Parametric estimation of complex mixed models based on meta-model approach

Pierre Barbillon[1,2], Célia Barthélémy[3], and Adeline Samson[4,5]

[1]AgroParisTech / UMR INRA MIA, F-75005 Paris
[2]INRA, UMR 518, F-75005 Paris
[3]INRIA Saclay, Popix team, Orsay
[4]Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France
[4]CNRS, LJK, F-38000 Grenoble, France

June 10, 2015

Abstract

Complex biological processes are usually experimented along time among a collection of individuals. Longitudinal data are then available and the statistical challenge is to better understand the underlying biological mechanisms. The standard statistical approach is mixed-effects model, with regression functions that are now highly-developed to describe precisely the biological processes (solutions of multi-dimensional ordinary differential equations or of partial differential equation). When there is no analytical solution, a classical estimation approach relies on the coupling of a stochastic version of the EM algorithm (SAEM) with a MCMC algorithm. This procedure needs many evaluations of the regression function which is clearly prohibitive when a time-consuming solver is used for computing it. In this work a meta-model relying on a Gaussian process emulator is proposed to replace this regression function. The new source of uncertainty due to this approximation can be incorporated in the model which leads to what is called a mixed meta-model. A control on the distance between the maximum likelihood estimates in this mixed meta-model and the maximum likelihood estimates obtained with the exact mixed model is guaranteed. Eventually, numerical simulations are performed to illustrate the efficiency of this approach.

Keywords: Mixed models, Stochastic EM algorithm, MCMC methods, Gaussian Process emulator.

1 Introduction

Mixed-effects model methodology (Pinheiro and Bates, 2000) allows to discriminate between inter- and intra-subjects variabilities which is essential when dealing with longitudinal data. Statistical methods for mixed models are now well established (see

references below) but can be time consuming depending on the complexity of the regression functions. Indeed sophisticated mathematical models have been developed to describe precisely biological processes: multi-dimensional ordinary differential equations (ODE) (see Wu et al., 2005; Guedj et al., 2007; Lavielle et al., 2011; Ribba et al., 2012, for modeling viral load decrease in HIV patients or tumor growth) or partial differential equation (PDE) (see Grenier et al., 2014; Chatterjee et al., 2012, for modeling tumor growth or HVC viral kinetic). These mathematical models have no analytical solution and only an approximate solution can be obtained with computationally intensive numerical methods. This induces a huge increase of the computation cost of the estimation method (up to 23 days according to Grenier et al., 2014). Therefore, there is a crucial need to develop new statistical approaches to reduce the computation time. The significant computation time of mixed models estimation methods is due to their iterative settings, compulsory to sidestep the intractability of the likelihood. This is true for methods based on linearisation (Pinheiro and Bates, 2000) or likelihood numerical approximation (Davidian and Giltinan, 1995; Wolfinger, 1993) and this is crucial for EM-type methods such as stochastic EM algorithms (Wei and Tanner, 1990; Kuhn and Lavielle, 2005).

Our objective is to propose a way of reducing the computation time of the SAEM-MCMC algorithm (Kuhn and Lavielle, 2005) for complex mixed models, together with a theoretical study of the resulting estimator. When the regression function is not analytically available (and we call it also computer model in the rest of the paper), extensions of SAEM have already been proposed. Donnet and Samson (2007) deal with the case of an ODE mixed model, approximating the solution with a numerical scheme and studying the influence of this scheme to the properties of the approximate maximum likelihood estimator. But this approach remains time consuming when the ODE is multi-dimensional. For a PDE mixed model, Grenier et al. (2014) propose to approximate the PDE with a numerical scheme on a predefined grid, and then to interpolate the solution of the PDE linearly between two points of the grid. This linear approximation allows substantially reducing the computation time from 23 days to around 30 minutes, but may lead to biased estimates depending on the non linearity of the model.

The keystone of providing an efficient estimation method with good statistical properties is the choice of the procedure approximating the regression function. More accurate surrogates of computer model than linear approximation rely on Gaussian process emulation which consists of modeling the computer model as the realization of a Gaussian process (Sacks et al., 1989; Santner et al., 2003; Fang et al., 2005). This technique is also known as Kriging. A cheap approximation of the model, the emulator, is obtained by conditioning the Gaussian process on some evaluations of the model corresponding to inputs of a well-chosen design of numerical experiments. This stochastic modeling of the computer model provides also a measure of uncertainty on the precision of the approximation as a supplementary variance-covariance function which can be integrated in the mixed model. This approach has been already coupled with a Stochastic EM algorithm (Barbillon et al., 2011) or with a Bayesian procedure (Fu et al., 2014) in regression models without random effects. In this paper, we propose to couple the SAEM algorithm with this Gaussian process emulator, incorporating this new source of uncertainty due

to the approximation. Thus, providing a confidence interval of the unknown parameters takes into account the error induced by the approximation. We will refer to this approach as the complete mixed meta-model. However, the supplementary variance-covariance function in the model induces a loss of independence of the observations obtained from the different subjects which increases the computational burdensome of the MCMC scheme. That is why we also propose two simplified versions: the first one (called intermediate) by considering a diagonal variance-covariance function of the approximation error; the second one (called simple) by only using the approximation of the computer model and not incorporating the variance-covariance function. These two last versions allow to assume the independence between the subjects, and to reduce significantly the computational cost.

The Gaussian process emulator can also be interpreted as an approximation of the computer model by kernel interpolation with radial basis function as in Schaback (1995, 2007). In this framework, a point-wise control on the error of approximation is provided. Hence, we are able to guarantee a control on the distance between the maximum likelihood estimates in the approximate mixed meta-models and the maximum likelihood estimates obtained with the exact computer model. This control is decreasing to zero as a function of the space-fillingness of the design of numerical experiments.

The paper is organized as follows. Section 2 introduces the standard non-linear mixed model and Section 3 recalls the principles and the main results of the Gaussian process emulation. Section 4 introduces three mixed models approximated by Gaussian process emulator. In Section 5, three versions of the SAEM algorithm coupled to a Gaussian process emulator are proposed. Theoretical results are given in Section 6. A simulation study illustrates these results (Section 7). Section 8 concludes the paper with some extensions. Proofs are gathered in Appendix.

2 Mixed model and notations

Let us define $\mathbf{y}_i = {}^t(y_{i1}, \ldots, y_{in_i})$ where $y_{ij} \in \mathbb{R}^p$ is the response for individual i at time t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, and let $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ be the vector of all observations, of size $n_{tot} = \sum_{i=1}^{N} n_i$. We assume that the individual vectors $\mathbf{y}_i$ are described by a non-linear mixed model, defined as follows, for $j = 1, \ldots, n_i$:

$$\begin{array}{rcl} y_{ij} & = & f(t_{ij}, \psi_i) + \sigma_\varepsilon \, \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0,1)\,, \\ \psi_i & \sim_{iid} & \mathcal{N}(\mu, \Omega), \end{array} \tag{1}$$

where $f(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^p$ is the regression function, ψ_i is a d-vector of individual parameters. The $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^t$ represents the Gaussian centered residual error, independent of ψ_i. The individual parameter ψ_i are assumed to be random, Gaussian with expectation μ and $d \times d$ covariance matrix Ω. Note that the individual vectors $(\mathbf{y}_i)_i$ are independent and identically distributed.
The quantities we want to estimate from the observations $\mathbf{y}$ are the population parameters $\theta = (\mu, \Omega, \sigma_\varepsilon^2)$. In the following, we restrict to the case of scalar observations ($p = 1$) to ease the reading, but the extension to a multidimensional observation with $p > 1$ is straightforward.

We want to estimate θ by maximum likelihood. The likelihood of model (1) is:

$$
\begin{aligned}
p(\mathbf{y},\theta) &= \int p(\mathbf{y},\boldsymbol{\psi};\theta)\,d\boldsymbol{\psi} = \prod_{i=1}^{N}\int p(\mathbf{y}_i|\psi_i;\theta)p(\psi_i;\theta)d\psi_i \\
&= \prod_{i=1}^{N}\int \left\{ \frac{1}{(2\pi\sigma_\varepsilon^2)^{n_i/2}} \times \exp\left(-\frac{1}{2}\,{}^t(\mathbf{y}_i-\mathbf{f}(\mathbf{t}_i,\psi_i))(\sigma_\varepsilon^2 I_{n_i})^{-1}(\mathbf{y}_i-\mathbf{f}(\mathbf{t}_i,\psi_i))\right)\right. \\
&\quad \left. \times \frac{1}{(2\pi)^{d/2}|\Omega|^{1/2}} \exp\left(-\frac{1}{2}\,{}^t(\psi_i-\mu)\Omega^{-1}(\psi_i-\mu)\right) d\psi_i \right\} \qquad (2)
\end{aligned}
$$

where $\mathbf{t}_i = (t_{i1},\ldots,t_{in_i})$ and
$\mathbf{f}(\mathbf{t}_i,\psi_i) = {}^t(f(t_{i1},\psi_i),\ldots,f(t_{in_i},\psi_i))$. When f is non linear with respect to $\boldsymbol{\psi}$, the maximum likelihood estimator has no closed form. Any estimation method adapted to non-linear mixed models would require a very large number of evaluations of f, which could be time consuming when the structural function f is a computer model. Therefore, there is a real need to consider approximations of the function f that are simple to evaluate at any point (t,ψ). For that purpose, we introduce the framework of meta-model in the next section.

3 Meta-model

We start with the point of view of conditioned Gaussian process emulation which has the nice feature of incorporating as a variance-covariance function the additional source of uncertainty due to the approximation. This will naturally lead to a mixed meta-model on which the SAEM-MCMC can be performed. We also link this framework to the kernel interpolation framework since we need the deterministic point-wise control on the approximation to obtain the control between the maximum likelihood estimates corresponding to the exact model and to its approximation.

3.1 Conditioned Gaussian process

In this framework, the function f is interpreted as a realization of a Gaussian Process. Let us denote F_λ a Gaussian process defined, for any $x=(t,\psi)$, as

$$
F_\lambda(x) = \sum_{j=1}^{L}\beta_j h_j(x) + \zeta(x) = {}^t H(x)\boldsymbol{\beta} + \zeta(x)\,, \qquad (3)
$$

where $h_1,\ldots,h_L$ are regression functions, $H=(h_1,\ldots,h_L)$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_L)$ is a vector of parameters, ζ is a centered Gaussian process with covariance function

$$
\mathrm{Cov}(\zeta(x),\zeta(x')) = \sigma^2 K^\phi(x,x')\,,
$$

where K^ϕ is a correlation function depending on some parameters ϕ and $\lambda=(\boldsymbol{\beta},\sigma,\phi)$ is the vector of all unknown parameters. For instance, the so-called Gaussian kernel is defined by $K^\phi(x,x') = \exp(-\phi\|x-x'\|^2)$. The regression functions $h_1,\ldots,h_L$ are

usually linear functions or low degree polynomials. The kernel K^ϕ has to be chosen with respect to the assumed regularity of the function f. Similarly, the regression functions H have to be chosen with respect to the supposed trend in the function f if some insights on the function are available. See Koehler and Owen (1996); Fang et al. (2005) for detailed discussions on the choice of the regression functions and kernels.

We assume that we are able, in a pre-computation step, to evaluate precisely the function f n_D times for a given design of numerical experiments, $D = \{x_1, \dots x_{n_D}\}$. These "exact" evaluations are denoted $z_1 = f(x_1), \dots, z_n = f(x_{n_D})$. The (z_k) are different from the (y_{ij}) considered in model (1) which are noisy observations of f in some unknown points $x = (t, \boldsymbol{\psi})$. The design of experiments D is usually chosen with respect to a space-filling criterion (Fang et al., 2005) in a bounded domain where the points $(t_{ij}, \psi_i)_{i,j}$ are assumed to be. For a given kernel K, and a given vector H of regression functions, the vector of parameters λ has to be estimated. Usually, λ is estimated by maximizing the log-likelihood ℓ_F of the Gaussian process F (3):

$$\begin{aligned} \ell_F(\lambda; \mathbf{z}_D) &= -\frac{1}{2}\log((2\pi\sigma^2)^{n_D}|\Sigma^\phi_{DD}|) \\ &\quad -\frac{1}{2}\,{}^t(\mathbf{z}_D - H_D\boldsymbol{\beta})(\sigma^2\Sigma^\phi_{DD})^{-1}(\mathbf{z}_D - H_D\boldsymbol{\beta})\,, \end{aligned} \tag{4}$$

where $(\Sigma^\phi_{DD})_{1\le k,j\le n_D} = (K^\phi(x_k, x_j))$, $(H_D)_{1\le k\le n_D, 1\le j\le L} = h_j(x_k)$. The Matlab toolbox DACE (Lophaven et al., 2002) provides an optimization algorithm to estimate directly λ. We denote by $\hat\lambda = (\hat{\boldsymbol{\beta}}, \hat\sigma, \hat\phi)$ the estimates. The Gaussian process is chosen to be $F = F_{\hat\lambda}$. It corresponds to a plug-in approach since from now, these parameters are considered as known.

Then f is not directly approximated by F, but rather by the conditional process denoted F^D, defined as the process F conditionally to $F(x_1) = z_1, \dots, F(x_{n_D}) = z_{n_D}$, in short $\mathbf{Z}_D = \mathbf{z}_D$. The process F^D is still a Gaussian process, defined by its mean and covariance functions, which can be exactly computed. Let us introduce the partial functions, for any $x \in \mathbb{R}^{d+1}$, $K^\phi_x : \mathbb{R}^{d+1} \to \mathbb{R}$ defined by $K^\phi_x(x') = K^\phi(x, x')$ for any x' and the vector $\Sigma^{\hat\phi}_{xD} = \left(K^{\hat\phi}_x(x_k)\right)_{1\le k\le n_D}$. Then, the mean $m_D(x)$ and covariance $C_D(x, x')$ of the process F^D are defined, for all x, x' as

$$m_D(x) = H(x)^t\hat{\boldsymbol{\beta}} + {}^t\Sigma^{\hat\phi}_{xD}(\Sigma^{\hat\phi}_{DD})^{-1}(\mathbf{z}_D - H_D\hat{\boldsymbol{\beta}})\,, \tag{5}$$

$$C_D(x, x') = \hat\sigma^2(K^{\hat\phi}_x(x') - {}^t\Sigma^{\hat\phi}_{xD}(\Sigma^{\hat\phi}_{DD})^{-1}\Sigma^{\hat\phi}_{x'D})\,. \tag{6}$$

The mean function m_D provides an approximation of the function f for any x and the variance function $x \mapsto C_D(x, x)$ measures the confidence in the accuracy of this approximation. The plug-in approach for $\lambda = (\beta, \sigma^2, \phi)$ may lead to underestimate the uncertainty on the quality of the approximation which is showed to be asymptotically negligible (Prasad and Rao, 1990). It can be used only for the parameter of the correlation kernel ϕ. In this case, when the process is not conditioned on $(\hat{\boldsymbol{\beta}}, \hat\sigma)$, the conditioned process is a Student T-process with still closed-form location and scale (Santner et al., 2003). However, we prefer the complete plug-in approach to deal with a Gaussian distribution which is easier to incorporate in the mixed model.

3.2 Kernel interpolation

We can interpret the previous meta-model approximation as a kernel interpolation. Indeed, a Reproducing Kernel Hilbert Space (RKHS) can be associated to K^ϕ as soon as the kernel K^ϕ is positive definite. Then the partial functions K_x^ϕ defined in Section 3.1 span a pre-Hilbert space with inner product $< K_x^\phi, K_{x'}^\phi >= K^\phi(x, x')$. Aronszajn's theorem states that there exists a unique completion $\mathcal{H}_K$ of this space with the reproducing property:

$$\forall v \in \mathcal{H}_K,\ x \in \mathbb{R}^{d+1},\ v(x) =< v, K_x^\phi > \ .$$

The space $\mathcal{H}_K$ is the RKHS associated to kernel K. Then, we focus on the function $g(x) = f(x) - {}^tH(x)\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is estimated as before. Note that this function can be seen as the residuals of the linear model (3) of the random variables $Z_1, \ldots, Z_{n_D}$ on the design $D = \{x_1, \ldots x_{n_D}\}$:

$$Z_i = f(x_i) = {}^tH(x_i)\hat{\boldsymbol{\beta}} + g(x_i).$$

We consider the following problem of seeking for the function in $\mathcal{H}_K$ which interpolates g on points of D with a minimal norm:

$$\left\{ \begin{array}{l} \min_{v \in \mathcal{H}_K} \|v\|_{\mathcal{H}_K} \\ g(x_k) = v(x_k),\ k = 1, \ldots n_D. \end{array} \right.$$

The solution to this problem is the orthogonal projection of g on the subspace spanned by $(K_{x_1}, \ldots, K_{x_{n_D}})$, denoted $s_{K,D}(g)$. If we assume that the function $g = f - {}^tH\hat{\boldsymbol{\beta}}$ belongs to $\mathcal{H}_K$, then $s_{K,D}(g)$ is defined as

$$\begin{aligned} s_{K,D}(g(x)) &= {}^t\Sigma_{xD}^{\hat{\phi}}(\Sigma_{DD}^{\hat{\phi}})^{-1}(\mathbf{z}_D - H_D\hat{\boldsymbol{\beta}}) \\ &= {}^t\mathbf{u}(x)(\mathbf{z}_D - H_D\hat{\boldsymbol{\beta}}) = \sum_{k=1}^{n_D} u_k(x) g(x_k)\,, \end{aligned} \tag{7}$$

with ${}^t\mathbf{u}(x) = {}^t\Sigma_{xD}^{\hat{\phi}}(\Sigma_{DD}^{\hat{\phi}})^{-1}$ (same notations as in Section 3.1). This provides an approximation of f which is the same than the function m_D (5). The approximation m_D belongs to the RKHS assuming that the regression functions H belong to the RKHS, which is true for linear or low degree polynomial regressors H.

The kernel interpolation framework yields to an upper bound on the point-wise error of this approximation using the reproducing property and a Cauchy-Schwarz inequality. For any x, we have

$$\begin{aligned} |f(x) - m_D(x)| &= |g(x) - s_{K,D}(g(x))| \\ &\leq \ | < g, K_x^{\hat{\phi}} - \sum_{k=1}^{n_D} u_k(x) K_{x_k}^{\hat{\phi}} > | \\ &\leq \ \|g\|_{\mathcal{H}_K} \cdot \|K_x^{\hat{\phi}} - \sum_{k=1}^{n_D} u_k(x) K_{x_k}^{\hat{\phi}}\|_{\mathcal{H}_K} \\ &=: \ \|g\|_{\mathcal{H}_K} P_D(x)\,. \end{aligned} \tag{8}$$

The norm $\|g\|_{\mathcal{H}_K}$ is unknown and depends on f. The norm $P_D(x)$ does not depend on f (or g) but on the design of experiments D only. It holds that

$$P_D(x) \quad = \quad (K^{\hat{\phi}}(x,x) - {}^t\Sigma_{xD}^{\hat{\phi}}(\Sigma_{DD}^{\hat{\phi}})^{-1}\Sigma_{xD}^{\hat{\phi}}) = \frac{1}{\hat{\sigma}^2}C_D(x,x)\,.$$

Again, there exists a link with the Gaussian process framework: up to the parameter $\hat{\sigma}^2$, we obtain the variance function of the conditioned Gaussian process (6).

For some usual kernels, a uniform upper-bound on $P_D(x)$ is available as a function of

$$a_D = \sup_{x\in\mathcal{X}^{d+1}} \min_{1\leq k\leq n_D} \|x - x_k\|$$

where $\mathcal{X}$ is a bounded subspace of $\mathcal{R}$. The value of a_D is related to the coverage of the space $\mathcal{X}$ by the design of experiments. A design of experiments which minimizes this quantity is said to be minimax (Johnson et al., 1990). The point-wise upper bound is given in the following Proposition (Schaback, 1995):

Proposition 1. *Assume that the experimental design D is minimax in $\mathcal{X}$. Let $\mathcal{H}_K$ denote the RKHS associated to the kernel K which is assumed to be derived from a radial basis function as proposed in Wu and Schaback (1992). Assume that f lies in $\mathcal{H}_K$. Let m_D denote the kernel approximation of the function f obtained on the design D. Then the point-wise error $|f(x) - m_D(x)|$ is uniformly upper-bounded in $\mathcal{X}$ by*

$$|f(x) - m_D(x)| \leq \|g\|_{\mathcal{H}_K} P_D(x) \leq \|g\|_{\mathcal{H}_K} G_K(a_D)\,.$$

where the function G_K is defined on $\mathbb{R}^+$ and is such that $\lim_{a\to 0^+} G_K(a) = 0$.
Furthermore, if the regressors $H \in \mathcal{H}_K$, then there exists a constant $C > 0$ such that

$$|f(x) - m_D(x)| \leq C\|f\|_{\mathcal{H}_K} G_K(a_D)\,.$$

For instance, when using a Gaussian kernel $K^{\phi}(x,x') = e^{-\phi\|x-x'\|^2}$, the function G_K is $G_K(a) = Ce^{-\delta/a^2}$ where C and δ are constants depending on ϕ.

4 Three mixed meta-models

The estimation of the population parameter θ is performed on the meta-model approximation of the mixed model (1). For computational reasons, we introduce three mixed meta-models.

4.1 Complete mixed meta-model

Let us introduce the so-called complete mixed meta-model that integrates the meta-model approximation. The regression function f in (1) is approximated by $F_D(t,\psi) = m_D(t,\psi) + r(t,\psi)$:

$$\begin{array}{rcll} y_{ij} & = & F_D(t_{ij},\psi_i) + \sigma_\varepsilon\,\varepsilon_{ij},\ \varepsilon_{ij} \sim_{iid} \mathcal{N}(0,1), & (9)\\ \psi_i & \sim_{iid} & \mathcal{N}(\mu,\Omega), & \\ F_D(t,\psi) & = & m_D(t,\psi) + r(t,\psi), \quad \text{with}, & \\ r(t,\psi) & \sim & \mathcal{GP}(0, C_D(t,\psi;t,\psi)). & \end{array}$$

Whereas model (1) is homoscedastic (constant error variance), the mixed metamodel (9) is heteroscedastic. Let us emphasize that *this is not a standard heteroscedastic error model*. Indeed, we have:

$$y_{ij}|\psi_i \sim \mathcal{N}(m_D(t_{ij},\psi_i), \Gamma_D(t_{ij},\psi_i))$$

with

$$\Gamma_D(t,\psi) = \sigma_\varepsilon^2 + C_D(t,\psi;t,\psi)\,,$$

but the $(y_{ij}|\psi_i)_j$ are *not independent*, as well as the individual vectors $(\mathbf{y}_i)_i$. This is due to the fact that the $(r(t_{ij},\psi_i))_{ij}$ are realizations of the same Gaussian process. This is a major difference with approximations that have already been proposed in the literature (Donnet and Samson, 2007; Grenier et al., 2014). Especially, this complicates the implementation of the MCMC scheme.

We propose to estimate θ as the maximum of the likelihood of model (9). We denote

$$\mathbf{m}_D(\mathbf{t},\boldsymbol{\psi}) = (m_D(t_{ij},\psi_i))_{1\le i\le N, 1\le j\le n_i}$$

the vector of the approximate mean, evaluated on $(\mathbf{t},\boldsymbol{\psi}) = (t_{ij},\psi_i)_{1\le i\le N,1\le j\le n_i}$. Similarly, we denote

$$\mathbf{C}_D(\mathbf{t},\boldsymbol{\psi}) = (C_D(t_{ij},\psi_i;t_{i'j'},\psi_{i'}))_{1\le i,i'\le N, 1\le j,j'\le n_i}\,.$$

The likelihood of model (9) is then:

$$\begin{aligned} p_D(\mathbf{y};\boldsymbol{\theta}) \;=\; & \int \left\{ p(\boldsymbol{\psi};\boldsymbol{\theta})\, \frac{1}{(2\pi)^{n_{tot}/2}|\sigma_\varepsilon^2\, I_{n_{tot}} + \mathbf{C}_D(\mathbf{t},\boldsymbol{\psi})|^{1/2}} \right. \\ & \left. \exp\left(-\frac{1}{2}\, {}^t(\mathbf{y}-\mathbf{m}_D(\mathbf{t},\boldsymbol{\psi}))(\sigma_\varepsilon^2\, I_{n_{tot}} + \mathbf{C}_D(\mathbf{t},\boldsymbol{\psi}))^{-1}(\mathbf{y}-\mathbf{m}_D(\mathbf{t},\boldsymbol{\psi}))\right) d\boldsymbol{\psi} \right\}. \end{aligned} \tag{10}$$

This likelihood is not explicit because function $m_D(t_{ij},\psi_i)$ is not linear in ψ_i. As said previously, this likelihood cannot be simplified as a product of individual likelihoods because the $\mathbf{y}_i$ are not independent (the matrix $\mathbf{C}_D(\mathbf{t},\boldsymbol{\psi})$ is a full matrix). The corresponding estimation algorithm requires to invert this $n_{tot}\times n_{tot}$-matrix at each iteration (at least $N\times 2d$ per iteration), which is highly computationally intensive. Therefore, we introduce an intermediate mixed meta-model by considering only the diagonal of $\mathbf{C}_D(\mathbf{t},\boldsymbol{\psi})$.

4.2 Intermediate mixed meta-model

In the intermediate mixed meta-model, the regression function f is approximated by $m_D(t,\psi)+\bar{r}(t,\psi)$, where $\bar{r}(t,\psi)$ has a diagonal covariance matrix $\Lambda_{i,\psi_i} = \mathrm{diag}(C_D(\mathbf{t}_i,\psi_i))$:

$$\begin{aligned} y_{ij} &= m_D(t_{ij},\psi_i) + \bar{r}(t_{ij},\psi_i) + \sigma_\varepsilon\,\varepsilon_{ij}, \\ \varepsilon_{ij} &\sim_{iid} \mathcal{N}(0,1)\,, \\ \psi_i &\sim_{iid} \mathcal{N}(\mu,\Omega) \\ \bar{r}(\mathbf{t}_i,\psi_i) &\sim_{ind} \mathcal{GP}(0,\Lambda_{i,\psi_i} = \mathrm{diag}(C_D(\mathbf{t}_i,\psi_i))). \end{aligned} \tag{11}$$

The likelihood of model (11) is then:

$$\bar{p}_D(\mathbf{y};\boldsymbol{\theta}) = \prod_{i=1}^{N} \int \left\{ p(\psi_i;\boldsymbol{\theta}) \frac{1}{(2\pi)^{n_i/2} \prod_{j=1}^{n_i} (\sigma_\varepsilon^2 + C_D((t_{ij},\psi_i),(t_{ij},\psi_i)))^{1/2}} \right. \tag{12}$$
$$\left. \exp\left(-\frac{1}{2}\, {}^t(\mathbf{y}_i - m_D(\mathbf{t}_i,\psi_i))(\sigma_\varepsilon^2 I_{n_i} + \Lambda_{i,\psi_i})^{-1}(\mathbf{y}_i - m_D(\mathbf{t}_i,\psi_i)) \right) d\psi_i \right\}.$$

This form of the likelihood is separable with respect to ψ_i and can be written as a product over the individuals which are independent. The covariance matrix $\sigma_\varepsilon^2 I_{n_i} + \Lambda_{i,\psi_i}$ is diagonal and can be easily inverted. This will substantially reduce the computation time of the estimation method. However the intermediate model is heteroscedastic, and σ_ε might be more difficult to estimate than in the exact model. This is why we introduce a simpler mixed meta-model.

4.3 Simple mixed meta-model

The simple mixed meta-model neglects the error of approximation of the computer model. The regression function is then m_D:

$$\begin{aligned} y_{ij} &= m_D(t_{ij},\psi_i) + \sigma_\varepsilon\, \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0,1), \\ \psi_i &\sim_{iid} \mathcal{N}(\mu,\Omega). \end{aligned} \tag{13}$$

The simple mixed meta-model (13) has similar properties than model (1): it is homoscedastic (constant error variance), *the vectors* $(\mathbf{y}_i)_i$ *are independent and identically distributed*, and for each individual i, conditionally to ψ_i, the $(y_{ij})_j$ are independent. The likelihood of model (13) is given by:

$$\tilde{p}_D(\mathbf{y};\boldsymbol{\theta}) = \prod_{i=1}^{N} \int \left\{ p(\psi_i;\boldsymbol{\theta}) \frac{1}{(2\pi\sigma_\varepsilon^2)^{n_i/2}} \right. \tag{14}$$
$$\left. \exp\left(-\frac{1}{2}\, {}^t(\mathbf{y}_i - m_D(\mathbf{t}_i,\psi_i))(\sigma_\varepsilon^2 I_{n_i})^{-1}(\mathbf{y}_i - m_D(\mathbf{t}_i,\psi_i)) \right) d\psi_i \right\},$$

which has the same form than likelihood of model (1).

5 Population parameter estimation

Likelihoods of the mixed meta-models being not explicit, we resort to the family of EM algorithm to estimate the parameters θ, which is a classical approach for models with non-observed or incomplete data. We start with the SAEM algorithm for the exact mixed model and then for the three mixed meta-models.

5.1 Estimation for the exact mixed model

The objective is to maximize the likelihood $p(\mathbf{y};\theta)$ of the exact mixed model (1). Let us briefly cover the EM principle (Dempster et al., 1977). The complete data of the

mixed model is $(\mathbf{y}, \boldsymbol{\psi})$. The EM algorithm maximizes the $Q(\theta|\theta') = \mathbb{E}(L(\mathbf{y}, \boldsymbol{\psi}; \theta)|\mathbf{y}; \theta')$ function in 2 steps, where $L(\mathbf{y}, \boldsymbol{\psi}; \theta)$ is the log-likelihood of the complete data for the mixed model (1) and $\mathbb{E}$ is the expectation under the conditional distribution $p(\boldsymbol{\psi}|\mathbf{y}; \theta')$. At the k-th iteration, the E step is the evaluation of $Q_k(\theta) = Q(\theta \,|\, \widehat{\theta}^{(k-1)})$, whereas the M step updates $\widehat{\theta}^{(k-1)}$ by maximizing $Q_k(\theta)$. For cases with a non analytic E step, Delyon et al. (1999) introduce a stochastic version SAEM of the EM algorithm which evaluates the integral $Q_k(\theta)$ by a stochastic approximation procedure. The E step is then divided into a simulation step (S step) of the missing data $\boldsymbol{\psi}^{(k)}$ under the conditional distribution $p(\boldsymbol{\psi}|\mathbf{y}; \widehat{\theta}^{(k-1)})$ and a stochastic approximation step (SA step) of the conditional expectation, using $(\gamma_k)_{k\geq 0}$ a sequence of positive numbers decreasing to 0:

$$Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k(L(\mathbf{y}, \boldsymbol{\psi}^{(k)}; \theta) - Q_{k-1}(\theta)).$$

In cases where the simulation of the non-observed vector ψ cannot be directly performed, Kuhn and Lavielle (2005) propose to combine the SAEM algorithm with a Markov Chain Monte Carlo (MCMC) procedure. The idea is to simulate a Markov chain $\boldsymbol{\psi}^{(k)}$ by use of a Metropolis-Hastings (M-H) algorithm with $p(\boldsymbol{\psi}|\mathbf{y}; \widehat{\theta}^{(k-1)})$ as the unique stationary distribution.
The complete data likelihood $L(\mathbf{y}, \boldsymbol{\psi}; \theta)$ of the exact mixed model belongs to the regular curved exponential family:

$$p(\mathbf{y}, \boldsymbol{\psi}; \theta) = \exp\{-\nu(\theta) + < S(\boldsymbol{\psi}), \lambda(\theta) >\}$$

where $< \cdot, \cdot >$ denotes the scalar product, the minimal sufficient statistic $S(\mathbf{y}, \boldsymbol{\psi})$ takes its values in an open subset $\mathcal{S}$ of $\mathbb{R}^m$, ν_D and λ are functions of θ. Then the SA step reduces to approximate $\mathbb{E}\left[S(\mathbf{y}, \boldsymbol{\psi})|\widehat{\theta}^{(k-1)}\right]$ at each iteration by the value s_k. The sufficient statistics for the exact mixed model are classically $S_1(\mathbf{y}, \boldsymbol{\psi}) = \sum_{i=1}^N \psi_i$, $S_2(\mathbf{y}, \boldsymbol{\psi}) = \sum_{i=1}^N \psi_i \, {}^t\psi_i$ and
$S_3(\mathbf{y}, \boldsymbol{\psi}) = \sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - f(t_{ij}, \psi_i))^2$ (Samson et al., 2007). Then the M step is explicit and easy to implement. The convergence of the SAEM-MCMC algorithm has been proved when the complete data likelihood belongs to the regular curved exponential family and under additional assumptions (see Proposition 2). Thus the exponential family plays a crucial role to obtain an efficient algorithm.

5.2 Estimation for the simple mixed meta-model

The objective is to maximize the likelihood $\tilde{p}_D(\mathbf{y}; \theta)$ of the simple mixed meta-model (13). In the following, all the quantities referring to this approximate likelihood $\tilde{p}_D(\mathbf{y}; \theta)$ are indexed by D with a tilde symbol. The corresponding complete data likelihood $\tilde{L}_D(\mathbf{y}, \boldsymbol{\psi}; \theta)$ belongs to the regular curved exponential family with minimal sufficient statistics $\tilde{S}_D(\mathbf{y}, \boldsymbol{\psi})$, which are the same as the exact mixed model. In that model, the MCMC algorithm is easy to implement because of the independence of the observations of the individuals. More precisely, the SAEM-MCMC is described as follows.

Algorithm* 1.** ***(SAEM-MCMC algorithm for the simple mixed meta-model)
At iteration k, given the current values of the estimators $\hat{\mu}^{(k-1)}, \hat{\Omega}^{(k-1)}, \hat{\sigma}_\varepsilon^{2\,(k-1)}$:

Simulation step: *For each individual i separately and successively, update $\psi_i^{(k)}$ with m iterations of an MCMC procedure with $\tilde{p}_D(\psi_i|\mathbf{y}_i;\widehat{\theta}^{(k-1)})$ as stationary distribution:*

For $l = 1\ldots,m$, given a current value ψ_i^{l-1} for individual i:

- *Simulate a candidate ψ_i^c with a proposal distribution $q_{\widehat{\theta}^{(k-1)}}(\cdot|\psi_i^{l-1})$.*
- ***Meta-model step:*** *Evaluate the meta-model $m_D(t_{ij},\psi_i^c)$ for all $j = 1,\ldots,n_i$.*
- *The candidate is accepted, $\psi_i^l = \psi_i^c$, with probability $\tilde{\alpha}_i(\psi_i^c,\psi_i^{l-1})$; otherwise the candidate is rejected, $\psi_i^l = \psi_i^{l-1}$ with probability $1-\tilde{\alpha}_i(\psi_i^c,\psi_i^{l-1})$, where*

$$\tilde{\alpha}_i(\psi_i^c,\psi_i^{l-1}) = \min\left(\frac{\tilde{p}_D(\mathbf{y}_i|\psi_i^c;\widehat{\theta}^{(k-1)})p(\psi_i^c;\widehat{\theta}^{(k-1)})}{\tilde{p}_D(\mathbf{y}_i|\psi_i^{l-1};\widehat{\theta}^{(k-1)})p(\psi_i^{l-1};\widehat{\theta}^{(k-1)})}\frac{q_{\widehat{\theta}^{(k-1)}}(\psi_i^{l-1}|\psi_i^c)}{q_{\widehat{\theta}^{(k-1)}}(\psi_i^c|\psi_i^{l-1})},1\right).$$

Set $\psi_i^{(k)} = \psi_i^m$.

Stochastic Approximation step: *update the sufficient statistics:*

$$\begin{aligned}
s_{k,1} &= s_{k-1,1} + \gamma_k\left(\sum_{i=1}^N \psi_i^{(k)} - s_{k-1,1}\right),\\
s_{k,2} &= s_{k-1,2} + \gamma_k\left(\sum_{i=1}^N \psi_i^{(k)\,t}\psi_i^{\,(k)} - s_{k-1,2}\right),\\
s_{k,3} &= s_{k-1,3} + \gamma_k\left(\sum_{i=1}^N\sum_{j=1}^{n_i}(y_{ij}-m_D(t_{ij},\psi_i^{(k)}))^2 - s_{k-1,3}\right).
\end{aligned}$$

Maximisation step: *update the population parameters*

$$\widehat{\mu}^{(k)} = \frac{s_{k,1}}{N},\quad \widehat{\Omega}^{(k)} = \frac{s_{k,2}}{N} - \frac{s_{k,1}\,^t s_{k,1}}{N^2},\quad \widehat{\sigma_\varepsilon}^{2\,(k)} = \frac{s_{k,3}}{n_{tot}}.$$

5.3 Estimation for the intermediate mixed meta-model

In the following, all the quantities referring to the approximate likelihood $\bar{p}_D(\mathbf{y};\theta)$ of the intermediate mixed meta-model are indexed by D with a bar symbol.
This model belongs to the exponential family when the Gaussian process $\bar{r}$ is considered in the hidden states. Then the complete data of the intermediate mixed meta-model are $(\mathbf{y},\boldsymbol{\psi},\bar{\mathbf{r}})$ where $\bar{\mathbf{r}} = (\bar{r}(t_{ij},\psi_i))_{i=1,\ldots,N,j=1,\ldots,n_i}$. The complete log-likelihood is thus:

$$\begin{aligned}
\bar{L}_D(\mathbf{y},\boldsymbol{\psi},\bar{\mathbf{r}};\theta) &= \log\bar{p}_D(\mathbf{y}|\bar{\mathbf{r}},\boldsymbol{\psi};\theta) + \log\bar{p}_D(\bar{\mathbf{r}}|\boldsymbol{\psi};\theta) + \log p(\boldsymbol{\psi};\theta)\\
&= cst - \frac{n_{tot}}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2}\sum_{ij}\frac{(y_{ij}-m_D(t_{ij},\psi_i)-\bar{r}(t_{ij},\psi_i))^2}{\sigma_\varepsilon^2}\\
&\quad -\frac{1}{2}\sum_i\log(|\Lambda_{i,\psi_i}|) - \frac{1}{2}\sum_i{}^t\bar{\mathbf{r}}\Lambda_{i,\psi_i}^{-1}\bar{\mathbf{r}} - \frac{N}{2}\log(|\Omega|)\\
&\quad -\frac{1}{2}\sum_i{}^t(\psi_i-\mu)\Omega^{-1}(\psi_i-\mu),
\end{aligned}$$

where *cst* denotes a constant term. The E-step is the computation of

$$\begin{aligned} Q(\theta|\widehat{\theta}^{(k-1)}) &= \mathbb{E}(\bar{L}_D(\mathbf{y},\boldsymbol{\psi},\bar{\mathbf{r}};\theta)|\mathbf{y};\widehat{\theta}^{(k-1)}) \\ &= \int\int \log \bar{p}_D(\mathbf{y},\bar{\mathbf{r}},\boldsymbol{\psi};\theta)\bar{p}_D(\bar{\mathbf{r}},\boldsymbol{\psi}|\mathbf{y};\widehat{\theta}^{(k-1)})d\bar{\mathbf{r}}d\boldsymbol{\psi} \\ &= \int \left(\int \log \bar{p}_D(\mathbf{y},\bar{\mathbf{r}},\boldsymbol{\psi};\theta)\bar{p}_D(\bar{\mathbf{r}}|\mathbf{y},\boldsymbol{\psi};\widehat{\theta}^{(k-1)})d\bar{\mathbf{r}}\right)\bar{p}_D(\boldsymbol{\psi}|\mathbf{y};\widehat{\theta}^{(k-1)})d\boldsymbol{\psi}\,. \end{aligned}$$

The conditional distribution $\bar{p}_D(\bar{\mathbf{r}}_i|\mathbf{y}_i,\psi_i;\widehat{\theta}^{(k-1)})$ is explicit, Gaussian, with mean and covariance defined by

$$\begin{aligned} \bar{m}_{r,\psi_i}^{(k-1)} &= \bar{\Gamma}_{r,\psi_i}^{(k-1)}(\mathbf{y}_i - m_D(\psi_i))/\widehat{\sigma_\varepsilon^2}^{(k-1)}\,, \\ \bar{\Gamma}_{r,\psi_i}^{(k-1)} &= (1/\widehat{\sigma_\varepsilon^2}^{(k-1)} + \Lambda_{i,\psi_i}^{-1})^{-1}. \end{aligned}$$

Integrated with respect to $\bar{\mathbf{r}}$ inside $Q(\theta|\widehat{\theta}^{(k-1)})$ yields to

$$\begin{aligned} Q(\theta|\widehat{\theta}^{(k-1)}) = \int \Big[& -\frac{n_{tot}}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2}\sum_i \frac{\|\mathbf{y}_i - m_D(\mathbf{t}_i,\psi_i) - \bar{m}_{r,\psi_i}^{(k-1)}\|^2}{\sigma_\varepsilon^2} \\ & -\frac{1}{2\sigma_\varepsilon^2}\sum_i \mathrm{Tr}(\bar{\Gamma}_{r,\psi_i}^{(k-1)}) - \frac{1}{2}\sum_i \log|\Lambda_{i,\psi_i}| \\ & -\frac{1}{2}\sum_i {}^t\bar{m}_{r,\psi_i}^{(k-1)}\Lambda_{i,\psi_i}^{-1}\bar{m}_{r,\psi_i}^{(k-1)} - \frac{1}{2}\sum_i \mathrm{Tr}(\Lambda_{i,\psi_i}^{-1}\bar{\Gamma}_{r,\psi_i}^{(k-1)}) \\ & -\frac{N}{2}\log(|\Omega|) - \frac{1}{2}\sum_i {}^t(\psi_i-\mu)\Omega^{-1}(\psi_i-\mu)\Big] \\ & p(\boldsymbol{\psi}|\mathbf{y};\widehat{\theta}^{(k-1)})d\boldsymbol{\psi} + cst\,. \end{aligned}$$

Then the sufficient statistic corresponding to σ_ε^2 is changed to $\bar{S}_{D,3}^{(k-1)}(\mathbf{y},\boldsymbol{\psi},\mathbf{r}) = \sum_{i=1}^N \|\mathbf{y}_i - m_D(\psi_i) - \bar{m}_{r,\psi_i}^{(k-1)}\|^2 + \mathrm{Tr}(\bar{\Gamma}_{r,\psi_i}^{(k-1)})$. The simulation step is a standard one, which can be applied to each individual separately. The MCMC algorithm targets $\bar{p}_D(\psi_i|\mathbf{y}_i;\widehat{\theta}^{(k-1)})$ as stationary distribution, where the process $\bar{\mathbf{r}}_i$ has been integrated out. The acceptance probability only requires the knowledge of $\bar{p}_D(\mathbf{y}_i|\psi_i;\theta^{(k-1)})$ which is a Gaussian density with covariance matrix $\sigma_\varepsilon^2 I_{n_i} + \Lambda_{i,\psi_i}$. As this matrix is diagonal, its inversion at each iteration is fast. Finally, the SAEM-MCMC proceeds as follows:

Algorithm 2. (SAEM-MCMC algorithm for the intermediate mixed metamodel)

At iteration k, given the current values of the estimators $\hat{\mu}^{(k-1)}, \hat{\Omega}^{(k-1)}, \hat{\sigma}_\varepsilon^{2\,(k-1)}$*:*

S step: *For each individual i separately and successively, update* $\psi_i^{(k)}$ *with m iterations of an MCMC procedure with* $\bar{p}_D(\psi_i|\mathbf{y}_i;\widehat{\theta}^{(k-1)})$ *as stationary distribution.*

SA step: *update the sufficient statistics* $s_{k,1}$ *and* $s_{k,2}$ *as usual and update*

$$s_{k,3} = s_{k-1,3} + \gamma_k\left(\sum_{i=1}^N \|\mathbf{y}_i - m_D(\psi_i) - \bar{m}_{r,\psi_i^{(k)}}^{(k-1)}\|^2 + \mathrm{Tr}\left(\bar{\Gamma}_{r,\psi_i^{(k)}}^{(k-1)}\right) - s_{k-1,3}\right)$$

M step: *as usual.*

5.4 Estimation for the complete mixed meta-model

In the following, all the quantities referring to the approximate likelihood $p_D(\mathbf{y};\theta)$ of the complete mixed meta-model (9) are indexed by D.
The main difficulty comes from the fact that model (9) is heteroscedastic but not in a standard way: the conditional distributions of $\mathbf{y}_i|\psi_i$ are not independent and all the subjects have to be treated together. Similarly as the intermediate model, we consider the Gaussian process r in the complete data and we have to integrate out with respect to r to compute the function Q. The conditional distribution $p_D(\mathbf{r}|\mathbf{y},\boldsymbol{\psi};\widehat{\theta}^{(k-1)})$ is explicit, Gaussian, with mean and covariance defined by:

$$\begin{aligned} m_{r,\boldsymbol{\psi}}^{(k-1)} &= \Gamma_{r,\boldsymbol{\psi}}^{(k-1)}(\mathbf{y} - m_D(\boldsymbol{\psi}))/\widehat{\sigma_\varepsilon^2}^{(k-1)}, \\ \Gamma_{r,\boldsymbol{\psi}}^{(k-1)} &= (1/\widehat{\sigma_\varepsilon^2}^{(k-1)} + \mathbf{C}_D(\mathbf{t},\boldsymbol{\psi})^{-1})^{-1}. \end{aligned}$$

The matrix $\Gamma_{r,\boldsymbol{\psi}}^{(k-1)}$ has dimension $n_{tot} \times n_{tot}$ and cannot be split as it was the case with the intermediate model. Thus the inversion of Γ_r and C_D increases dramatically the computation time of the estimation algorithm.
Moreover, the MCMC step is also more complex. Indeed, the conditional distributions $p_D(\boldsymbol{\psi}|\mathbf{y})$ cannot be written as a product of individual conditional distributions. But the MCMC kernels are applied to each subject i successively. The corresponding target distribution is the conditional distribution $p_D(\psi_i|\mathbf{y},\boldsymbol{\psi}_{-i})$ where $\boldsymbol{\psi}_{-i} = (\psi_1,\ldots,\psi_{i-1},\psi_{i+1},\ldots,\psi_N)$ is the vector of individual parameters except individual i (with obvious notations when $i = 1$ or $i = N$) and not the distribution $p_D(\psi_i|\mathbf{y}_i)$ as in a standard heteroscedastic mixed model. This increases the difficulty of implementation of the MCMC: the whole covariance function $\mathbf{C}_D(\mathbf{t},\boldsymbol{\psi})$, evaluated at each point (t_{ij},ψ_i), has to be evaluated and inverted at each iteration of the MCMC scheme.

Algorithm 3. (SAEM-MCMC algorithm for the complete mixed meta-model)
At iteration k, given the current values of the estimators $\hat{\mu}^{(k-1)}, \hat{\Omega}^{(k-1)}, \hat{\sigma}_\varepsilon^{2\,(k-1)}$:

S step: *for each individual i successively, given the current values $\boldsymbol{\psi}_{-i}^{(k)} = (\psi_1^{(k)},\ldots,\psi_{i-1}^{(k)},\psi_{i+1}^{(k-1)},\ldots,\psi_N^{(k-1)})$ of all the other individuals, update $\psi_i^{(k)}$ with m iterations of an MCMC procedure with $p_D(\psi_i|\mathbf{y},\boldsymbol{\psi}_{-i}^{(k)};\widehat{\theta}^{(k-1)})$ as stationary distribution:*

For $l = 1\ldots,m$, given a current value ψ_i^{l-1} for individual i and a current vector $\boldsymbol{\psi}^{(k)l-1} = (\psi_1^{(k)},\ldots,\psi_{i-1}^{(k)},\psi_i^{l-1},\psi_{i+1}^{(k-1)},\ldots,\psi_N^{(k-1)})$ for all individuals:

- *Simulate a candidate ψ_i^c with a proposal distribution $q_{\widehat{\theta}^{(k-1)}}(\cdot|\psi_i^{l-1})$.*
- *Set $\boldsymbol{\psi}^c = (\psi_1^{(k)},\ldots,\psi_{i-1}^{(k)},\psi_i^c,\psi_{i+1}^{(k-1)},\ldots,\psi_N^{(k-1)})$.*
- ***Meta-model step:*** *For all $j = 1,\ldots,n_i$, evaluate the meta-model $m_D(t_{ij},\psi_i^c)$. For all subjects $i',i'' = 1,\ldots,N$ (including subject i) and all observations j',j'', evaluate the covariance functions $C_D(t_{i'j'},\boldsymbol{\psi}_{i'}^c;t_{i''j''},\boldsymbol{\psi}_{i''}^c)$ and invert the obtained matrix C_D.*

– *The candidate is accepted,* $\psi_i^l = \psi_i^c$*, with probability* $\alpha_i(\psi_i^c, \psi_i^{l-1})$*; otherwise* $\psi_i^l = \psi_i^{l-1}$ *with probability* $1 - \alpha_i(\psi_i^c, \psi_i^{l-1})$ *where*

$$\alpha(\psi_i^c, \psi_i^{l-1}) = \min\left(\frac{p_D(\mathbf{y}|\boldsymbol{\psi}^c; \widehat{\theta}^{(k-1)}) p(\psi_i^c; \widehat{\theta}^{(k-1)})}{p_D(\mathbf{y}|\boldsymbol{\psi}^{(k)l-1}; \widehat{\theta}^{(k-1)}) p(\psi_i^{(l-1)}; \widehat{\theta}^{(k-1)})} \frac{q_{\widehat{\theta}^{(k-1)}}(\psi_i^{(l-1)}|\psi_i^c)}{q_{\widehat{\theta}^{(k-1)}}(\psi_i^c|\psi_i^{(l-1)})}, 1\right).$$

Set $\psi_i^{(k)} = \psi_i^m$.

SA step: *update the sufficient statistics* $s_{k,1}$ *and* $s_{k,2}$ *as before and update:*

$$s_{k,3} = s_{k-1,3} + \gamma_k \left(\|\mathbf{y} - m_D(\boldsymbol{\psi}) - m_{r,\boldsymbol{\psi}^{(k)}}^{(k-1)}\|^2 + \mathrm{Tr}\left(\Gamma_{r,\boldsymbol{\psi}^{(k)}}^{(k-1)}\right) - s_{k-1,3} \right).$$

M step: *as usual.*

Let us emphasize that the MCMC in the S step is difficult to implement due to the heteroscedasticity of the complete mixed meta-model. This MCMC algorithm may have poor mixing properties because the vectors ψ_i are updated successively while they are highly correlated through this non-diagonal matrix $\mathbf{C}_D(\mathbf{t}, \boldsymbol{\psi})$. Another solution could be to design a proposal in the MCMC algorithm for the whole vector $\boldsymbol{\psi}$. However, such a proposal is quite complicated to construct since the dimension of $\boldsymbol{\psi}$ is high: $d \times N$.

5.5 Fisher Information matrix estimates

Using formula in Louis (1982) and estimation scheme proposed in Delyon et al. (1999), confidence intervals can be obtained on the parameters implementing a stochastic approximation scheme of the Fisher Information matrix. It is only necessary to approximate the gradient and the Hessian matrix of the log-likelihood of the complete data:

$$\log p(\boldsymbol{\psi}, \mathbf{y}; \boldsymbol{\theta}) = \log p(\mathbf{y}|\boldsymbol{\psi}; \boldsymbol{\theta}) + \log p(\boldsymbol{\psi}; \boldsymbol{\theta}). \tag{15}$$

Actually, $\log p(\mathbf{y}|\boldsymbol{\psi}; \boldsymbol{\theta})$ does not depend on μ et Ω, hence the gradient and Hessian computations are only about $\log p(\boldsymbol{\psi}; \boldsymbol{\theta})$ which is a multivariate normal $\mathcal{N}(\mu, \Omega)$. Thus, this implementation does not depend on the mixed model and remains the same for the standard mixed model and the three mixed meta-models.

6 Convergence of the SAEM algorithm to the maximum likelihood of the meta-model

Since the SAEM-MCMC algorithm is not applied to model (1), but to an approximate mixed model, it is not possible to prove the convergence of the algorithm toward a local maximum of the exact likelihood $p(\mathbf{y}; \boldsymbol{\theta})$. However, it is possible to apply the results of Kuhn and Lavielle (2005) for the three mixed meta-models. Hence, the algorithms converge toward a local maximum of the likelihood $p_D(\mathbf{y}; \theta)$, $\bar{p}_D(\mathbf{y}; \theta)$ and $\tilde{p}_D(\mathbf{y}; \theta)$ when applied to the complete, intermediate or the simple mixed meta-models (9), (11) and (13), respectively. This is given by Kuhn and Lavielle (2005) that we briefly recall, without detailing their assumptions (M1)-(M5) and (SAEM1)-(SAEM4).

Proposition 2 (Kuhn and Lavielle)**.** *Under assumptions (M1)-(M5) and (SAEM1)-(SAEM4) for the complete, intermediate or simple mixed meta-model, if the sequence* (s_k) *stays in a compact set, the SAEM algorithm produces a sequence* $(\hat{\theta}^{(k)})_{k\geq 1}$ *which converges to the (local) maximum of the approximate likelihood* $p_D(\mathbf{y};\theta)$*,* $\bar{p}_D(\mathbf{y};\theta)$ *or* $\tilde{p}_D(\mathbf{y};\theta)$*, respectively.*

Now we study the impact of the meta-model approximations on the likelihoods. Our goal is to obtain a uniform control on the distance between the likelihood of the exact model $p(\mathbf{y};\boldsymbol{\theta})$ and the likelihoods of the three mixed meta-models $p_D(\mathbf{y};\boldsymbol{\theta})$, $\bar{p}_D(\mathbf{y};\boldsymbol{\theta})$ and $\tilde{p}_D(\mathbf{y};\boldsymbol{\theta})$ as a function of the quality of the meta-model. We start by the simple mixed meta-model.

Proposition 3. *Let us consider the likelihoods* $p(\mathbf{y};\theta)$ *(2) of the mixed model (1) and* $\tilde{p}_D(\mathbf{y};\theta)$ *(14) of the simple mixed meta-model (13) associated to a minimax design* D*. Assume that the support of the distribution of* ψ *is compact. Assume that the functions* f *and* m_D *are uniformly bounded on the support of the distribution of* ψ*. Assume that* f *lies in the RKHS associated with the kernel* K *satisfying to the same hypotheses as in Proposition 1. Then, there exists a constant* $\tilde{C}_y$ *which depends only on* $\mathbf{y}$ *such that*

$$|p(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y};\theta)| \leq \tilde{C}_y \frac{n_{tot}}{\sigma_\varepsilon^{n_{tot}+2}} G_K(a_D)$$

where the function G_K *tends to* 0 *when* $a \to 0$ *(defined in Proposition 1) and the constant* a_D *is the covering distance of the design of experiments* D*.*

Recall that, when using a Gaussian kernel $K(x,x')$ for the meta-model approximation, the function G_K is defined by $G_K(a) = Ce^{-\delta/a^2}$. Then, to ensure that this covering distance is small, we need a global upper-bound, uniformly in ψ. This is true when the support of the distribution of ψ is compact. Under this assumption, we obtain that the covering distance $G_K(a_D)$ can be as small as required provided there is a sufficient number of points n_D in the design. Thus providing a rich design D during the pre-computation step allows to control as finely as we want the error induced on the likelihoods.
Now, we can study the distance between the three mixed meta-models.

Proposition 4. *Let us consider the likelihoods* $p_D(\mathbf{y};\theta)$ *(10) of the complete mixed meta-model (9),* $\bar{p}_D(\mathbf{y};\theta)$ *(12) of the intermediate mixed meta-model (13) and* $\tilde{p}_D(\mathbf{y};\theta)$ *(14) of the simple mixed meta-model (13) associated to a minimax design* D*.*
Under the same hypotheses as Proposition 3, there exist two constants C_y *and* $\bar{C}_y$*which depend only on* $\mathbf{y}$ *such that*

$$\begin{aligned} |p_D(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y})| &\leq C_y \frac{n_{tot}}{\sigma_\varepsilon^{n_{tot}+2}} G_K(a_D), \\ |\bar{p}_D(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y})| &\leq \bar{C}_y \frac{n_{tot}}{\sigma_\varepsilon^{n_{tot}+2}} G_K(a_D). \end{aligned}$$

Therefore, this guarantees a control between the likelihood of any of the mixed meta-model and the likelihood of the exact mixed model.
With regularity hypotheses on the Hessian matrix of each likelihood, results similar to Donnet and Samson (2007) can be obtained: The distance between the maximum of

Parameter		Intermediate meta-model			Simplified meta-model			Exact model
	n_D	25	50	100	25	50	100	
$\mu_{\log V}$	Bias	-0.508	-0.025	0.121	-0.483	-0.015	0.089	-0.390
	RMSE	0.048	0.043	0.042	0.048	0.042	0.041	0.063
	Cov.	92.8	93.3	93.3	92.1	93.6	93.8	86.3
$\mu_{\log k_a}$	Bias	-4.920	-1.389	-0.794	-4.872	-1.396	-0.904	-2.079
	RMSE	0.860	0.545	0.476	0.870	0.541	0.502	0.797
	Cov.	79.2	86.9	89.1	77.8	85.9	87.4	81.0
$\mu_{\log V_m}$	Bias	-2.067	-0.599	0.014	-1.930	-0.577	-0.138	-1.566
	RMSE	0.392	0.333	0.314	0.401	0.328	0.327	0.680
	Cov.	87.3	88.1	88.9	85.5	88.9	89.2	82.5
$\omega^2_{\log V}$	Bias	4.569	-2.408	-2.185	4.270	-2.108	-2.215	-4.393
	RMSE	6.487	5.526	5.276	6.359	5.445	5.287	9.461
	Cov.	94.5	92.2	91.9	94.8	93.3	93.5	84.8
$\omega^2_{\log k_a}$	Bias	1.755	-3.822	-6.797	1.935	-5.022	-6.799	-1.705
	RMSE	17.398	16.382	16.590	17.465	16.305	16.870	23.416
	Cov.	84.4	82.6	81.4	84.9	81.6	80.4	80.0
$\omega^2_{\log V_m}$	Bias	-33.721	-30.387	-30.408	-33.916	-29.946	-30.148	1.867
	RMSE	15.975	13.236	13.039	15.981	12.914	13.236	17.039
	Cov.	62.6	65.6	65.8	62.3	69.6	67.7	83.8
σ^2_ϵ	Biais	2.449	1.975	2.337	5.054	2.648	2.450	0.308
	RMSE	0.354	0.302	0.370	0.650	0.426	0.397	0.177

Table 1: Michaelis-Menten pharmacokinetic simulations: relative bias (%), relative MSE (%) and coverage rate (%) computed over 1000 simulations, with the intermediate meta-, the simple meta- and the exact mixed models. Meta-models are built with either $n_D = 25$, $n_D = 50$ or $n_D = 100$ design points. Coverage rate (Cov.) is the coverage rate of the 95% confidence interval based on the stochastic approximation of the Fisher matrix.

the exact likelihood $p(\mathbf{y};\theta)$ and the maximum of the approximate likelihoods $p_D(\mathbf{y};\theta)$, $\bar{p}_D(\mathbf{y};\theta)$ or $\tilde{p}_D(\mathbf{y};\theta)$ can be as small as we want, as soon as the design D is rich enough.

7 Simulation study

The objective of this study is to compare the main statistical properties of the estimation with the mixed meta-models and compare them to the exact mixed model. Two examples are illustrated below, using standard ODE pharmacokinetics (PK) models.

Parameter		Intermediate meta-model		Simple meta-model		Exact model
	n_D	50	100	50	100	
$\mu_{\log k_e}$	Bias	0.101	0.007	-0.320	0.007	0.003
	RMSE	0.004	0.005	0.005	0.005	0.005
	Cov.	94.2	94.4	90.6	94.6	93.9
$\mu_{\log k_a}$	Bias	-2.441	0.001	-8.380	0.008	-0.220
	RMSE	0.222	0.162	0.910	0.160	0.160
	Cov.	90.9	95.6	59.6	95.3	95.6
$\mu_{\log C_l}$	Bias	0.388	0.036	0.160	0.036	-0.004
	RMSE	0.004	0.003	0.003	0.003	0.003
	Cov.	87.6	95.1	93.4	94.7	94.9
$\omega^2_{\log k_e}$	Bias	-12.113	-2.745	-23.200	-2.780	-3.400
	RMSE	7.131	6.404	9.730	6.530	6.460
	Cov.	83.2	91.5	65.7	90.5	90.3
$\omega^2_{\log k_a}$	Bias	-20.485	-3.442	20.900	-3.320	-2.440
	RMSE	10.696	5.911	13.500	5.930	6.050
	Cov.	72.3	89.7	96.9	89.2	90.2
$\omega^2_{\log C_l}$	Bias	0.375	-1.145	-8.100	-1.100	-2.660
	RMSE	5.944	5.726	5.810	5.690	5.650
	Cov.	92.6	92.0	87.5	92.8	91.1
σ^2_ϵ	Biais	-45.262	-0.612	16.000	-0.009	-0.023
	RMSE	20.719	0.232	2.950	0.220	0.220

Table 2: One compartment simulations: relative bias (%), relative MSE (%) and coverage rate (%) computed over 1000 simulations, with the intermediate meta-, the simple meta- and the exact mixed models. Meta-models are built with either $n_D = 50$ or $n_D = 100$ design points. Coverage rate (Cov.) is the coverage rate of the 95% confidence interval based on the stochastic approximation of the Fisher matrix.

7.1 Michaelis-Menten pharmacokinetic model

7.1.1 Simulation settings

Let us now consider a one-compartment pharmacokinetic model, first order absorption and Michaelis Menten elimination. A dose D of a drug is given to a patient by intravenous bolus. The concentration of the drug in the body along time in then described by the following ordinary differential equation

$$\frac{df}{dt} = -\frac{V_m \cdot f}{k_m + f} + k_a \cdot \frac{D}{V} \cdot \exp(-k_a \cdot t), \quad f(t_0) = 0$$

where V is the volume of distribution, V_m is the maximum elimination rate (in amount per time unit), k_m is the Michaelis-Menten constant (in concentration unit) and k_a is the absorption constant. We consider $\log k_m$ as fixed to -2.5. The individual parameter ψ consists in $\log V$, $\log k_a$ and $\log V_m$. We assume a Gaussian distribution on the logarithm of these parameters with mean $(\mu_{\log V}, \mu_{\log k_a}, \mu_{\log V_m}) = (2.5, 1, -0.994)$ and a diagonal covariance matrix with terms $(\omega^2_{\log V}, \omega^2_{\log k_a}, \omega^2_{\log V_m}) = (0.09, 0.09, 0.09)$. Then a homoscedastic additive error model is simulated with a standard error $\sigma_\varepsilon = 0.1$. We implement the four algorithms: SAEM on the original mixed model, SAEM on the complete, intermediate and simple mixed meta-model. For the meta-model SAEM algorithms, we use successively $n_D = 25$, $n_D = 50$ and $n_D = 100$ number of points in the design of experiments for the Gaussian process emulator. The covariance is Gaussian, and the regression functions H are linear functions. More sophisticated choices in the regression functions and in the kernel can be made. However, our goal in this section is to illustrate in a quite simple case the efficiency of the combination of the Gaussian process emulation with the SAEM-MCMC algorithm. In the pre-computation step, for a given value ψ, the ODE solver provides $f(t, \psi)$ for each time of measurement. Thus, the design of numerical experiment is only built over the values of ψ and not t. We have compared two approaches one where t is considered as an additional input and the other where a meta-model is built for each time t. Based on the comparison of the quality of the approximations, we have kept the second one which is quite simple to deal with. However, more sophisticated approaches can be tested (see Rougier, 2008, for a review). The approximation is built over the domain: $[1.6; 3.3]$ for $\log V$, $[0; 2.1]$ for $\log k_a$ and $[-1.6; -0.3]$ for $\log V_m$. The starting values for the parameters are $\widehat{\mu}^{(0)}_{\log V} = 2$, $\widehat{\mu}^{(0)}_{\log k_a} = 0.5$, $\widehat{\mu}^{(0)}_{\log V_m} = -0.5$, $\widehat{\omega}^{2(0)}_{\log V} = \widehat{\omega}^{2(0)}_{\log k_a} = \widehat{\omega}^{2(0)}_{\log V_m} = 0.1$ and $\widehat{\sigma}^{(0)}_\varepsilon = 0.3$.

7.1.2 Results

The computation times for one run of SAEM (100 iterations of SAEM, with 15 iterations of MCMC at each SAEM iteration) were the following: around 15 min for the exact mixed model (requiring solving the ODE at each iteration of MCMC), around 30 min for the complete mixed meta-model with $n_D = 50$ (requiring inverting the $\mathbf{C}(\mathbf{t}, \boldsymbol{\psi})$ at each iteration of MCMC), around 80 sec for the intermediate model and 30 sec for the simple one. Therefore, in the following, we only present the results for the exact mixed model (as a benchmark) and the intermediate and simple mixed meta-models.
Relative bias and relative root mean square error (RMSE) are computed for each population parameter from 1000 replications and presented in Table 1. The 95% coverage

rates correspond to the coverage rate of the confidence interval on parameters based on the stochastic approximation of the Fisher Information matrix. In this example, the two meta-models have good performances even with only 25 points in the design of experiments and increasing n_D decreases the bias. The parameter $\omega_{\log V_m}$ is biased when using a meta-model (whatever n_D) while it is not with the exact model. This may be due to the error of approximation that is not completely taken into account.

7.2 First order pharmacokinetic model

7.2.1 Simulation settings

Let us consider a one-compartment PK model with first order absorption and elimination, with a dose D of a drug. The concentration of the drug in the body along time in then described by the following ordinary differential equation

$$\frac{df}{dt} = D\frac{k_a k_e}{C_l} exp(-k_a t) - k_e f, \quad f(t_0) = 0$$

where k_a and k_e are the absorption and elimination constants, C_l is the clearance. We consider the PK parameters of theophyllin (Pinheiro and Bates, 2000): $\log k_e = -2.52$, $\log k_a = 0.4$, $\log C_l = -3.22$. One dataset of 36 patients is simulated with a dose D=6 $mmol$ and measurements at time $t = 0.25, 0.5, 1, 2, 3.5, 5, 7, 9, 12$ hours. The random effects were simulated assuming a Gaussian distribution for the logarithm of the parameters with a diagonal variance-covariance matrix Ω with the following diagonal elements: $\omega^2_{\log k_e} = \omega^2_{\log k_a} = \omega^2_{\log C_l} = 0.01$. Then a homoscedastic additive error model is simulated with a standard error $\sigma_\varepsilon = 0.1$.
The same SAEM algorithms were run as in the first example. The Gaussian process emulators were built with $n_D = 50$ and $n_D = 100$ points, linear regression functions and a Gaussian covariance kernel, in the same fashion as in subsection 7.1.1. The domain where the approximation is built is $[-4; -1]$ for $\log k_e$, $[0; 2]$ for $\log k_a$ and $[-4.5; 2]$ for $\log C_l$. The starting values for the parameters are $\mu^{(0)}_{\log k_e} = -3$, $\mu^{(0)}_{\log k_a} = 1$, $\mu^{(0)}_{\log C_k} = -3$, $\omega^{2(0)}_{\log k_e} = \omega^{2(0)}_{\log k_a} = \omega^{2(0)}_{\log C_l} = 0.1$ and $\sigma^{2(0)}_\varepsilon = 0.3$.

7.2.2 Results

The computation times are the same as before. Relative bias, relative RMSE and coverage rate computed from 1000 replications are presented in Table 2. When the design of numerical experiments in the pre-computation step has 100 points, the estimates obtained with the mixed meta-models have similar performance to the ones obtained with the exact mixed model. With only 50 points in the design, the estimates with the mixed meta-models are less accurate, especially σ_ε with the intermediate mixed model. There is a clear improvement of the quality of the estimates with the intermediate mixed meta-model for the parameters concerning the means. Recall that the simple mixed meta-model neglects the approximation of the function f. Therefore, taking into account the errors of the approximation of the Gaussian process emulator in the model prevents from a systematic bias in the estimates. However, since the correlation between the Gaussian process emulator approximation errors are set to zero for the sake

of simplicity, the estimation of σ_ε may be less accurate but usually this parameter if of less interest.

8 Concluding remarks

In the case of a mixed model where the regression function is a non-analytical solution of an ODE or of a PDE, we proposed to build a so-called meta-model which is obtained thanks to a pre-computation step. It consists in running the ODE/PDE solver on a well chosen design of numerical experiments. Once this meta-model is obtained, we use it as a surrogate of the regression function in the estimation procedure which is based on a SAEM-MCMC algorithm. We derived three mixed meta-models depending on whether the additional source of uncertainty due to the approximation by the meta-model is taken into account totally, partially or not at all (complete, intermediate and simple mixed meta-model). In the complete mixed meta-model, there is a full covariance matrix accounting for dependencies induced by the meta-modeling errors which slows down the SAEM-MCMC algorithm. That is why we have renounced to test it in the simulation study. Further works are needed to design MCMC algorithm adapted to this case of non-independent individuals. In the intermediate and simple mixed meta-model, the individuals are still independent thus the SAEM-MCMC algorithm does not suffer from any computational burden. We showed examples where even with a very few design points for the meta-model approximation, the estimation results are very satisfactory. We also showed an example where the intermediate meta-model improves the quality of the estimates of the parameters especially those accounting for the mean of the population parameters.

Since the quality of the approximation provided by the meta-model directly depends on the density of the numerical design of experiments in the neighborhood of the input where the approximation is made, a sequential strategy for building an adaptive design reinforcing the meta-model where the SAEM-MCMC algorithm identifies likely region for the parameters should improve the estimates. However, this strategy would make the Markov property in the SAEM-MCMC procedure no longer to be valid. Therefore, there are theoretical questions which will be interesting to solve in order to ensure guarantees in this case.

Acknowledgements

Adeline Samson has been supported by the LabEx
PERSYVAL-Lab (ANR-11-LABX-0025-01). Les recherches menant aux présents résultats ont bénéficié d'un soutien financier du septième programme-cadre de l'Union européenne (7ePC/2007-2013) en vertu de la convention de subvention n 266638.

References

Barbillon, P., Celeux, G., Grimaud, A., Lefebvre, Y., and Rocquigny, E. D. (2011). Nonlinear methods for inverse statistical problems. *Computational Statistics & Data*

Analysis, 55(1):132 – 142.

Chatterjee, A., Guedj, J., and Perelson, A. S. (2012). Mathematical modelling of HCV infection: what can it teach us in the era of direct-acting antiviral agents? *Antivir Ther*, 17(6 Pt B):1171–1182.

Davidian, M. and Giltinan, D. (1995). *Nonlinear models to repeated measurement data*. Chapman and Hall.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27:94–128.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jr. R. Stat. Soc. B*, 39:1–38.

Donnet, S. and Samson, A. (2007). Estimation of parameters in incomplete data models defined by dynamical systems. *J. Statist. Plann. Inference*, 137:2815–2831.

Fang, K., Li, R., and Sudjianto, A. (2005). *Design and Modeling for Computer Experiments (Computer Science & Data Analysis)*. Chapman & Hall/CRC.

Fu, S., Celeux, G., Bousquet, N., and Couplet, M. (2014). Bayesian inference for inverse problems occurring in uncertainty analysis. *International Journal for Uncertainty Quantification*.

Grenier, E., Louvet, V., and Vigneaux, P. (2014). Parameter estimation in non-linear mixed effects models with SAEM algorithm: extension from ODE to PDE. *ESAIM: Mathematical Modelling and Numerical Analysis*.

Guedj, J., Thiébaut, R., and Commenges, D. (2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics*, 63:1198–2006.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131 – 148.

Koehler, J. R. and Owen, A. B. (1996). Computer experiments. In *Design and analysis of experiments*, volume 13 of *Handbook of Statist.*, pages 261–308. North-Holland, Amsterdam.

Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Statist. Data Anal.*, 49:1020–1038.

Lavielle, M., Samson, A., Fermin, A., and Mentre, F. (2011). Maximum likelihood estimation of long term HIV dynamic models and antiviral response. *Biometrics*, 67(1):250–259.

Lophaven, N., Nielsen, H., and Sondergaard, J. (2002). DACE, a matlab kriging toolbox. Technical Report IMM-TR-2002-12, DTU. Available to : http://www2.imm.dtu.dk/ hbn/dace/dace.pdf.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 44(2):226–233.

Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus*. Springer-Verlag.

Prasad, N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85:163–171.

Ribba, B., Kaloshi, G., Peyre, M., Ricard, D., Calvez, V., Tod, M., Cajavec-Bernard, B., Idbaih, A., Psimaras, D., Dainese, L., Pallud, J., Cartalat-Carel, S., Delattre, J., Honnorat, J., Grenier, E., and Ducray, F. (2012). A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clin Cancer Res*, 18:5071–5080.

Rougier, J. (2008). Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, 17(4):827–843.

Sacks, J., Schiller, S. B., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409–435.

Samson, A., Lavielle, M., and Mentré, F. (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Stat. Med.*, 26(27):4860–4875.

Santner, T. J., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.

Schaback, R. (1995). Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3(3):251–264.

Schaback, R. (2007). Kernel-based meshless methods. Technical report, Institute for Numerical and Applied Mathematics, Georg-August-University Goettingen.

Wei, G. C. G. and Tanner, M. A. (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika*, 77(3):649–652.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80(4):791–795.

Wu, H., Huang, Y., Acosta, E., Rosenkranz, S., Kuritzkes, D., Eron, J., Perelson, A., and Gerber, J. (2005). Modeling long-term HIV dynamics and antiretroviral response: effects of drug potency, pharmacokinetics, adherence, and drug resistance. *Journal of Acquired Immune Deficiency Syndromes*, 39:272–283.

Wu, Z.-M. and Schaback, R. (1992). Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal*, 13:13–27.

Appendix

Proof of proposition 3

We have

$$|p(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y};\theta)| \leq \int |p(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})|p(\boldsymbol{\psi})d\boldsymbol{\psi}\,.$$

Therefore, we start by studying $|p(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})|$:

$$
\begin{aligned}
& (2\pi\sigma_\varepsilon^2)^{n_{tot}/2}|p(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})| \\
= & \left|\exp\Big(-\frac{1}{2\sigma_\varepsilon^2}\sum_{ij}(y_{ij}-f(t_{ij},\psi_i))^2\Big) - \exp\Big(-\frac{1}{2\sigma_\varepsilon^2}\sum_{ij}(y_{ij}-m_D(t_{ij},\psi_i))^2\Big)\right| \\
= & \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\sum_{ij}(y_{ij}-f(t_{ij},\psi_i))^2\right) \\
& \times\left|1-\exp\Big(-\frac{1}{2\sigma_\varepsilon^2}\sum_{ij}\Big((y_{ij}-m_D(t_{ij},\psi_i))^2-(y_{ij}-f(t_{ij},\psi_i))^2\Big)\Big)\right| \\
\leq & \left|1-\exp\Big(-\frac{1}{2\sigma_\varepsilon^2}\sum_{ij}\Big(f(t_{ij},\psi_i)-m_D(t_{ij},\psi_i))(2y_{ij}-f(t_{ij},\psi_i)-m_D(t_{ij},\psi_i)\Big)\Big)\right|.
\end{aligned}
$$

Under the assumption that the functions f and m_D are uniformly bounded on the support of ψ, there exists a constant C_y which is uniform according to ψ such that $|2y_{ij} - f(t,\psi) - m_D(t,\psi)| \leq C_y$. Proposition 1 implies that the approximation error due to the metamodel $|f(t_{ij},\psi_i) - m_D(t_{ij},\psi_i)|$ is controlled by inequality (8):

$$|f(t_{ij},\psi_i) - m_D(t_{ij},\psi_i)| \leq \|f\|_{\mathcal{H}_K} G_K(a_D)\,.$$

Then there exists a constant C_y depending only on $\mathbf{y}$ such that

$$(2\pi\sigma_\varepsilon^2)^{n_{tot}/2}|p(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})| \leq C_y\frac{n_{tot}}{2\sigma_\varepsilon^2}\|f\|_{\mathcal{H}_K}G_K(a_D).$$

Finally

$$|p(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y};\theta)| \leq \frac{C_y}{(2\pi\sigma_\varepsilon^2)^{n_{tot}/2}}\frac{n_{tot}}{2\sigma_\varepsilon^2}\|f\|_{\mathcal{H}_K}G_K(a_D)\,.$$

□

8.1 Proof of proposition 4

We study the distance between the two likelihoods p_D and $\tilde{p}_D$. As in Proposition 3, we start by studying $|p_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})|$. We consider two Gaussian distributions with same expectations and different covariance matrix. Thus this distance is maximum

when $\sum(y_{ij} - m_D(t_{ij}, \psi_i))^2 = 0$. This yields

$$\begin{aligned}
& (2\pi)^{n_{tot}/2}|p_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})| \\
\leq & \left|\frac{1}{\sigma_\varepsilon^{n_{tot}}} - \frac{1}{\sqrt{|\sigma_\varepsilon^2 I_{n_{tot}} + \mathbf{C}_D|}}\right| \\
= & \frac{1}{\sigma_\varepsilon^{n_{tot}}}\left|1 - \frac{\sigma_\varepsilon^{n_{tot}}}{|\sigma_\varepsilon^2 I_{n_{tot}} + \mathbf{C}_D|^{1/2}}\right| \\
\leq & \frac{1}{\sigma_\varepsilon^{n_{tot}}}\left|1 - \frac{\sigma_\varepsilon^{n_{tot}}}{(\sigma_\varepsilon^2 + \frac{1}{n_{tot}}\sum_{ij} C_D(t_{ij}, \psi_i; t_{ij}, \psi_i))^{n_{tot}/2}}\right| \\
\leq & \frac{1}{\sigma_\varepsilon^{n_{tot}}}\left|1 - \frac{1}{(1 + \frac{1}{\sigma_\varepsilon^2 n_{tot}}\sum_{ij} C_D(t_{ij}, \psi_i; t_{ij}, \psi_i))^{n_{tot}/2}}\right|
\end{aligned}$$

where we use that the determinant, as a product of eigen values, is smaller than a function of the trace of the matrix. Thus, the sum is over the diagonal of the matrix $\mathbf{C}_D$ i.e. the sum of the variances. Then, we obtain that there exists a constant C such that

$$\begin{aligned}
|p_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta}) - \tilde{p}_D(\mathbf{y}|\boldsymbol{\psi};\boldsymbol{\theta})| & \leq C\frac{1}{\sigma_\varepsilon^{n_{tot}}}\left|\frac{1}{2\sigma_\varepsilon^2}\sum_{ij} C_D(t_{ij}, \psi_i; t_{ij}, \psi_i)\right| \\
& \leq C(2\pi)^{n_{tot}/2}\frac{n_{tot}}{\sigma_\varepsilon^{n_{tot}+2}} G_K(a_D)
\end{aligned}$$

where the last inequality holds using Proposition 1. Finally, we obtain

$$|p_D(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y};\theta)| \leq C_y\frac{n_{tot}}{\sigma_\varepsilon^{n_{tot}+2}} G_K(a_D).$$

The proof is similar for the distance between the two likelihoods $\bar{p}_D$ and $\tilde{p}_D$. □