentration in the plasma. The scatter plot of the posterior predictive residuals versus the mean predictive concentrations

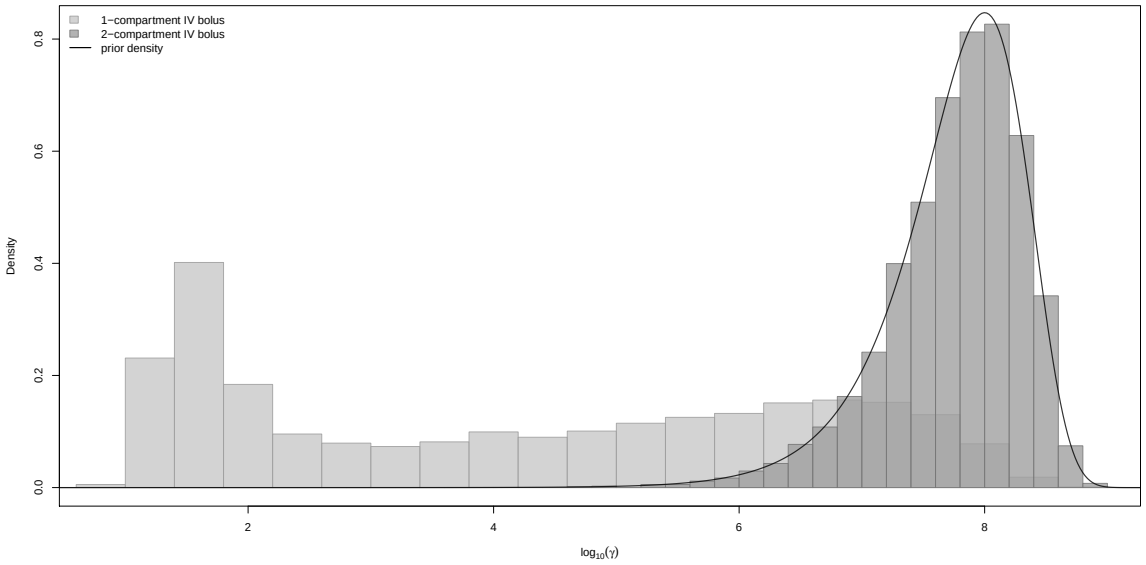


Figure 9: Histogram based on a sample of the posterior distribution of the ODE-adhesion parameter $\gamma_{11}$ (light gray, log10 scale) in the 1-compartment IV bolus model, histogram based on a sample of the posterior distribution of the ODE-adhesion parameter $\gamma_{21}$ (dark gray, log10 scale) in the 2-compartment IV bolus model and the common prior distribution for the corresponding ODE-adhesion parameter (black solid line).

does not clearly indicate any trend (Figure 10). The two outliers are for subject 2. Figure 11 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 12 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 compatible with the expected coverage.

| | Model | |
|---|---|---|
| Divergence measure | 1-compartment IV bolus | 2-compartment IV bolus |
| Kolmogorov-Smirnov | 7.62e-1 | 4.47e-2 |
| Cramér-von Mises | 2.56e+3 | 7.50 |
| Anderson-Darling | 4.76e+4 | 6.11e+1 |
| Kullback-Leibler | 6.03 | 2.62e-2 |

Table 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 $\log_{10}(\gamma_{21})$.

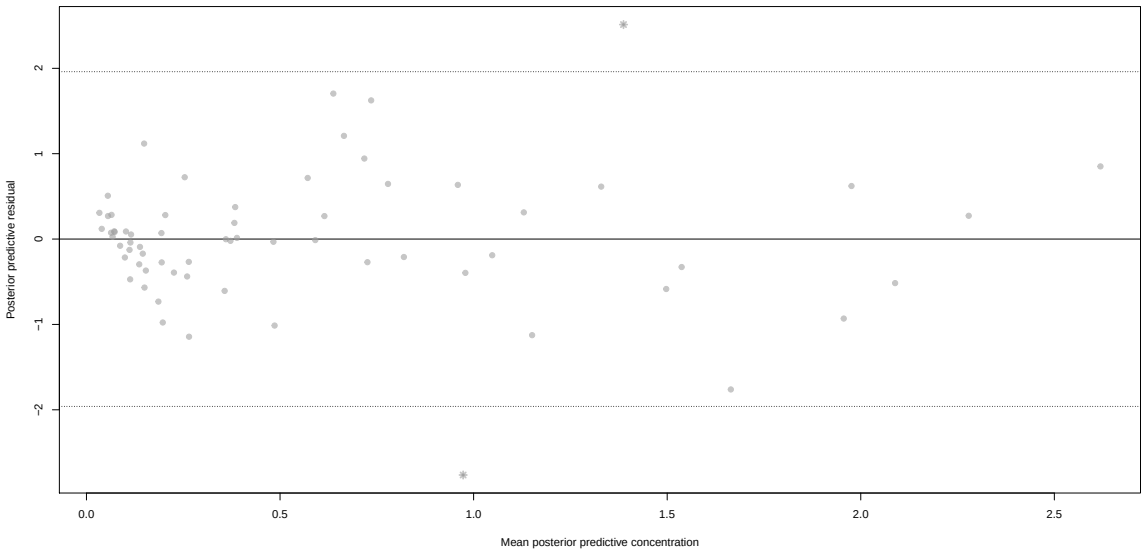


Figure 10: Scatter plot of posterior predictive residuals versus mean predictive concentrations. Outliers are represented by stars.

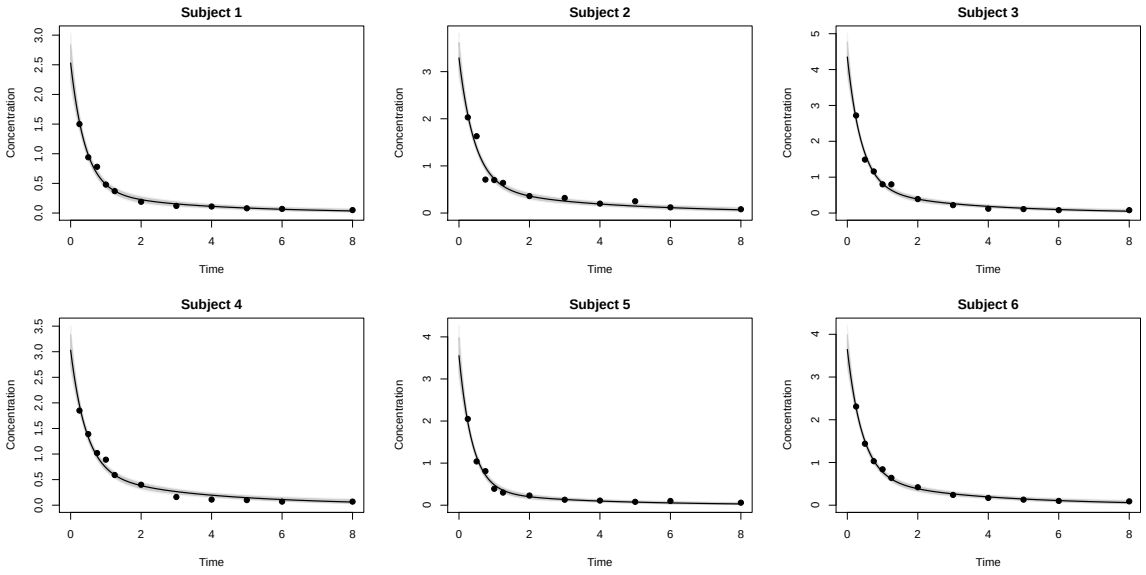


Figure 11: Pointwise 80% (dark grey region) and 95% (light grey region) credibility intervals for the fitted plasma concentration of Indomethacin with pointwise median (black solid curve) and observed concentration of Indomethacin (solid circles) versus time for the six subjects.

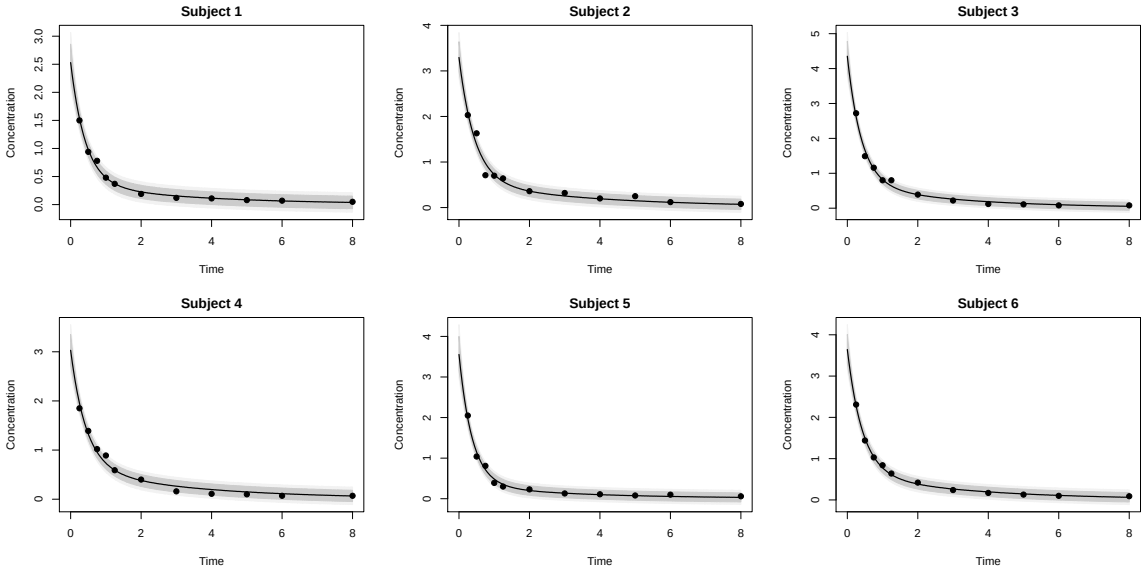


Figure 12: Pointwise 80% (dark grey region) and 95% (light grey region) posterior predictive credibility intervals for the plasma concentration of Indomethacin with pointwise posterior predictive median (black solid curve) and observed concentration of Indomethacin (solid circles) versus time for the six subjects.

# 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 of 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 of 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 of the ODE-adhesion parameter. We recommend to combine that strategy with other established analysis such as residual inspection. Some extensions are desirable. The proposed Bayesian approach was deliberately focused on the affine ODE case. Jaeger and Lambert (2011) highlight the major role of the normalization constant in the prior distribution of the spline coefficients in the inferential process. This is an essential information to deal with more advanced problems such as systems of nonlinear differential equations where by simplicity one might be tempted to ignore such a constant as it has no explicit analytical form. Analytical or numerical approximations will be required and should be assessed in these challenging settings.

# References


T Anderson and D Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49: 765–769, 1954.

P L Bonate. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer, 2011.

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

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

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

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.

S Kullback and R Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

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*, 4:255–280, 1976. ISSN 1567-567X.

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

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.

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

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*, 69:741–796, 2007.

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

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.

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.