Parameter estimation in nonlinear mixed effect models based on ordinary differential equations: An optimal control approach

Quentin Clairon, Chloé Pasin, Irene Balelli, Rodolphe Thiébaut, Mélanie Prague

Parameter estimation in nonlinear mixed effect models based on ordinary differential equations: An optimal control approach

Quentin Clairon

University of Bordeaux, Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR1219, F-33000 Bordeaux, France

Vaccine Research Institute, F-94000 Créteil, France

Chloé Pasin

Institute of Medical Virology, University of Zurich
Department of Infectious Diseases and Hospital Epidemiology, University Hospital, Zurich, Switzerland

Irene Balelli

Université Côte d'Azur, INRIA Sophia Antipolis, EPIONE Research Project, Valbonne, France.

Rodolphe Thiébaut and Mélanie Prague

University of Bordeaux, Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR1219, F-33000 Bordeaux, France
Vaccine Research Institute, F-94000 Créteil, France

Abstract

We present a parameter estimation method for nonlinear mixed effect models based on ordinary differential equations (NLME-ODEs). The method presented here aims at regularizing the estimation problem in presence of model misspecifications, practical identifiability issues and unknown initial conditions. For doing so, we define our estimator as the minimizer of a cost function which incorporates a possible gap between the assumed model at the population level and the specific individual dynamic. The cost function computation leads to formulate and solve optimal control problems at the subject level. This control

*Corresponding author
Email address: quentin.clairon@u-bordeaux.fr (Quentin Clairon)

theory approach allows to bypass the need to know or estimate initial conditions for each subject and it regularizes the estimation problem in presence of poorly identifiable parameters. Comparing to maximum likelihood, we show on simulation examples that our method improves estimation accuracy in possibly partially observed systems with unknown initial conditions or poorly identifiable parameters with or without model error. We conclude this work with a real application on antibody concentration data after vaccination against Ebola virus coming from phase 1 trials. We use the estimated model discrepancy at the subject level to analyze the presence of model misspecification.

Keywords: Dynamic population models, Ordinary differential equations, Optimal control theory, Clinical trial analysis

---

1. Introduction

ODE models are standard in population dynamics, epidemiology, virology, pharmacokinetics, or genetic regulation networks analysis due to their ability to describe the main mechanisms of interaction between different biological components of complex systems, their evolution in time and to provide reasonable approximations of stochastic dynamics [1, 2, 3]. In cases of experimental designs involving a large number of subjects and limited number of individual measurements, non-linear mixed-effect models may be more relevant than subject-by-subject model to gather information from the whole population while allowing between-individual variability. For example, clinical trials and pharmacokinetics/pharmacodynamics studies often fall into this category [4, 5]. Formally, we are interested in a population where the dynamics of the compartments of each subject $i \in [\![1,\, n]\!]$ is modeled by the d-dimensional ODE:

$$\begin{cases} \dot{x}_i(t) = f_{\theta,b_i}(t, x_i(t)) \\ x_i(0) = x_{i,0} \end{cases} \tag{1}$$

where f is a $d-$dimensional vector field, θ is a $p-$dimensional parameter, $b_i \sim N(0,\, \Psi)$ is a $q-$dimensional random effect where Ψ is a variance-covariance matrix, $x_{i,0} \sim \Gamma_i$ is the initial condition for subject i belonging to $\mathbb{R}^d$ where Γ_i

is a possibly subject dependent distribution. We denote $X_{\theta,b_i,x_{i,0}}$ the solution of (1) for a given set $(\theta, b_i, x_{i,0})$. In (1), we can also incorporate covariate functions z_i which are omitted here for the purpose of clarity.

Our goal is to estimate the true population parameters $(\theta^*,\ \Psi^*)$ as well as the true subject specific realizations $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ from partial and noisy observations coming from n subjects and described by the following observational model:

$$y_{ij} = CX_{\theta^*,b_i^*,x_{i,0}^*}(t_{ij}) + \epsilon_{ij}$$

where t_{ij} is the j-th measurement time-point for the $i-$th subject on the observation interval $[0,\,T]$. Here C is a $d^o \times d$ sized observation matrix emphasizing the potentially partially observed nature of the process and $\epsilon_{ij} \sim \sigma^* \times N(0, I_{d^o})$ is the measurement error. We also assume only a subset of the true initial condition $x_{i,0}^* \sim \Gamma_i^*$, denoted $x_{i,0}^{k*}$, is known, the other ones, denoted $x_{i,0}^{u*}$, being unknown. For the sake of clarity, we order the state variables as follows: $x_{i,0} = \left(\left(x_{i,0}^u\right)^T, \left(x_{i,0}^k\right)^T\right)^T$. We denote n_i the number of observations for the i-th subject, $\mathbf{y_i} = \{y_{ij}\}_{j\in[\![1,\,n_i]\!]}$ its corresponding set of observations and $\mathbf{y} = \{\mathbf{y_i}\}_{i\in[\![1,\,n]\!]}$ the set of all observations in the population. We consider a Bayesian framework where a priori knowkedge about population parameters can be available under the form of a prior $\mathbb{P}\left[\theta, \Psi\right]$.

Our problem belongs to the class parameter estimation problem in nonlinear mixed effect models. In this context, frequentist methods based on likelihood maximization (via different numerical procedures: Laplace approximation [6], Gaussian quadrature [5] or SAEM [7, 8]) and Bayesian ones aiming to reconstruct the a posteriori distribution or to derive the maximum a posteriori estimator (via MCMC algorithms [9, 10], importance sampling [11], approximation of the asymptotic posterior distribution [12]) have been proposed. In particular, dedicated methods/softwares using the structure of ODE models have been implemented to increase numerical stability and speed up convergence rate [13], to reduce the computational time [14] or to avoid the repeated model integration and estimation of initial conditions [15]. However, all the preceding methods face similar pitfalls due to specific features of population models based on ODEs

(with the exception of [15]):

1. They do not account for model misspecification presence, a common feature in ODE models used in biology. Indeed, the ODE modeling process suffers from model inadequacy, understood as the discrepancy between the mean model response and real world process, and residual variability issues, that is subject specific stochastic perturbations or missed elements which disappear by averaging over the whole population [16]. As examples of model inadequacy causes, one can think of ODE models used in epidemiology and virology which are derived by approximations where for instance, interactions are modeled by pairwise products while higher order terms and/or the influence of unknown/unmeasured external factors are neglected. Regarding residual variability, let us remind that biological processes are often stochastic and the justification of deterministic modeling comes from the approximation of stochastic processes [17, 18]. Moreover, in the context of population models, new sources of model uncertainties emerge. Firstly, error measurement in covariates z_i can leads to use a proxy function $\widehat{z}_i$ instead of z_i [10]. Secondly, the sequential nature of most inference methods leads to estimate $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ based on an approximation $\widehat{\theta}$ instead of θ^*. Thus, the structure of mixed-effect models spread measurement uncertainty into the mechanistic model structure during the estimation. It turns classical statistical uncertainties into model error causes. Estimation of θ^*, Ψ^* and $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ has to be done in presence of model error presence although it is known to dramatically impair the accuracy of methods which do not take it into account [19].

2. They have to estimate or make assumptions on $\left(x_{i,0}^{u*}, \Gamma_i^*\right)$. In ODE models, the initial conditions are generally nuisance parameters in the sense that knowing their values does not bring answers to the scientific questions which motivate the model construction but the estimation of the relevant parameters requires $x_{i,0}^*$ inference as well. For example, partially observed compartmental models used in pharmacokinetics/pharmacodynamics of-

ten involve unknown initial conditions which needs to be inferred to estimate the transmission rates between compartments which are the true parameters of interest. Unknown initial conditions imply either: assumptions on $\left(x_{i,0}^{u*}, \Gamma_i^*\right)$ values [4], another potential cause of model misspecifications, or the need to estimate them [20] which increases the optimization problem dimension and degrades estimation accuracy due to covariance effect between (θ^*, Ψ^*) and $\left(x_{i,0}^{u*}, \Gamma_i^*\right)$ estimate.

3. They can face accuracy degradation when the inverse problem of parameter estimation is ill-posed [2] due to practical identifiability issues. Ill-posedness in ODE models is often due to the geometry induced by the mapping $(\theta, b_i, x_{i,0}) \longmapsto CX_{\theta,b_i,x_{i,0}}$, where there can be a small number of relevant directions of variation skewed from the original parameter axes [21]. This problem, called sloppiness, often appears in ODE models used in biology [21, 22] and leads to an ill-conditioned Fisher Information Matrix. For maximum likelihood estimators this is a cause for high variance due to the Cramér-Rao bound. For Bayesian inference, it leads to a nearly singular asymptotic a posteriori distribution because of Bernstein–von Mises theorem (see [23] for the computational induced problems). Despite this problem is in part mitigated by the population approach which merges different subjects for estimating (θ^*, Ψ^*) and uses distribution of $b_i \mid \Psi$ as prior at the subject level [24], estimation accuracy can benefit from the use of regularization techniques.

These specific features of ODE-based population models limit the amount of information classic approaches can extract for estimation purposes from observations no matter their qualities or abundances. This advocates for the development of new estimation procedures. Approximate methods [25, 26] have already proven to be useful for ODE models facing these issues with observations coming from one subject. These approaches rely on an approximation of the solution of the original ODE (1) which is expected to have a smoother dependence with respect to the parameters and to relax the constraint imposed by the model

during the estimation process. The interest of such approximations is twofold. Firstly, they produce estimators with a better conditioned variance matrix comparing to classic likelihood based approaches. Secondly, they reduce the effect of model error on estimator accuracy. Also, some of these approximations bypass the need to estimate initial conditions [26, 27]. In this work, we develop a new estimation method specific to NLME-ODEs integrating such approximations to mitigate the effect of model misspecification and poorly identifiable parameters on estimation accuracy, while avoiding the need to estimate $x_{i,0}^{u*}$. We propose here a nested estimation procedure where population parameters $(\theta^*, \Psi^*, \sigma^*)$ are estimated through the maximization of an outer criterion. This requires in turn an estimator for the $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ obtained through the repeated optimization of inner criteria. We consider that the actual dynamic for each subject is described by a perturbed version of the ODE (1) where the added perturbation captures different sources of errors at the subject level [19, 28]. We control the magnitude of the acceptable perturbations by defining the inner criteria through a cost function balancing the two contrary objectives of fidelity to the observations and to the original model: to this end, we introduce a model discrepancy penalization term. The practical computation of the $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ estimators require to solve optimal control problems [29] known as tracking problems. This is done using a method inspired by [30]. In addition, our method does not need to know/infer $x_{i,0}^{u*}$ but can provide an estimator of it if needed with no additional computational costs.

In section 2, we present the estimation method and derive the inner and outer criteria. In section 3, we introduce the numerical method used for solving the control problems appearing in the inner criteria. In section 4, we analyse the asymptotic behavior of (θ^*, Ψ^*) estimator and derive an approximation of its asymptotic Variance-Covariance matrix from it. In section 5, we compare our approach with classic maximum likelihood in simulations. We then proceed to the real data analysis coming from clinical studies and a model of the antibody concentration dynamics following immunization with an Ebola vaccine in East African participants [31]. Section 7 concludes and discuss future extensions of

the method.

2. Construction of the estimator: definition of the inner and outer criteria

From now on, we use the following Choleski decomposition $\sigma^2\Psi^{-1} = \triangle^T\triangle$ and the parametrization (θ, Δ, σ) instead of (θ, Ψ, σ). This parametrization will allows us to enforce positiveness and symmetry of Ψ and to derive an explicit estimator of σ given a value for (θ, Δ). The norm $\|.\|_2$ will denote the classic Euclidean one defined by $\|b\|_2 = \sqrt{b^T b}$. Similarly as in the Expectation-Maximization (EM) algorithm, we estimate the population and individual parameters via a nested procedure:

- Estimation of $\widehat{b_i} := \widehat{b_i}(\theta, \Delta)$ for each subject i by minimization of an **inner criterion** g_i based on an approximation of $\max_{x^u_{0,i}} \ln \mathbb{P}(\mathbf{y_i}, b_i \mid x^u_{0,i}, \theta, \Delta, \sigma)$, the log joint-distribution of the data and the random effects profiled on unknown initial conditions.

- Estimation of (θ, Δ) by maximization of an **outer criterion** $G(\theta, \Delta, \sigma)$ based on an approximation of $\max_\sigma \max_b \ln \mathbb{P}[\theta, \Delta, \sigma, b \mid \mathbf{y}]$, the log joint-distribution of $(\theta, \Delta, \sigma, b)$ sequentially profiled on σ and b.

2.1. Inner criteria

In this section, we describe the procedure used to estimate the random effects $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ for a given (θ, Δ, σ) value. A straightforward approach would be to look for the minimum of the log joint-likelihood function of the data and $\left\{b_i, x^u_{0,i}\right\}$. However, we want to:

1. avoid estimation of unknown initial conditions,

2. allow for each subject an acceptable departure from the assumed model at the population level to take into account possible model misspecifications.

To solve the first point, we define our estimator as the maximizer of the joint conditional likelihood $\mathbb{P}(\mathbf{y_i}, b_i \mid x^u_{0,i}, \theta, \Delta, \sigma)$ profiled on the unknown initial condition. Since

$$\begin{aligned}\mathbb{P}(\mathbf{y_i}, b_i \mid x^u_{0,i}, \theta, \Delta, \sigma) &= \mathbb{P}(\mathbf{y_i} \mid b_i, x^u_{0,i}, \theta, \Delta, \sigma)\mathbb{P}(b_i \mid \theta, \Delta, \sigma) \\ &= (2\pi)^{-(d^o n_i + q)/2}\, \sigma^{-(d^o n_i + q)}\, |\triangle|\, e^{-0.5\left(\sum_j \left\|CX_{\theta,b_i,x_{0,i}}(t_{ij}) - y_{ij}\right\|_2^2 + b_i^T\left(\triangle^T\triangle\right)b_i\right)/\sigma^2}\end{aligned}$$

by using $\mathbb{P}(\mathbf{y_i} \mid b_i, \theta, \Delta, \sigma) = \prod_j \mathbb{P}(y_{ij} \mid b_i, \theta, \Delta, \sigma) = \prod_j (2\pi)^{-d^o/2}\, \sigma^{-d^o} e^{-0.5\left\|CX_{\theta,b_i,x_{0,i}}(t_{ij}) - y_{ij}\right\|_2^2/\sigma^2}$, $\mathbb{P}(b_i \mid \theta, \Delta, \sigma) = (2\pi)^{-q/2}\, |\Psi|^{-1/2}\, e^{-0.5 b_i^T \Psi^{-1} b_i}$ and $\sigma^{2q}\, |\Psi|^{-1} = |\triangle|^2$, a straightforward mixed-effect estimator would be $\widehat{b_i} = \arg\min_{b_i} \min_{x^u_{0,i}} \left\{\sum_j \left\|CX_{\theta,b_i,x_{0,i}}(t_{ij}) - y_{ij}\right\|_2^2 + \|\Delta b_i\|_2^2\right\}$ that is, the classic maximum likelihood criteria profiled on $x^u_{0,i}$. Concerning the second point, we allow perturbations comparing to the original model, by assuming that the dynamic of each subject i follows a perturbed version of ODE (1):

$$\begin{cases} \dot{x}_i(t) = f_{\theta,b_i}(t, x_i(t)) + Bu_i(t) \\ x_i(0) = x_{i,0} \end{cases} \tag{2}$$

with the addition of the forcing term $t \mapsto Bu_i(t)$ with B a $d \times d_u$ matrix and u_i a function in $L^2\left([0,T], \mathbb{R}^{d_u}\right)$. We denote $X_{\theta,b_i,x_{i,0},u_i}$ the solution of this new ODE (2). However, to ensure the possible perturbations remain small, we replace the data fitting criterion $\sum_j \left\|CX_{\theta,b_i,x_{0,i}}(t_{ij}) - y_{ij}\right\|_2^2$ by $\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)$ where $\mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U) = \sum_j \left\|CX_{\theta,b_i,x_{0,i},u_i}(t_{ij}) - y_{ij}\right\|_2^2 + \|u_i\|^2_{U,L^2}$ and $\|u_i\|^2_{U,L^2} = \int_0^T u_i(t)^T U u_i(t) dt$ is the weighted Euclidean norm. Therefore the magnitude of the allowed perturbations is controlled by a positive definite and symmetric weighting matrix U. Finally, we obtain:

$$\widehat{b_i}\,(\theta, \Delta) := \arg\min_{b_i} g_i(b_i \mid \theta, \Delta, U) \tag{3}$$

where $g_i(b_i \mid \theta, \Delta, U) = \min_{x^u_{0,i}} \left\{\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U) + \|\Delta b_i\|_2^2\right\}$.This requires to solve the infinite dimensional optimization problem $\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)$ in $L^2\left([0,T], \mathbb{R}^{d_u}\right)$. This problem belongs to the field of optimal control theory for which dedicated approaches have been developed to solve them [32, 33, 29]. Here we use the same method as in [27] which is detailed in section 3. The perturbation u_i corresponding to the solution of $\min_{x^u_{0,i}} \{\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)\}$

is named optimal control and denoted $\overline{u}_{i,\theta,b_i}$. The corresponding solution of (2) for $u_i := \overline{u}_{i,\theta,b_i}$ is denoted $\overline{X}_{\theta,b_i}$ and named optimal trajectory. In particular, $\overline{X}_{\theta,b_i}$ and $\overline{u}_{i,\theta,b_i}$ are respectively the subject specific state variable and perturbation such that:

$$g_i(b_i \mid \theta, \Delta, U) = \sum_j \left\| C\overline{X}_{\theta,b_i}(t_{ij}) - y_{ij} \right\|_2^2 + \|\overline{u}_{i,\theta,b_i}\|_{U,L^2}^2 + \|\Delta b_i\|_2^2 . \quad (4)$$

To incorporate possible model errors in the estimation process, e.g. due to subject specific exogenous perturbations, $\overline{X}_{\theta,b_i}$ is now assumed to be the subject specific regression function, defined as the state-variable which needs the smallest perturbation in order to get close to the observations.

Remark 2.1. *The definition of the optimal control $\overline{u}_{i,\theta,b_i}$ has an interpretation in terms of Bayesian inference in an infinite dimensional space. According to [34] (theorem 3.5 and Corollary 3.10), $\overline{u}_{i,\theta,b_i}$ is a maximum a posteriori estimator where the chosen prior measure is a centered Gaussian random field with the covariance operator determined by U.*

2.2. Outer criteria definition

We focus in this section on population parameter estimation. Classic approaches rely on maximum a posteriori distribution or the likelihood of the observations in which they get rid of the unknown subject specific parameters by taking the mean value of $\mathbb{P}[\theta, \Delta, \sigma, b \mid \mathbf{y}]$ or $\mathbb{P}[\mathbf{y} \mid \theta, \Delta, \sigma, b]$, $\mathbb{E}_b[\mathbb{P}[\theta, \Delta, \sigma, b \mid \mathbf{y}]]$ or $\mathbb{E}_b[\mathbb{P}[\mathbf{y} \mid \theta, \Delta, \sigma, b]]$ respectively, as outer criteria. This generally requires the numerical approximation of integrals of possibly high dimensions (the same as b), a source of approximation and computational issues [6]. To avoid this, we consider the random effects as nuisance parameters and rely on a classic profiling approach for $(\theta^*, \triangle^*)$ estimation [35]. Instead of taking the mean, we rely on the maximal value of the joint distribution with respect to b, or equivalently $\max_b \ln \mathbb{P}[\theta, \Delta, \sigma, b \mid \mathbf{y}]$. Bayes formula gives us $\mathbb{P}[\theta, \Delta, \sigma, b \mid \mathbf{y}] \propto \mathbb{P}[\mathbf{y} \mid \theta, \Delta, \sigma, b]\,\mathbb{P}[\theta, \Delta, \sigma, b]$. Since $\mathbb{P}[\theta, \Delta, \sigma, b] = \mathbb{P}[b \mid \theta, \Delta, \sigma]\,\mathbb{P}[\theta, \Delta]$, we get $\mathbb{P}[\theta, \Delta, \sigma, b \mid \mathbf{y}] \propto \left(\prod_i \mathbb{P}[\mathbf{y_i} \mid \theta, \Delta, \sigma, b_i]\,\mathbb{P}[b_i \mid \theta, \Delta, \sigma]\right)\mathbb{P}[\theta, \Delta]$ by conditional independence of subject by subject observations and subject specific parameters.

It follows that

$$\max_{b} \ln \mathbb{P}\left[\theta, \Delta, \sigma, b \mid \mathbf{y}\right] \propto \sum_{i} \max_{b_i} \left(\ln \mathbb{P}\left[\mathbf{y_i} \mid \theta, \Delta, \sigma, b_i\right] + \ln \mathbb{P}\left[b_i \mid \theta, \Delta, \sigma\right]\right) + \ln \mathbb{P}\left[\theta, \Delta\right]$$

and now we construct a suitable approximation of this last expression to estimate population parameters. As said in the previous section, we define the optimal trajectory $\overline{X}_{\theta,b_i}$ as the regression function for each subject. Therefore, we approximate $\mathbb{P}\left[\mathbf{y_i} \mid \theta, \Delta, \sigma, b_i\right]$ by $\widetilde{\mathbb{P}}\left[\mathbf{y_i} \mid \theta, \Delta, \sigma, b_i\right] \simeq \prod_j (2\pi)^{-d^o/2} \sigma^{-d^o} e^{-0.5\left\|C\overline{X}_{\theta,b_i}(t_{ij}) - y_{ij}\right\|_2^2/\sigma^2}$ and we derive from this:

$$\arg\max_{b_i} \left(\ln \widetilde{\mathbb{P}}\left[\mathbf{y_i} \mid \theta, \Delta, \sigma, b_i\right] + \ln \mathbb{P}\left[b_i \mid \theta, \Delta, \sigma\right]\right) = \arg\max_{b_i} \left(\sum_j \left\|C\overline{X}_{\theta,b_i}(t_{ij}) - y_{ij}\right\|_2^2 + \left\|\Delta b_i\right\|_2^2\right).$$

We regularize this estimation problem by approximating it via the addition of the Tikhonov penalization term on perturbation magnitude $\left\|\overline{u}_{i,\theta,b_i}\right\|_{U,L^2}^2$, thus $\arg\max_{b_i} \left(\ln \widetilde{\mathbb{P}}\left[\mathbf{y_i} \mid \theta, \Delta, \sigma, b_i\right] + \ln \mathbb{P}\left[b_i \mid \theta, \Delta, \sigma\right]\right) \simeq \arg\max_{b_i} g_i(b_i \mid \theta, \Delta, U) = \widehat{b_i}\left(\theta, \Delta\right)$ by using definition (4). From this, we derive

$$\overline{G}\left[\theta, \Delta, \sigma \mid \mathbf{y}\right] = \sum_i \left(\ln \widetilde{\mathbb{P}}\left[\mathbf{y_i} \mid \theta, \Delta, \sigma, \widehat{b_i}\left(\theta, \Delta\right)\right] + \ln \mathbb{P}\left[\widehat{b_i}\left(\theta, \Delta\right) \mid \theta, \Delta, \sigma\right]\right) + \ln \mathbb{P}\left[\theta, \Delta\right]$$

as suitable approximation. Moreover, for each (θ, Δ), the maximizer in σ^2 of $\overline{G}$ has a closed form expression:

$$\sigma^2\left(\theta, \Delta\right) = \frac{1}{\left(d^o \sum_i n_i + qn\right)} \sum_i \left(\sum_j \left\|C\overline{X}_{\theta,\widehat{b_i}(\theta,\Delta)}(t_{ij}) - y_{ij}\right\|_2^2 + \left\|\triangle \widehat{b_i}\left(\theta, \Delta\right)\right\|_2^2\right). \tag{5}$$

By using this expression for $\sigma^2\left(\theta, \Delta\right)$, we get that $\arg\max_{(\theta,\Delta)} \max_{\sigma^2} \overline{G}(\theta, \Delta, \sigma \mid \mathbf{y}) = \arg\max_{(\theta,\Delta)} \left\{G\left[\theta, \Delta \mid \mathbf{y}\right]\right\}$ where:

$$G\left[\theta, \Delta \mid \mathbf{y}\right] = -0.5 \left(d^o \sum_i n_i + qn\right) \ln\left(\sigma^2\left(\theta, \Delta\right)\right) + n \ln|\triangle| + \ln \mathbb{P}\left[\theta, \Delta\right].$$

Thus we can profile $\overline{G}$ on sigma σ^2 and define our estimator as:

$$\left(\widehat{\theta}, \widehat{\Delta}\right) \quad = \quad \arg\max_{(\theta,\Delta)} \left\{G\left[\theta, \Delta \mid \mathbf{y}\right]\right\} \tag{6}$$

to reduce the optimization problem dimension and focus on the structural parameters. An estimator of σ^* is obtained from there by computing $\sigma^2\left(\widehat{\theta}, \widehat{\Delta}\right)$

given by equation (5). The details of the outer criteria derivation are left in appendix A.

3. Numerical procedure for $\overline{u}_{i,\theta,b_i}$, $\overline{X}_{\theta,b_i}$ and g_i computation

In this section we explain how to get numerical approximations for $\min_{x^u_{0,i}} \{\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)\}$ and $\overline{u}_{i,\theta,b_i}$ linked to the perturbed ODE (2) which are then used to evaluate $\overline{X}_{\theta,b_i}$ and g_i. Firstly, we approximate g_i with a special type of optimal control problem, known as 'tracking problem', in a discrete time setting. Secondly, we adapt the method proposed by [30, 36] to obtain $\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)$. This presents the advantage of formulating $\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)$ as a quadratic form (or a sequence of quadratic forms) with respect to $x^u_{0,i}$. Thus, the computation of $\min_{x^u_{0,i}} \{\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)\}$ does not add any computational complexity comparing to $\min_{u_i} \mathcal{C}_i(b_i, x_{i,0}, u_i \mid \theta, U)$.

All it requires for the user is to specify a pseudo-linear representation of ODE (1), i.e a possibly state-dependent matrix $A_{\theta,b_i}(t, x_i(t))$ and state-independent vector $r_{\theta,b_i}(t)$ such that:

$$f_{\theta,b_i}(t, x_i(t)) = A_{\theta,b_i}(t, x_i(t))\, x_i(t) + r_{\theta,b_i}(t). \tag{7}$$

This formulation is crucial for solving the optimal control problem in a computationally efficient way. Linear models already fit in this formalism with $A_{\theta,b_i}(t) := A_{\theta,b_i}(t, x_i(t))$. For nonlinear models, the pseudo-linear representation is not unique but always exists [36] (in order to exploit this non-uniqueness as an additional degree of freedom, see [37] section 6).

3.1. g_i expression as an optimal control problem

We now rely on the pseudo-linear version of model (2):

$$\begin{cases} \dot{x}_i(t) = A_{\theta,b_i}(t, x_i(t))\, x_i(t) + r_{\theta,b_i}(t) + B u_i(t) \\ x_i(0) = x_{i,0} \end{cases} \tag{8}$$

and it discretized version:

$$\begin{cases} x_i(t_{k+1}^d) = \left(I_d + \Delta_k A_{\theta,b_i}(t_k^d, x_i(t_k^d))\right) x_i(t_k^d) + \Delta_k r_{\theta,b_i}(t_k^d) + B\Delta_k u_i(t_k^d) \\ x_i(0) = x_{i,0} \end{cases} \tag{9}$$

where the discretization is made at $K_i + 1$ time points $\left\{t_k^d\right\}_{0\leq k\leq K_i}$ with $t_0^d = 0$ and $t_{K_i}^d = t_{in_i}$. This set contains the observations time points i.e. $\{t_{ij}\}_{0\leq j\leq n_i} \subset \left\{t_k^d\right\}_{0\leq k\leq K_i}$, but can be bigger and patient specific, allowing to accurately approximate $X_{\theta,b_i,x_{i,0}}$ even when the observations are sparse on $[0,\, T]$. We define:

- $\Delta_k = t_{k+1}^d - t_k^d$, the mesh size between two discretization time-points,
- u_i^d the set of discrete values taken by the control at each time step i.e $u_i^d = \left(u(t_k^d), \ldots, u(t_{K_i-1}^d)\right)$,
- $w_k = 1_{\left\{\exists t_{ij} \,|\, t_{ij}=t_k^d\right\}}/(t_{k+1}^d - t_k^d)$ i.e. w_k is equal to $1/(t_{k+1}^d - t_k^d)$ if t_k^d corresponds to an observation time t_{ij}, otherwise $w_k = 0$,
- y_k^d=y_{ij} if $t_k^d = t_{ij}$, 0 otherwise,
- $X_{\theta,b_i,x_{i,0},u_i^d}^d$ the solution of (9).

The weights w_k and the set of extended data $\left\{y_k^d\right\}$ are introduced to have a vector of observations with the same length as $\left\{t_k^d\right\}_{0\leq k\leq K_i}$. We now introduce the discretized version of the cost $\mathcal{C}_i$ to be minimized:

$$\begin{aligned} \mathcal{C}_i^d(b_i, x_{i,0}, u_i^d \mid \theta, U) &= \sum_{j=0}^{n_i} \left\| C X_{\theta,b_i,x_{i,0},u_i^d}^d(t_{ij}) - y_{ij} \right\|_2^2 + \sum_{k=0}^{K_i-1} \triangle_k u_i(t_k)^T U u_i(t_k) \\ &= \left\| C X_{\theta,b_i,x_{i,0},u_i^d}^d(t_{in_i}) - y_{in_i} \right\|_2^2 \\ &+ \sum_{k=0}^{K_i-1} \triangle_k \left(\left\| C X_{\theta,b_i,x_{i,0},u_i^d}^d(t_k^d) - y_k^d \right\|_2^2 w_k + u_i(t_k)^T U u_i(t_k) \right). \end{aligned} \tag{10}$$

such that our inner criteria g_i can be approximated by:

$$g_i(b_i \mid \theta, \Delta, U) \simeq \min_{x_{0,i}^u} \min_{u_i^d} \mathcal{C}_i^d(b_i, x_{i,0}, u_i^d \mid \theta, U) + \|\Delta b_i\|_2^2 .$$

The solution of this discrete control problem will be denoted $\overline{u}_{i,\theta,b_i}^d$, and the related optimal trajectory $\overline{X}_{\theta,b_i}^d$: they will be used as numerical approximations of $\overline{u}_{i,\theta,b_i}$ and $\overline{X}_{\theta,b_i}$ respectively.

3.2. Numerical methods for solving the tracking problem

We present how to numerically obtain $\min_{x^u_{0,i}} \min_{u_i^d} \mathcal{C}_i^d(b_i, x_{i,0}, u_i^d \mid \theta, U)$ as well as the corresponding minimizer $\overline{u}^d_{i,\theta,b_i}$. We start with linear ODE models, then we consider nonlinear models following the steps detailed in [36].

3.2.1. Linear models

Here, we suppose $A_{\theta,b_i}(t) := A_{\theta,b_i}(t,x)$ in the pseudo-linear model formulation. For a given set $(\theta, b_i, x_{i,0})$, Linear-Quadratic theory ensures the existence and uniqueness of the optimal control $\overline{u}^d_{i,\theta,b_i}$ and that $\min_{x^u_{0,i}} \min_{u_i^d} \mathcal{C}_i^d(b_i, x_{i,0}, u_i^d \mid \theta, U)$ can be computed by solving a discrete final value problem, called the Riccati equation (e.g. [32, 33]).

Proposition 3.1. *Let us introduce $(R_{\theta,b_i,k},\, h_{\theta,b_i,k})$ for $1 \leq k \leq K_i$, the solution of the discrete Riccati equation:*

$$\left\{\begin{array}{rcl} R_{\theta,b_i,k} & = & R_{\theta,b_i,k+1} + \triangle_k w_k C^T C + \Delta_k \left(R_{\theta,b_i,k+1} A_{\theta,b_i}(t_k^d) + A_{\theta,b_i}(t_k^d)^T R_{\theta,b_i,k+1} \right) \\ & + & \triangle_k^2 A_{\theta,b_i}(t_k^d)^T R_{\theta,b_i,k+1} A_{\theta,b_i}(t_k^d) \\ & - & \triangle_k (I_d + \triangle_k A_{\theta,b_i}(t_k^d)^T) R_{\theta,b_i,k+1} BG(R_{\theta,b_i,k+1}) B^T R_{\theta,b_i,k+1} (I_d + \triangle_k A_{\theta,b_i}(t_k^d)) \\ h_{\theta,b_i,k} & = & h_{\theta,b_i,k+1} - \triangle_k w_k C^T y_k^d + \triangle_k A_{\theta,b_i}(t_k^d)^T h_{\theta,b_i,k+1} \\ & + & \Delta_k \left(I_d + \Delta_k A_{\theta,b_i}(t_k^d) \right)^T R_{\theta,b_i,k+1} r_{\theta,b_i}(t_k^d) \\ & - & \Delta_k (I_d + \Delta_k A_{\theta,b_i}(t_k^d))^T R_{\theta,b_i,k+1} BG(R_{\theta,b_i,k+1}) B^T \left(h_{\theta,b_i,k+1} + \Delta_k R_{\theta,b_i,k+1} r_{\theta,b_i}(t_k^d) \right) \end{array}\right. \tag{11}$$

with final condition $(R_{\theta,b_i,K_i},\, h_{\theta,b_i,K_i}) = (C^T C,\, -C^T y_{in_i})$ and $G(R_{\theta,b_i,k+1}) := \left[U + \triangle_k B^T R_{\theta,b_i,k+1} B \right]^{-1}$. Hence we get:

$$\begin{array}{l} g_i(b_i \mid \theta, \Delta, U) = \|\Delta b_i\|_2^2 + y_{in_i}^T y_{in_i} \\ - \left(R^{uk}_{\theta,b_i,0} x^k_{0,i} + h^u_{\theta,b_i,0} \right)^T \left(R^u_{\theta,b_i,0} \right)^{-1} \left(R^{uk}_{\theta,b_i,0} x^k_{0,i} + h^u_{\theta,b_i,0} \right) + \left(x^k_{0,i} \right)^T R^k_{\theta,b_i,0} x^k_{0,i} + 2 \left(h^k_{\theta,b_i,0} \right)^T x^k_{0,i} \\ + \sum_{k=0}^{K_m - 1} \triangle_k \left(w_k \left(y_k^d \right)^T y_k^d + \left(2 \left(h_{\theta,b_i,k+1} \right)^T + \Delta_k r_{\theta,b_i}(t_k^d)^T R_{\theta,b_i,k+1} \right) r_{\theta,b_i}(t_k^d) \right) \\ - \sum_{k=0}^{K_m - 1} \triangle_k \left(h_{\theta,b_i,k+1} + \Delta_k R_{\theta,b_i,k+1} r_{\theta,b_i}(t_k^d) \right)^T BG(R_{\theta,b_i,k+1}) B^T \left(h_{\theta,b_i,k+1} + \Delta_k R_{\theta,b_i,k+1} r_{\theta,b_i}(t_k^d) \right) \end{array} \tag{12}$$

where $R^u_{\theta,b_i,0}, R^{uk}_{\theta,b_i,0}, R^k_{\theta,b_i,0}, h^u_{\theta,b_i,0}$ and $h^k_{\theta,b_i,0}$ are given by the following decomposition $R_{\theta,b_i,0} := \begin{pmatrix} R^u_{\theta,b_i,0} & R^{uk}_{\theta,b_i,0} \\ \left(R^{uk}_{\theta,b_i,0} \right)^T & R^k_{\theta,b_i,0} \end{pmatrix}$ and $h_{\theta,b_i,0} := \begin{pmatrix} h^u_{\theta,b_i,0} & h^k_{\theta,b_i,0} \end{pmatrix}$.

Moreover, the control $\overline{u}^d_{i,\theta,b_i}$ *which minimizes the cost (10) is unique and equal to:*

$$\overline{u}^d_{i,\theta,b_i}(t^d_k) = -G(R_{\theta,b_i,k+1})B^T\left(R_{\theta,b_i,k+1}\left(\left(I_d + \triangle_k A_{\theta,b_i}(t^d_k)\right)\overline{X}^d_{\theta,b_i}(t^d_k) + \Delta_k r_{\theta,b_i}(t^d_k)\right) + h_{\theta,b_i,k+1}\right) \tag{13}$$

where $\overline{X}^d_{\theta,b_i}$ *is the optimal trajectory, i.e. the solution of the initial value problem:*

$$\begin{cases} \overline{X}^d_{\theta,b_i}(t^d_{k+1}) &= \left(I_d + \triangle_k A_{\theta,b_i}(t^d_k)\right)\overline{X}^d_{\theta,b_i}(t^d_k) + \Delta_k r_{\theta,b_i}(t^d_k) \\ &- \triangle_k BG(R_{\theta,b_i,k+1})B^T R_{\theta,b_i,k+1}\left(\left(I_d + \triangle_k A_{\theta,b_i}(t^d_k)\right)\overline{X}^d_{\theta,b_i}(t_k) + \Delta_k r_{\theta,b_i}(t^d_k)\right) \\ &- \triangle_k BG(R_{\theta,b_i,k+1})B^T h_{\theta,b_i,k+1} \end{cases} \tag{14}$$

with estimator $\widehat{x^u_{i,0}} = -\left(R^u_{\theta,b_i,0}\right)^{-1}\left(R^{uk}_{\theta,b_i,0}x^k_0 + h^u_{\theta,b_i,0}\right)$ *for* $x^u_{i,0}$.

3.2.2. Non-linear models

We adapt the method proposed by [36] to solve tracking problem for discrete time models. The outline of the method is the following: we replace the original problem (10) by a recursive sequence of problems, where the l-th one is defined by:

$$\begin{array}{rcl} \min_{u^d_i} \mathcal{C}^{d,l}_i(b_i, x_{i,0}, u^d_i \mid \theta, U) & := & \left\| CX^{d,l}_{\theta,b_i,x_{i,0},u^d_i}(t_{in_i}) - y_{in_i}\right\|^2_2 \\ & + & \sum_{k=0}^{K_i-1}\triangle_k\left(\left\|CX^{d,l}_{\theta,b_i,x_{i,0},u^d_i}(t^d_k) - y^d_k\right\|^2_2 w_k + u_i(t_k)^T U u_i(t_k)\right) \end{array}$$

$$\text{such that } \begin{cases} x_i(t^d_{k+1}) = \left(I_d + \Delta_k A_{\theta,b_i}(t^d_k, \overline{X}^{d,l-1}_{\theta,b_i}(t^d_k))\right)x_i(t^d_k) + \Delta_k r_{\theta,b_i}(t^d_k) + B\Delta_k u_i(t_k) \\ x_i(0) = x_{i,0}. \end{cases} \tag{15}$$

where $\overline{X}^{d,l-1}_{\theta,b_i}$ is the solution of problem (15) at iteration $l-1$. Thus, for each l, the matrix $A_{\theta,b_i}(t^d_k, \overline{X}^{d,l-1}_{\theta,b_i}(t^d_k))$ does not depend on x_i and the problem (15) is solved using proposition 3.1. We then construct the following algorithm:

1. Initialization phase: $\overline{X}^{u,d,0}_{\theta,b_i}(t^d_k) = x^{u,r}_{i,0}$ for all $k \in [\![0,\, n_i]\!]$ where $x^{u,r}_{i,0}$ is an arbitrary starting point for the unknown initial condition and $\overline{X}^{k,d,0}_{\theta,b_i}(t^d_k) = x^k_{i,0}$.

2. At iteration l: use proposition 3.1 to obtain $(R^l_{\theta,b_i},\, h^l_{\theta,b_i})$, $\overline{u}^{d,l}_{i,\theta,b_i}$,$\overline{X}^{d,l}_{\theta,b_i}$ and $g^l_i(b_i \mid \theta, \Delta, U)$.

3. If $\sum_{k=1}^{K_i} \left\| \overline{X_{\theta,b_i}^{d,l}}(t_k^d) - \overline{X_{\theta,b_i}^{d,l-1}}(t_k^d) \right\|_2^2 < \varepsilon_1$ and $\left| g_i^l(b_i \mid \theta, \Delta, U) - g_i^{l-1}(b_i \mid \theta, \Delta, U) \right| < \varepsilon_2$, then step 4; otherwise get back to step 2.

4. Set $(R_{\theta,b_i}, h_{\theta,b_i}) = (R_{\theta,b_i}^l, h_{\theta,b_i}^l)$, $\overline{u}_{i,\theta,b_i}^d = \overline{u}_{i,\theta,b_i}^{d,l}$, $\overline{X}_{\theta,b_i}^d = \overline{X}_{\theta,b_i}^{d,l}$ and $g_i(b_i \mid \theta, \Delta, U) = g_i^l(b_i \mid \theta, \Delta, U)$.

4. Asymptotic Variance-Covariance matrix estimator for $\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Delta}}\right)$

In this section, we derive an estimator of the asymptotic variance of $\left(\widehat{\theta}, \widehat{\Delta}\right)$. We highlight that in practice the matrix Δ is parametrized by a vector δ of dimension q', i.e $\triangle := \triangle(\delta)$. We give here a variance estimator of $\left(\widehat{\theta}, \widehat{\delta}\right)$. The variance of $\widehat{\Delta}$ can be obtained using classic delta-methods (see [38] chapter 3). We introduce the function $h(b_i, \theta, \Delta, \mathbf{y_i}) = \|\Delta b_i\|_2^2 + \sum_j \left\| C\overline{X}_{\theta,b_i}(t_{ij}) - y_{ij} \right\|_2^2$ in order to present sufficient conditions ensuring our estimator is asymptotically normal:

1. the function $\widetilde{G}\left[\theta, \Delta(\delta)\right] = -0.5\left(d^o \mathbb{E}\left[n_1\right] + q\right) \ln\left(\frac{\lim_n \frac{1}{n} \sum_i^n \mathbb{E}\left[h(\widehat{b}(\theta,\Delta(\delta)),\theta,\Delta(\delta),y_i)\right]}{d^o \mathbb{E}[n_1]+q} \right) + \ln\left|\Delta(\delta)\right|$ has a well separated minimum $\left(\overline{\theta}, \overline{\delta}\right)$ belonging to the interior of a compact $\Theta \times \Omega$,

2. the true initial condition distributions $\{\Gamma_i^*\}_{i \in [\![1,n]\!]}$ have finite variance and either

   (a) they are identicals $\Gamma_i^* = \Gamma^*$ or

   (b) are such that for $\nu = 0$ and $\nu = 1$:

   $$\lim_{n \longrightarrow \infty} \frac{1}{\left(V^{(\nu)}\right)^2} \mathbb{E}\left[\sum_{i=1}^n \left(\overline{h}^{(\nu)}(\mathbf{y_i}) - \mathbb{E}\left[\overline{h}^{(\nu)}(\mathbf{y_i}) \right] \right)^2 1_{\left\{ \overline{h}(\mathbf{y_i}) - \mathbb{E}\left[\overline{h}(\mathbf{y_i})\right] > \varepsilon \sqrt{V^{(\nu)}} \right\}} \right]$$

   where $\overline{h}^{(\nu)}(\mathbf{y_i}) = \frac{d^{(\nu)}h}{d^{(\nu)}(\theta,\delta)}(\widehat{b}_i(\overline{\theta}, \Delta(\overline{\delta})), \overline{\theta}, \Delta(\overline{\delta}), \mathbf{y_i})$ and $V^{(\nu)} = \sqrt{\sum_i Var(\overline{h}^{(\nu)}(\mathbf{y_i}))^2}$,

3. the subject specific number of observations $\{n_i\}_{i \in [\![1,n]\!]}$ are i.i.d and uniformly bounded,

4. for all possibles values (θ, b_i), the solution $X_{\theta,b_i,x^*_{0,i}}$ belongs to a compact χ of $\mathbb{R}^d$, and for all (t,θ,x), the mapping $b_i \longmapsto f_{\theta,b_i}(t,x)$ has a compact support Θ_b,

5. $(\theta, b_i, t, x) \longmapsto f_{\theta,b_i}(t,x)$ belongs to $C^1(\Theta \times \Theta_b \times [0,\, T] \times \chi, \mathbb{R}^d)$,

6. the matrices $\frac{\partial^2}{\partial^2 b_i} g_i(\widehat{b_i}\left(\overline{\theta}, \Delta(\overline{\delta})\right) \mid \overline{\theta}, \Delta(\overline{\delta}), U)$ and $\frac{\partial^2 \mathcal{C}_i}{\partial^2 x_{0,i}}(\widehat{b_i}\left(\overline{\theta}, \Delta(\overline{\delta})\right), \overline{X}_{\overline{\theta},\widehat{b}_i(\overline{\theta},\Delta(\overline{\delta}))}(0), \overline{u}_{\overline{\theta},\widehat{b}_i(\overline{\theta},\Delta(\overline{\delta}))} \mid \overline{\theta}, U)$ are of full rank almost surely for every sequence $\mathbf{y_i}$,

7. there is a neighborhood $\Theta_{\overline{\theta}}$ of $\overline{\theta}$ such that $(\theta, b_i, t, x) \longmapsto f_{\theta,b_i}(t,x) \in C^5(\Theta_{\overline{\theta}} \times \Theta_b \times [0,\, T] \times \chi, \mathbb{R}^d)$.

Condition 2b is here to ensure asymptotic normality for non identically distributed random variables via Lindeberg-Feller theorem. Conditions 1-4 are used to derive the consistency of our estimator toward $\left(\overline{\theta}, \overline{\delta}\right)$ by following classic steps for M-estimator by proving 1/the uniform convergence of our stochastic cost function to a deterministic one, 2/the existence of a well-separated minimum for this deterministic function [38]. Conditions 6-7 ensures that our cost function is asymptotically smooth enough in the vicinity of $\left(\overline{\theta}, \overline{\delta}\right)$ to proceed to a Taylor expansion and transfer the regularity of the cost function to the asymptotic behavior of $\sqrt{n}(\widehat{\theta} - \overline{\theta}, \widehat{\delta} - \overline{\delta})$. Less restrictive conditions can be established under which our estimator is still asymptotically normal, in particular regarding f_{θ,b_i} regularity with respect to t.

Theorem 4.1. *Under conditions 1-7, there is a model dependent lower bound λ such that if $\|U\|_2 > \lambda$ then the estimator $\left(\widehat{\theta}, \widehat{\delta}\right)$ is asymptotically normal and:*

$$\sqrt{n}(\widehat{\theta} - \overline{\theta}, \widehat{\delta} - \overline{\delta}) \rightsquigarrow N\left(0, A(\overline{\theta}, \overline{\delta})^{-1} B(\overline{\theta}, \overline{\delta}) \left(A(\overline{\theta}, \overline{\delta})^{-1}\right)^T\right)$$

where $A(\overline{\theta}, \overline{\delta}) = \lim_n \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \widetilde{J}(\overline{\theta}, \overline{\delta}, \mathbf{y_i})}{\partial(\theta, \delta)}\right]$, $B(\overline{\theta}, \overline{\delta}) = \lim_n \frac{1}{n} \left[\sum_i \widetilde{J}(\overline{\theta}, \overline{\delta}, \mathbf{y_i}) \widetilde{J}(\overline{\theta}, \overline{\delta}, \mathbf{y_i})^T\right]$
and the vector valued function $\widetilde{J}(\theta, \delta, \mathbf{y_i}) = \begin{pmatrix} \widetilde{J}_\theta(\theta, \delta, \mathbf{y_i}) \\ \widetilde{J}_\delta(\theta, \delta, \mathbf{y_i}) \end{pmatrix}$ *is given by:*

$$\widetilde{J}_\theta(\theta, \delta, \mathbf{y_i}) = \frac{d}{d\theta} h(\widehat{b}(\theta, \Delta(\delta)), \theta, \Delta(\delta), y_i)$$
$$\widetilde{J}_\delta(\theta, \delta, \mathbf{y_i}) = \frac{d}{d\delta} h(\widehat{b}_i(\theta, \Delta(\delta)), \theta, \Delta(\delta), y_i) - \frac{2}{d^o \mathbb{E}[n_1] + q} Tr\left(\triangle(\delta)^{-1} \frac{\partial \triangle(\delta)}{\partial \delta_k}\right) h(\widehat{b}_i(\theta, \Delta(\delta)), \theta, \Delta(\delta), y_i).$$

The proof is left in appendix C. The practical interest of this theorem is to give an estimator of Variance-Covariance $V(\widehat{\theta},\widehat{\delta}) \simeq \widehat{A}(\widehat{\theta},\widehat{\delta})^{-1}\widehat{B}(\widehat{\theta},\widehat{\delta})\left(\widehat{A}(\widehat{\theta},\widehat{\delta})^{-1}\right)^T/n$ with the matrices $\widehat{A}(\widehat{\theta},\widehat{\delta}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial J(\widehat{\theta},\widehat{\delta},\mathbf{y_i})}{\partial(\theta,\delta)}$ and $\widehat{B} = \frac{1}{n}\sum_{i=1}^{n} J(\widehat{\theta},\widehat{\delta},\mathbf{y_i})J(\widehat{\theta},\widehat{\delta},\mathbf{y_i})^T$ where the $(p+q)$ components of the vector valued function J for $1 \leq k \leq p$ are given by

$$J_k(\theta,\delta,\mathbf{y_i}) = \frac{d}{d\theta_k}h(\widehat{b}(\theta,\Delta(\delta)),\theta,\Delta(\delta),\mathbf{y_i})$$

and for $p+1 \leq k \leq p+q$ by

$$J_k(\theta,\delta,\mathbf{y_i}) = \frac{d}{d\delta_k}h(\widehat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),\mathbf{y_i}) - \frac{2n}{d^o\sum_i n_i + qn}Tr\left(\triangle(\delta)^{-1}\frac{\partial\triangle(\delta)}{\partial\delta_k}\right)h(\widehat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),\mathbf{y_i}).$$

Now that we have proven the existence of the variance matrix $V(\theta^*,\delta^*)$ such that $\widehat{\delta} - \delta^* \rightsquigarrow N(0, V(\theta^*,\delta^*))$, we can use the Delta method to derive the asymptotic normality of the original matrix $\Psi\left(\widehat{\delta}\right) = \sigma^2\left(\Delta(\widehat{\delta})^T\Delta(\widehat{\delta})\right)^{-1}$ as well as an estimator of its asymptotic variance. In the case of a diagonal matrix Ψ, composed of the elements $\left(\Psi_1^2,\dots\Psi_q^2\right)$ and of the parametrization $\triangle(\delta) = \begin{pmatrix} e^{\delta_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{\delta_q} \end{pmatrix}$ used in section 5, we derive:

$$\begin{pmatrix} \Psi_1(\widehat{\delta}) \\ \vdots \\ \Psi_q(\widehat{\delta}) \end{pmatrix} - \begin{pmatrix} \Psi_1(\delta^*) \\ \vdots \\ \Psi_q(\delta^*) \end{pmatrix} \rightsquigarrow N\left(0, \sigma^2\begin{pmatrix} e^{-\delta_1^*} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{-\delta_q^*} \end{pmatrix} V(\theta^*,\delta^*) \begin{pmatrix} e^{-\delta_1^*} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{-\delta_q^*} \end{pmatrix}\right).$$

Remark 4.1. *The previous theorem 4.1 states that we retrieve a parametric convergence rate despite a number of nuisance parameter increasing with the number of subjects. We avoid the pitfall described in [39] for profiled methods, thanks to the i.i.d structure of the nuisance parameters. This allows us to prevent bias accumulation for score functions among subjects by using the central limit theorem.*

5. Results on simulated data

We compare the accuracy of our approach with maximum likelihood (ML) in different models and experimental designs reflecting the problems exposed in in-

troduction, that is estimation in 1/presence of model error, 2/partially observed framework with unknown initial conditions and 3/presence of poorly identifiable parameters. For the fairness of comparison with ML, we choose a non-informative one i.e. $\ln \mathbb{P}[\theta, \Delta] = 0$ for our method throughout this section. If the ODE (1) has an analytical solution, the ML estimator is computed via SAEM algorithm (SAEMIX package [7]). Otherwise, it is done via a restricted likelihood method dedicated to ODE models implemented in the nlmeODE package [13]. We proceed to Monte-Carlo simulations based on $N_{MC} = 100$ runs. At each run, we generate n_i observations coming from n subjects on an observation interval $[0,\, T]$ with Gaussian measurement noise of standard deviation σ^*. From these data, we estimate θ^*, Ψ^* and b_i^* with both estimation methods. We quantify the accuracy of each entry $\widehat{\psi_p}$ of the population parameters estimate $\widehat{\psi} = \left(\widehat{\theta}, \widehat{\Psi}\right)$ via Monte-Carlo computation of the bias $Bias(\widehat{\psi_p}) = \mathbb{E}\left[\widehat{\psi_p}\right] - \psi_p^*$, the empirical variance $V^e(\widehat{\psi_p}) = \mathbb{E}\left[\left(\mathbb{E}\left[\widehat{\psi_p}\right] - \psi_p^*\right)^2\right]$, the mean square error $MSE(\widehat{\psi_p}) = Bias(\widehat{\psi_p})^2 + V_{emp}(\widehat{\psi_p})$, the estimated variance $\widehat{V}\left(\widehat{\psi_p}\right)$ as well as the coverage rate of the 95%-confidence interval derived from it. This coverage rate, denoted CR in the following results, corresponds to the frequency at which the interval $\left[\widehat{\psi_p} \pm z_{0.975}\sqrt{\widehat{V}\left(\widehat{\psi_p}\right)}\right]$ contains ψ_p^* with $z_{0.975}$ the $0.975-$quantile of the centered Gaussian law. We compute the previous quantities for the normalized values $\widehat{\psi}_p^{norm} := \frac{\widehat{\psi_p}}{\psi_p^*}$ to make relevant comparisons among parameters with different order of magnitude. For b_i^*, we estimate the mean square error $MSE(\widehat{b_i}) = \mathbb{E}\left[\left\|b_i^* - \widehat{b_i}\right\|_2^2\right]$. For each subsequent examples, we give the results for $n = 50$ and present in appendix B the case $n = 20$ to analyze the evolution of each estimator accuracy with respect to data sparsity.

For our method, we need to select U balancing model and data fidelity in the inner criteria (4). We use the method presented in [40] to compute $\mathrm{EP}_i(U)$, the prediction error for the subject i corresponding to the estimators $\widehat{\theta_U}$, $\left\{\widehat{b_{i,U}}\right\}_{i\in[\![1,\,n]\!]}$ obtained for a given matrix U. From this, we compute $\mathrm{EP}(U) = \sum_i \mathrm{EP}_i(U)$ the global prediction error for the whole population. We retain the matrix U minimizing EP among a trial of tested values and we de-

note $\widehat{\theta}, \widehat{\Psi}, \left\{\widehat{b_{i,}}\right\}_{i\in[\![1,\,n]\!]}$ the corresponding estimator. In the following, we use the superscript ML to denote the ML estimator.

For solving the optimization problems required for computing our inner and outer criteria, we use the Nelder-Mead algorithm implemented in the optimr package [41]. All optimization algorithms used here require a starting guess value. We start from the true parameter value for each of them. By doing so, we aim to do not mix two distinct problems: 1)the numerical stability of the estimation procedures, 2)the intrinsic accuracy of the different estimators. These two problems are correlated, but we aim to adress only the latter which corresponds to the issues raised in introduction. Still, we check on preliminary analysis that local minima presence was not an issue in the vicinity of $(\theta^*, \triangle^*)$ by testing different starting points for all methods. No problem appears for our method and SAEMIX. A negligible number of non convergence cases appear for nlmODE which have been discarded thanks to the convergence criteria embedded in the package.

5.1. Partially observed linear model

We consider the population model where each subject i follows the ODE:

$$\begin{cases} \dot{X}_{1,i} = \phi_{2,i}X_{2,i} - \phi_{1,i}X_{1,i} \\ \dot{X}_{2,i} = -\phi_{2,i}X_{2,i} \\ (X_{1,i}(0), X_{2,i}(0)) = (x_{1,0}, x_{2,0,i}) \end{cases} \tag{16}$$

with the following parametrization: $\log(\phi_{1,i}) = \theta_1 + b_i$ and $\log(\phi_{2,i}) = \theta_2$ where $b_i \sim N(0, \Psi)$. The true population parameter values are $\theta^* = (\theta_1^*, \theta_2^*) = (\log(0.5), \log(2))$ and $\Psi^* = 0.5^2$ and we are in a partially observed framework where only $X_{1,i}$ is accessible. The true initial conditions are distributed with $x_{1,0,i}^* \sim N(2,\, 0.5)$ and $x_{2,0,i}^* \sim N(3,\, 1)$. ODE (16) has an analytic solution given by $X_{1,i}(t) = e^{-\phi_{1,i}t}(x_{1,0} + \frac{x_{2,0}\phi_{2,i}}{\phi_{1,i}-\phi_{2,i}}(e^{(\phi_{1,i}-\phi_{2,i})t} - 1))$ for its first component which will be used for estimation with the SAEMIX package. We generate $n_i = 11$ observations per subject on $[0,\, T] = [0,\, 10]$ with measurement noise of standard deviation $\sigma = 0.05$. An example of observations and corresponding solution is plotted in figure 1.

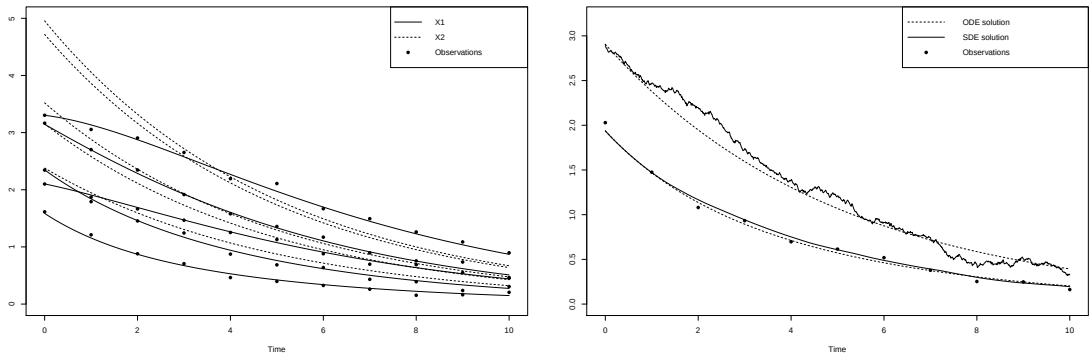

Figure 1: Left: Examples of (16) solutions and corresponding observations. Right: Solution of (16) and a realization of (17) for the same parameter values.

We want to investigate the impact of initial condition, especially the unobserved one $x^*_{2,0,i}$, on the ML estimator accuracy. Indeed, our method does not need to estimate $x^*_{2,0,i}$ and thus no additional difficulties appear in this partially observed framework. For the ML, however, it is nuisance subject-specific parameter that should be estimated and for which no observations are available. For this, we compute $\widehat{\theta}^{ML}_{x_0}$, $\widehat{\theta}^{ML}_{x_{0,2}}$ and $\widehat{\theta}^{ML}$ the ML estimator respectively when: 1) both initial conditions are perfectly known, 2) $x^*_{1,0,i}$ is replaced by the measured value, 3)in addition $x^*_{2,0,i}$ has to be estimated.

5.1.1. Correct model case

We present the estimation results in table 1. For ML, the results are goods in terms of accuracy and consistent in terms of asymptotic confidence interval coverage rate when both initial conditions are known: 95% for θ_1 and θ_2 in accordance with theoretical results. However, there is a significant drop in accuracy when $x^*_{2,0,i}$ has to be estimated, especially for θ_2. In particular, the coverage rate drops to 86% and 80% for θ_1 and θ_2 respectively. Interestingly, ML inaccuracy is driven by bias and under-estimated variance when initial conditions are not known. In this case our method provides a relevant alternative: it gives accurate estimations with a good coverage rate for all parameters while

		Well-specified						Misspecified					
		MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i	MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i
θ_1	$\widehat{\theta}^{ML}_{x_0}$	0.01	0.01	0.01	0.01	0.95		0.01	4e-4	0.01	0.01	0.91	
	$\widehat{\theta}^{ML}_{x_{0,2}}$	0.01	0.01	0.01	0.01	0.94		0.01	-3e-4	0.01	1e-4	0.89	
	$\widehat{\theta}^{ML}$	0.04	-0.04	0.04	0.01	0.86		0.05	0.02	0.05	0.01	0.81	
	$\widehat{\theta}$	**5e-3**	**8e-3**	**8e-3**	**1e-2**	**0.97**		**0.01**	**-8e-3**	**7e-3**	**0.05**	**0.97**	
θ_2	$\widehat{\theta}^{ML}_{x_0}$	4e-5	1e-3	4e-5	4e-5	0.95		1e-4	-1e-3	1e-4	1e-4	0.83	
	$\widehat{\theta}^{ML}_{x_{0,2}}$	6e-5	1e-3	6e-5	8e-5	0.94		1e-4	-1e-3	2e-4	0.01	0.82	
	$\widehat{\theta}^{ML}$	4e-3	-0.01	3e-3	1e-4	0.80		4e-3	-2e-3	4e-3	2e-4	0.63	
	$\widehat{\theta}$	**5e-5**	**2e-3**	**4e-5**	**4e-5**	**0.93**		**1e-4**	**2e-5**	**1e-4**	**1e-4**	**0.92**	
Ψ	$\widehat{\theta}^{ML}_{x_0}$	0.01	-0.03	0.01	7e-3	1	5e-3	0.01	-0.003	0.01	0.01	1	0.01
	$\widehat{\theta}^{ML}_{x_{0,2}}$	0.02	-0.03	0.01	7e-3	1	5e-3	0.01	-0.005	0.01	0.01	1	0.01
	$\widehat{\theta}^{ML}$	0.05	0.17	0.02	0.02	1	0.10	0.09	0.21	0.04	0.03	1	0.12
	$\widehat{\theta}$	**0.01**	**-0.01**	**0.01**	**0.01**	**0.92**	**0.01**	**0.02**	**-0.02**	**0.02**	**0.01**	**0.90**	**0.01**

Table 1: Results of estimation for model (16). The different subscripts stand for the following estimation scenarios: 1)x_0 when both initial conditions are set to $\left(x^*_{0,1}, x^*_{0,2}\right)$, 2)$x_{0,2}$ when $x_{0,i}$ is set to $y_{i,0}$ and $x_{0,2}$ to $x^*_{0,2}$, 3/absence of subscript when $x_{0,i}$ is set to $y_{i,0}$ and $x_{0,2}$ is estimated. Results from our method are in bold.

avoiding the estimation of $x^*_{2,0,i}$. Estimation of individual random effects is also more accurate with our method, with a decrease of more than 90% of MSE for b_i comparing to ML.

5.1.2. Estimation in presence of model error at the subject level

To mimic misspecification presence, we now generate the observations from the hypoelliptic stochastic model:

$$\begin{cases} dX_{1,i} = \phi_{2,i} X_{2,i} dt - \phi_{1,i} X_{1,i} dt \\ dX_{2,i} = -\phi_{2,i} X_{2,i} dt + \alpha dB_t \\ (X_{1,i}(0), X_{2,i}(0)) = (x_{1,0}, x_{2,0,i}) \end{cases} \tag{17}$$

with B_t a Wiener process and $\alpha = 0.1$ the diffusion coefficient. For the sake of comparison, a solution of (16) and a realization of its perturbed counterpart

given by (17) are plotted in figure 1. This framework where stochasticity only affects the unmeasured compartment is known to be problematic for parameter estimation and inference procedures are yet to be developed for sparse sampling case. From figure 1 it is easy to see the diffusion α will be hard to estimate when we only have observations for $X_{1,i}$. Thus, we still estimate the parameters from the model (16) which is now seen as a deterministic approximation of the true stochastic process. Still, it is expected that our method will mitigate the effect of stochasticity on the estimation accuracy by taking into account model error presence. Results are presented in table 1. The differences between the two methods are similar to the previous well-specified case with an additional loss of accuracy coming from model error for both estimators. However, the misspecification effect for SAEM is more pronounced than for our method which manages to limit the damages done. This confirms the benefits of taking into account model uncertainty for the regularization of the inverse problem, in particular when model error occurs in the unobserved compartment, a situation in which classic statistical criteria for model assessment based on a data fitting criterion are difficult to use.

5.2. Partially observed nonlinear model

We consider a simplified version of the model used in [13] for the analysis of glucose and insulin regulation:

$$\begin{cases} \dot{G}_i = S_G(G_B - G_i) - X_i G_i \\ \dot{I}_i = \gamma t(G_i - h) - n_i(I_i - I_B) \\ \dot{X}_i = -p_2(X_i + S_I(I_i - I_B)). \end{cases} \tag{18}$$

We are in a partially observed case where only the glucose (G_i) and insulin (I_i) concentration are measured. The values of parameters $(p_2, \gamma, h, G_B, I_B)$ are set to $(-4.93, -6.85, 4.14, 100, 100)$ and we aim to estimate $\theta = (\theta_{S_G}, \theta_{S_I}, \theta_n)$, linked to the original model via the parametrization: $\log(S_G) = \theta_{S_G}$, $\log(S_I) = \theta_{S_I}$ and $\log(n_i) = \theta_n + b_i$ where $b_i \sim N(0, \Psi)$. The true population parameter values are $\theta^* = (-3.89, -7.09, -1.81)$ and $\Psi^* = 0.26^2$. The true subject-specific

initial conditions $x^*_{i,0} = \left(G^*_{0,i}, I^*_{0,i}, X^*_{0,i}\right)$ are distributed according $\ln(x^*_{i,0}) \sim N(l_{x^*_0}, \Psi_{l_{x^*_0}})$ with $l_{x^*_0} = (5.52, 4.88, -7)$ and $\Psi_{l_{x^*_0}} = \left(0.17^2, 0.1^2, 10^{-4}\right)$. We generate $n_i = 5$ observations on $[0, T] = [0, 180]$ with Gaussian measurement noise of standard deviation $\sigma^* = 3$. As in the previous example, we investigate the impact of unknown initial conditions on estimators accuracy. We are particularly interested by the joint estimation of θ_{S_I}, which appears only in the equation ruling the unobserved state variable X_i, and $x^*_{0,i}$ required for each subject by the maximum likelihood based method. For this, we distinguish two cases, 1)when θ_{S_I} is known, 2)when θ_{S_I} has to be estimated and we respectively denote $\widehat{\theta_{S_i}}$ and $\widehat{\theta}$ the corresponding estimators. Finally, since the model is non-linear we have to specify a pseudo-linear representation of the vector field as in (7):

$$A_{\theta,b_i}(t, G_i, I_i, X_i) = \begin{pmatrix} -S_G & 0 & -G_i \\ \gamma t & -n_i & 0 \\ 0 & -p_2 S_I & -p_2 \end{pmatrix}, \; r_{\theta,b_i}(t) = \begin{pmatrix} S_G G_B \\ -\gamma t h + n_i I_B \\ -p_2 S_I I_B \end{pmatrix}.$$

5.2.1. Correct model case

We present the estimation results in table 2. Our method obtains smaller MSE than ML and escapes the drop in coverage rate of the confidence interval in the case of $\theta^*_{S_I}$ estimation. The difference between the two estimators behavior is explained by the fact that they are defined through the construction of two different optimization problems. At the population level, our approach leads to minimize a cost function depending on a 4-dimensional parameter whereas ML, due to its need to estimate $x^*_{i,0}$, considers a 10-dimensional one. Thus, the topology of the parameter spaces explored by each method to look for the minimum are very different.

		Well-specified						Misspecified					
		MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i	MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i
θ_{S_G}	$\widehat{\theta}^{ML}_{S_i}$	5e-5	2e-3	4e-5	9e-6	0.95		6e-5	3e-3	6e-5	2e-5	0.85	
	$\widehat{\theta}^{ML}$	2e-3	0.03	1e-3	8e-5	0.85		2e-3	3e-3	1e-3	2e-4	0.54	
	$\widehat{\theta}_{S_i}$	**1e-5**	**4e-4**	**1e-5**	**8e-6**	**0.95**		**2e-5**	**-2e-5**	**2e-5**	**2e-5**	**0.93**	
	$\widehat{\theta}$	**2e-4**	**-6e-4**	**2e-4**	**2e-4**	**0.96**		**3e-4**	**-1e-3**	**3e-4**	**4e-4**	**0.93**	
θ_{S_I}	$\widehat{\theta}^{ML}_{S_i}$	known						known					
	$\widehat{\theta}^{ML}$	2e-3	0.03	1e-3	6e-5	0.90		0.01	0.04	0.01	1e-3	0.55	
	$\widehat{\theta}_{S_i}$	**known**						**known**					
	$\widehat{\theta}$	**1e-4**	**-7e-4**	**1e-4**	**1e-4**	**0.96**		**3e-4**	**-1e-3**	**3e-4**	**3e-4**	**0.92**	
θ_n	$\widehat{\theta}^{ML}_{S_i}$	7e-4	3e-3	6e-4	5e-4	0.94		8e-4	-3e-3	8e-4	5e-4	0.89	
	$\widehat{\theta}^{ML}$	9e-4	8e-3	8e-4	5e-4	0.86		5e-3	-5e-3	5e-3	5e-4	0.88	
	$\widehat{\theta}_{S_i}$	**5e-4**	**6e-3**	**5e-4**	**5e-4**	**0.95**		**4-4**	**7e-4**	**4e-4**	**5e-4**	**0.95**	
	$\widehat{\theta}$	**6e-4**	**6e-3**	**5e-4**	**5e-4**	**0.95**		**4e-4**	**6e-4**	**4e-4**	**5e-4**	**0.96**	
Ψ	$\widehat{\theta}^{ML}_{S_i}$	0.02	7e-4	0.02	0.02	0.95	0.02	0.03	-3e-3	0.03	0.02	0.93	0.03
	$\widehat{\theta}^{ML}$	0.04	-0.09	0.03	0.02	0.88	0.02	0.03	-8e-3	0.02	0.02	0.87	0.03
	$\widehat{\theta}_{S_i}$	**0.01**	**-2e-3**	**0.01**	**0.01**	**0.95**	**0.01**	**0.01**	**-4e-3**	**0.01**	**0.02**	**0.94**	**0.01**
	$\widehat{\theta}$	**0.01**	**3e-3**	**0.01**	**0.01**	**0.94**	**0.01**	**0.02**	**-7e-3**	**0.02**	**0.02**	**0.94**	**0.02**

Table 2: Results of estimation for model (18). The different subscripts stand for the following estimation scenarios: 1)S_i when S_i is set to S_i^*, 2)absence of subscript when S_i is estimated. Results from our method are in bold.

5.2.2. Estimation in presence of model error at the subject level

To mimic misspecification presence, we generate the observations from the stochastic model:

$$\begin{cases} dG_i = (S_G(G_B - G_i) - X_i G_i)\, dt + \alpha_1 dB_{1,t} \\ dI_i = (\gamma t(G_i - h) - n_i(I_i - I_B))dt + \alpha_2 dB_{2,t} \\ dX_i = (-p_2(X_i + S_I(I_i - I_B)))\, dt + \alpha_3 dB_{3,t} \end{cases} \tag{19}$$

where the $B_{i,t}$ are Wiener processes and $(\alpha_1, \alpha_2, \alpha_3) = \left(2, 2, 2 \times 10^{-4}\right)$ their diffusion coefficients. We present the estimation results in table 2. For ML, the drop in coverage rate for $\theta^*_{S_G}$ and $\theta^*_{S_I}$ is even more striking when $\theta^*_{S_I}$ needs to be estimated. This is explained by the effect of model misspecification which increases bias and the fact that ML does not take into account this new source of uncertainty which leads to under-estimation of variance and too narrow confidence intervals.

5.3. Antibody concentration evolution model

We consider the model presented in [31] to analyze the antibody concentration, denoted A_i, generated by two populations of antibody secreting cells: the short lived, denoted S_i, and the long-lived, denoted L_i:

$$\begin{cases} \dot{S}_i = -\delta_S S_i \\ \dot{L}_i = -\delta_L L_i \\ \dot{A}_i = \vartheta_{S,i} S_i + \vartheta_{L,i} L_i - \delta_{Ab,i} A_i \\ (S_i(0), L_i(0), A_i(0)) = (S_{0,i}, L_{0,i}, A_{0,i})\,. \end{cases} \tag{20}$$

This model is used to quantify the humoral response on different populations after an Ebola vaccine injection with a 2 doses regimen seven days after the second injection when the antibody secreting cells enter in a decreasing phase. These cells being unobserved, the preceding equation can be simplified to focus on antibody concentration evolution:

$$\dot{A}_i = \phi_{S,i} e^{-\delta_S t} + \phi_{L,i} e^{-\delta_L t} - \delta_{Ab,i} A_i \tag{21}$$

with $\phi_{S,i} := \vartheta_{S,i} S_{0,i}$ and $\phi_{L,i} := \vartheta_{L,i} L_{0,i}$. This equation has an analytic solution which will be used for maximum likelihood estimation with SAEMIX. We

Parameters		Biological interpretation	Values
δ_L		long-lived B-cells declining rate	$\log(2)/(364 \times 6)$
θ^*	$\theta^*_{\delta_S}$	Mean log-value for δ_S, the short-lived cells declining rate	$\log(\log(2)/1.2) \simeq -0.54$
	$\theta^*_{\phi_S}$	Mean log-value for ϕ_S, the antibodies influx from short-lived cells	$\log(2755) \simeq 7.92$
	$\theta^*_{\phi_L}$	Mean log-value for ϕ_L, the antibodies influx from long-lived cells	$\log(16) \simeq 2.78$
	$\theta^*_{\delta_{Ab}}$	Mean log-value for δ_{Ab}, the antibodies declining rate	$\log(\log(2)/24) \simeq -3.54$
Ψ^*	$\Psi^*_{\phi_S}$	Inter individual variance for $\log(\phi_{S,i})$	0.92^2
	$\Psi^*_{\phi_L}$	Inter individual variance for $\log(\phi_{L,i})$	0.85^2
	$\Psi^*_{\delta_{Ab}}$	Inter individual variance for $\log(\delta_{Ab,i})$	0.3^2

Table 3: Biological interpretation and parameter values

consider the following parametrization: $\log(\delta_S) = \theta_{\delta_S}$, $\log(\phi_{S,i}) = \theta_{\phi_S} + b_{\phi_S,i}$, $\log(\phi_{L,i}) = \theta_{\phi_L} + b_{\phi_L,i}$ and $\log(\delta_{Ab,i}) = \theta_{\delta_{Ab}} + b_{\delta_{Ab},i}$. The true parameter values are presented in table 3. According to [31], the parameter δ_L was non-identifiable and only a lower bound has been derived for it via profiled likelihood. So, to make fair comparisons between our approach and maximum likelihood, we do not estimate it. Regarding population parameters, we are particularly interested in the behavior of estimation methods for θ_{δ_S} and θ_{ϕ_S}. Indeed, a parameter sensitivity analysis shows the symmetric role of θ_{δ_S} and θ_{ϕ_S} on the ODE solution (see [42]). Thus, they are likely to face practical identifiability problems. To investigate this effect, we estimate the parameters when 1) $\theta^*_{\delta_S}$ is known (the corresponding estimators will be denoted with the subscript δ_S), 2) it has to be estimated as well.

5.3.1. Correct model case

We generate $n_i = 11$ observations on the interval $[0, T] = [0, 364]$ with measurement noise of standard deviation $\sigma^* = 100$. For each subject i, the initial condition has been generated according to $A^*_{0,i} \sim N(\overline{A_0}, \sigma^2_{\overline{A_0}})$ with $\overline{A_0} = 500$ and $\sigma_{\overline{A_0}} = 260$ to reflect the dispersion observed in data presented in [31]. We present the estimation results in table 4. Our method gives an improved estimation with reduced variance of $\theta^*_{\delta_S}$ comparing to the ML. Our approach

		Well-specified						Misspecified					
		MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i	MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i
θ_{δ_S}	$\widehat{\theta}^{ML}_{\delta_S}$	known						known					
	$\widehat{\theta}_{ML}$	2.13	0.78	1.51	70.64	0.92		3.88	1.48	1.68	4.10	0.80	
	$\widehat{\theta}_{\delta_S}$	**known**						**known**					
	$\widehat{\theta}$	**0.62**	**-0.34**	**0.50**	**0.66**	**0.92**		**0.93**	**-0.40**	**0.77**	**0.62**	**0.90**	
θ_{ϕ_S}	$\widehat{\theta}^{ML}_{\delta_S}$	4e-4	0.01	3e-4	3e-4	0.94		1e-3	0.02	1e-3	5e-4	0.91	
	$\widehat{\theta}_{ML}$	0.01	-0.05	7e-3	0.40	0.92		0.02	-0.10	0.01	0.02	0.88	
	$\widehat{\theta}_{\delta_S}$	**2e-3**	**-0.05**	**2e-4**	**1e-3**	**0.94**		**7e-4**	**-0.02**	**3e-4**	**1e-3**	**0.92**	
	$\widehat{\theta}$	**2e-3**	**1e-3**	**2e-3**	**2e-3**	**0.93**		**4e-3**	**-6e-3**	**3e-3**	**0.01**	**0.90**	
θ_{ϕ_L}	$\widehat{\theta}^{ML}_{\delta_S}$	3e-3	0.02	3e-3	2e-3	0.95		5e-3	0.03	4e-3	3e-3	0.93	
	$\widehat{\theta}_{ML}$	4e-3	0.03	4e-3	3e-3	0.90		9e-3	0.05	7e-3	4e-3	0.90	
	$\widehat{\theta}_{\delta_S}$	**7e-4**	**-0.01**	**5e-4**	**3e-3**	**0.95**		**2e-3**	**-0.02**	**3e-3**	**2e-3**	**0.97**	
	$\widehat{\theta}$	**3e-3**	**-3e-3**	**3e-3**	**2e-3**	**0.91**		**6e-3**	**-8e-3**	**6e-3**	**7e-3**	**0.90**	
$\theta_{\delta_{Ab}}$	$\widehat{\theta}^{ML}_{\delta_S}$	7e-4	-0.02	5e-4	3e-4	0.93		2e-3	-0.03	1e-3	1e-3	0.92	
	$\widehat{\theta}_{ML}$	2e-3	-0.02	1e-3	4e-4	0.88		4e-3	-0.04	3e-3	7e-4	0.88	
	$\widehat{\theta}_{\delta_S}$	**2e-4**	**0.01**	**1e-4**	**3e-4**	**0.95**		**3e-4**	**2e-3**	**3e-4**	**3e-4**	**0.96**	
	$\widehat{\theta}$	**4e-4**	**0.01**	**3e-4**	**2e-4**	**0.90**		**3e-4**	**8e-3**	**3e-4**	**2e-3**	**0.89**	
Ψ_{ϕ_S}	$\widehat{\theta}^{ML}_{\delta_S}$	0.04	-1e-3	0.04	0.07	1	0.15	0.05	0.03	0.05	0.08	1	0.17
	$\widehat{\theta}_{ML}$	0.11	0.01	0.11	0.05	1	0.17	0.13	0.01	0.13	0.25	1	0.21
	$\widehat{\theta}_{\delta_S}$	**0.02**	**8e-3**	**0.02**	**0.01**	**0.94**	**0.06**	**0.02**	**2e-3**	**0.02**	**0.02**	**0.94**	**0.11**
	$\widehat{\theta}$	**0.02**	**-0.03**	**0.02**	**0.02**	**0.94**	**0.07**	**0.02**	**-0.05**	**0.02**	**0.03**	**0.92**	**0.08**
Ψ_{ϕ_L}	$\widehat{\theta}^{ML}_{\delta_S}$	0.03	0.04	0.02	0.04	1	0.30	0.05	0.03	0.05	0.06	1	0.73
	$\widehat{\theta}_{ML}$	0.03	0.05	0.02	0.04	1	0.60	0.03	0.05	0.02	0.07	1	0.74
	$\widehat{\theta}_{\delta_S}$	**0.02**	**-0.1**	**5e-3**	**8e-3**	**0.93**	**0.07**	**0.02**	**-0.10**	**0.01**	**0.02**	**0.91**	**0.10**
	$\widehat{\theta}$	**0.03**	**-0.06**	**0.02**	**0.01**	**0.92**	**0.08**	**0.03**	**-0.06**	**0.02**	**0.03**	**0.87**	**0.12**
$\Psi_{\delta_{Ab}}$	$\widehat{\theta}^{ML}_{\delta_S}$	0.11	0.18	0.08	0.02	1	0.10	0.33	0.41	0.17	0.05	1	0.56
	$\widehat{\theta}_{ML}$	0.20	0.29	0.11	0.02	1	0.50	0.30	0.34	0.19	0.05	1	0.69
	$\widehat{\theta}_{\delta_S}$	**0.10**	**-0.30**	**0.01**	**0.01**	**0.95**	**0.03**	**0.10**	**-0.16**	**0.08**	**0.06**	**0.91**	**0.04**
	$\widehat{\theta}$	**0.11**	**-0.27**	**0.04**	**0.04**	**0.95**	**0.04**	**0.15**	**-0.29**	**0.06**	**0.10**	**0.88**	**0.06**

Table 4: Results of estimation for model (21). The different subscripts stand for the following estimation scenarios: 1)δ_S when θ_{δ_S} is set to $\theta^*_{\delta_S}$, 2)absence of subscript when θ_{δ_S} is estimated. Results from our method are in bold.

provides an improved estimate for the $\{b_i^*\}_{i\in[\![1,\,n]\!]}$. We assume that is due to the committed estimation error for θ^*, as it causes model error for $\{b_i^*\}_{i\in[\![1,\,n]\!]}$ estimation, which is not taken into account by exact methods. This in turn explains why their variance Ψ^* is better estimated with our approach. In this mixed-effect context, this cause of model error is systematically present and claims for the use of estimation methods taking into account modeling uncertainties when subject specific parameters are critical for the practitioner.

5.3.2. Estimation in presence of model error at the subject level

The data are now generated with a stochastic perturbed version of the original model:

$$dA_i = \left(\phi_{S,i}e^{-\delta_S t} + \phi_{L,i}e^{-\delta_L t} - \delta_{Ab,i}A_i\right)dt + \alpha dB_t \tag{22}$$

where B_t is a Wiener process and $\alpha = 10$ its diffusion coefficient. The value for α has been chosen big enough to produce significantly perturbed trajectories but small enough to ensure that ODE (21) is still a relevant approximation for estimation purpose. The results are presented in table 4. Our method still outperforms the maximum likelihood for $\theta^*_{\delta_S}$ as well as the $\{b_i^*\}_{i\in[1,\,n]}$ estimation and their variances. In addition, we mitigate the effect of model error on estimation accuracy.

6. Real data analysis

We now proceed to the estimation starting from real data presented in [31] from which the parameter values given in table 3 come from. In [31], the estimation is made from cohorts coming from three phase I trials performed in African and European countries. Each subject was vaccinated with two doses, Ad26.ZEBOV (Janssen Vaccines and Prevention) and MVA-BN-Filo (Bavarian Nordic). In these cohorts, both the effect of injection order, either Ad26.ZEBOV first and MVA-BN-Filo second, or MVA-BN-Filo first and Ad26.ZEBOV second, and the delay between, 28 or 56 days, were evaluated. In this study, we focus on an east African subpopulation where Ad26.ZEBOV

was injected first and then MVA-BN-Filo with a delay of 28 days between the two doses. As in [31] to stay in the temporal domain of validity of the model, we use the 5 measurements per subject made seven days after the second dose injection. The estimation in the original work has been done using the NIMROD software [12] and log-transformed antibody concentration measurement. We now estimate the parameters with our method with the aim to compare our results with the existing one. We used the same prior distribution

$$\pi(\theta) \sim N\left(\begin{pmatrix} -1 \\ 0 \\ 0 \\ -4.1 \end{pmatrix}, \begin{pmatrix} 25 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right) \text{ for } \theta = (\theta_{\delta_S}, \theta_{\phi_S}, \theta_{\phi_L}, \theta_{\delta_{Ab}})$$

as them. We set our mesh-size to get 200 discretization points for each subject on the observation interval and we use $U = 10$ i.e. a value lower than in the simulated data case because of the model error presence. We also proceed to the log-transformation of the data to stabilize the measurement noise variance. This drives us to use the nonlinear model:

$$\dot{\widetilde{A}}_i(t) = \frac{1}{\ln(10)}\left(\phi_{S,i}e^{-\delta_S t} + \phi_{L,i}e^{-\delta_L t}\right) 10^{-\widetilde{A}_i(t)} - \frac{\delta_{Ab,i}}{\ln(10)} \tag{23}$$

describing the dynamic of $\widetilde{A_i}(t) := \log_{10} A_i(t)$ for parameter estimation purpose. We use $A_{\theta,b_i}(t, x, z_i(t)) = \frac{1}{\ln(10)}\left(\phi_{S,i}e^{-\delta_S t} + \phi_{L,i}e^{-\delta_L t}\right)\frac{10^{-x}}{x}$ and $r_{\theta,b_i}(t, z_i(t)) = -\frac{\delta_{Ab,i}}{\ln(10)}$ for the pseudo-linear formulation of the model. Our estimation and the one from the original paper [31] are presented in Table 5. In the following, we denote $\left(\widehat{\theta}^P, \widehat{b_i}^P\right)$ (respectively $\left(\widehat{\theta}, \widehat{b_i}\right)$) the estimation obtained by [31] (respectively our approach). Both methods produce estimations with overlapping confidence intervals for θ. Still, significant differences appear for $(\Psi_{\phi_S}, \Psi_{\phi_L}, \Psi_{\delta_{Ab}})$ estimation which quantifies the dispersion of random effects. We only consider a subset of the subjects used in [31] for estimation. This has an effect on the observed diversity within the cohort of patients and thus on $(\Psi_{\phi_S}, \Psi_{\phi_L}, \Psi_{\delta_{Ab}})$ estimation. Regarding the predictions, we present in figure 2 examples of estimated trajectories.

The confidence intervals are computed via Monte-Carlo sampling from the

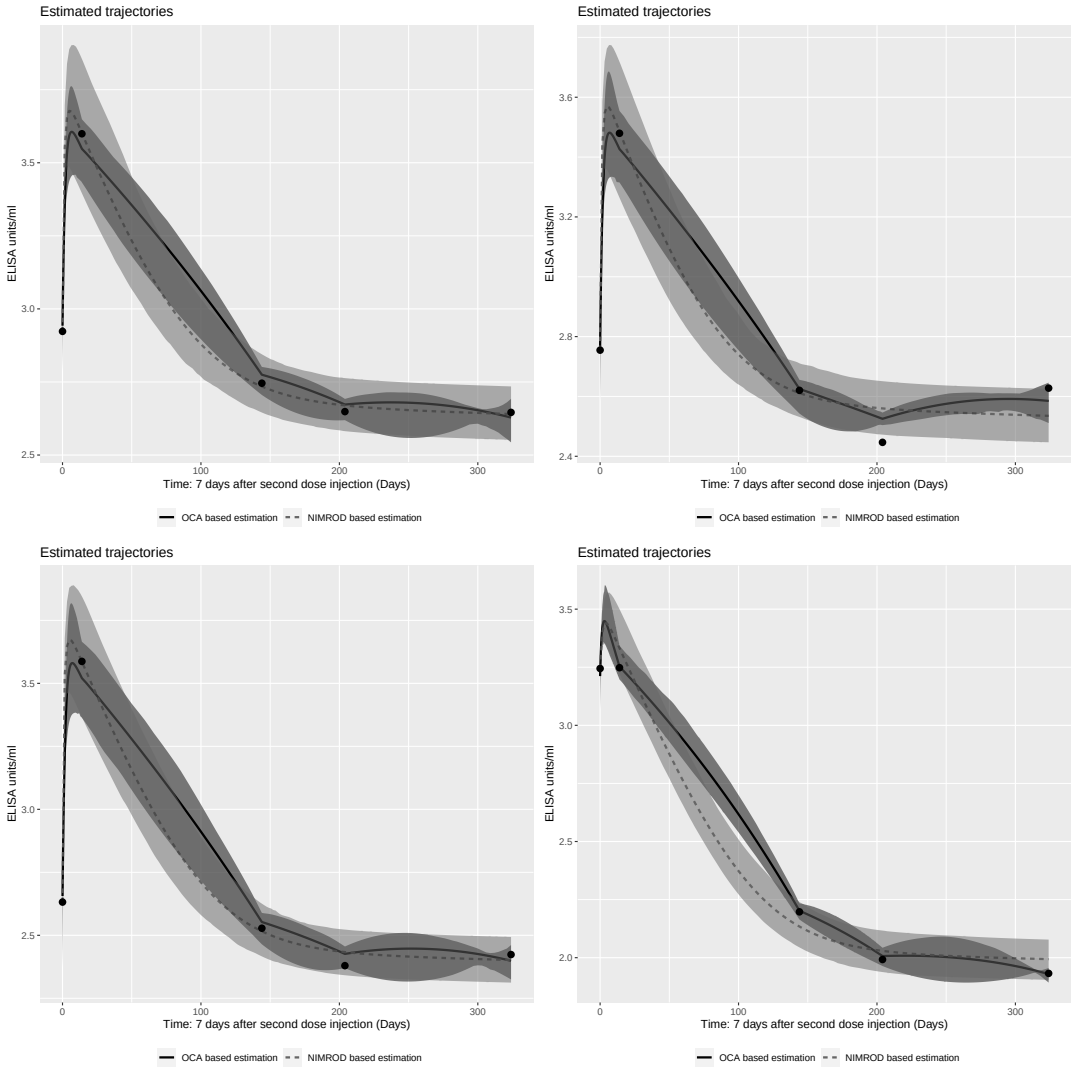

Figure 2: Examples of fitted trajectories for both methods for different subjects. Here Time=0 is the 7th day post-second dose. Dashed lines: fitted ODE solutions (23) with $\left(\widehat{\theta}^P, \widehat{b_i}^P\right)$. Solid line: optimal trajectories $\overline{X}_{\widehat{\theta},\widehat{b_i}}$. Shaded area are the 95% confidence intervals.

	θ_{δ_S}	θ_{ϕ_S}	θ_{ϕ_L}	$\theta_{\delta_{Ab}}$
Pasin et al.	-0.57 [-1.02, -0.02]	7.92 [7.52, 8.30]	2.78 [2.62, 3.01]	-3.54 [-3.62, -3.45]
OCA	-0.18 [-0.58, 0.22]	7.45 [6.85, 7.96]	2.58 [2.15, 3.01]	-3.48 [-3.95, -3.01]
		Ψ_{ϕ_S}	Ψ_{ϕ_L}	$\Psi_{\delta_{Ab}}$
Pasin et al.		0.92 [0.83, 1.01]	0.85 [0.78, 0.92]	0.3 [0.24, 0.36]
OCA		0.64 [0.60, 0.70]	0.70 [0.55, 0.90]	0.25 [0.19, 0.31]

Table 5: Estimation presented in [31]and via our approach.

approximated normal laws $\mathcal{N}(\widehat{\theta}, V(\widehat{\theta}))$ and $\mathcal{N}(\widehat{\theta}^P, V(\widehat{\theta}^P))$ to quantify the effect of estimation uncertainy on θ on the prediction. For NIMROD estimation, for a given sampled value $\widetilde{\theta}^P \sim \mathcal{N}(\widehat{\theta}^P, V(\widehat{\theta}^P))$ and subject i, the sampled regression function $X_{\widetilde{\theta}^P, \widehat{b_i}^P, y_{0,i}}$ is obtained by solving ODE (23) for parameter values $(\theta, b_i, x_{0,i}) = \left(\widetilde{\theta}^P, \widehat{b_i}^P, y_{0,i}\right)$. Regarding our approach, for $\widetilde{\theta} \sim \mathcal{N}(\widehat{\theta}, V(\widehat{\theta}))$ the sampled regression function for subject i is the optimal trajectory $\overline{X}_{\widetilde{\theta}, \widehat{b_i}}$ obtained via the minimization of the cost function $\mathcal{C}_i(\widehat{b_i}, x_{i,0}, u_i \mid \widetilde{\theta}, U)$. This imposes a common goal of data fidelity to each sampled $\overline{X}_{\widetilde{\theta}, \widehat{b_i}}$ which limits their inter-variability. This explain the differences between the two confidence intervals in terms of shape and width and why our method gives narrower intervals. Still, despite these differences in shapes, both prediction intervals cover the same points. Morever, on the long-term our intervals are nearly always contained in the ones given by NIMROD.

Our estimation of θ supports the parameter inference obtained in [31] via another method and the subsequent analysis made on the antibody concentration dynamics. In addition to this parametric comparison, we want to assess the model adequacy via the temporal evolution analysis of the optimal controls $\overline{u}_{i, \widehat{\theta}, b_i(\widehat{\theta})}$ estimated as byproducts of our method. Indeed, they quantify the exogenous perturbations u_i we need to add to model (23) so that the solution of its perturbed counterpart,

$$\dot{\widetilde{A_{i,u}}}(t) = \frac{1}{\ln(10)} \left(\phi_{S,i} e^{-\delta_S t} + \phi_{L,i} e^{-\delta_L t}\right) 10^{-\widetilde{A_{i,u}}(t)} - \frac{\delta_{Ab,i}}{\ln(10)} + u_i \qquad (24)$$

reproduce the observations. This approach is similar to the one developed in [43] where control theory replaces non-parametric procedures to estimate u_i. For comparison, we also quantify the committed model error for $\left(\widehat{\theta}^P, \widehat{b_i}^P\right)$. To do so, we compute $\overline{u}_i^P$, the solution of the optimal control problem: $\overline{u}_i^P = \arg\min_{u_i} \left\{ \sum_j \left\| \widetilde{A}_{i,\widehat{\theta}^P,\widehat{b_i}^P,y_{i0},u_i}(t_{ij}) - y_{ij} \right\|_2^2 + \|u_i\|_{U,L^2}^2 \right\}$. In the last expression $\widetilde{A}_{i,\widehat{\theta}^P,\widehat{b_i}^P,y_{i0},u_i}$ is the solution of the perturbed ODE (24) for $(\theta, b_i) = \left(\widehat{\theta}^P, \widehat{b_i}^P\right)$ and y_{i0} is the measured concentration at $t = 0$ used a surrogate value for the initial condition (as they did in [31]). We still use $U = 10$ for this optimal control problem to allow for the same level of perturbation magnitude for both methods. In figure (3), we plot $\overline{u}_{i,\widehat{\theta},b_i(\widehat{\theta})}$ and $\overline{u}_i^P$ as well as their mean values and confidence intervals.

Our method leads to residual perturbations of smaller magnitudes and narrower confidence intervals. This means our approach produces an estimation which minimizes the committed model error for each subject comparing to a method based only on a data fitting criteria. This is particularly clear at the beginning of the observation interval. In this case, our narrower confidence interval clearly excludes a null perturbation and advocates for an over-estimation of the predicted antibody concentration by the model. This makes sense because model (20) assumes that both populations of antibody secreting cells decrease with time, and that is probably not true at the beginning of the dynamic. Thus, despite similar results regarding parameter values between our estimation and [31], the insight given by our method at the dynamic scale leads us to the additional conclusion of model misspecification presence at the beginning of the observation interval.

7. Conclusion

In this work, we propose an estimation method addressing issues encountered by classic approaches for the problem of parameter estimation in NLME-ODEs. We identify three potential sources of problems for exact methods such as likelihood based inference: their difficulties in presence of model error, their need

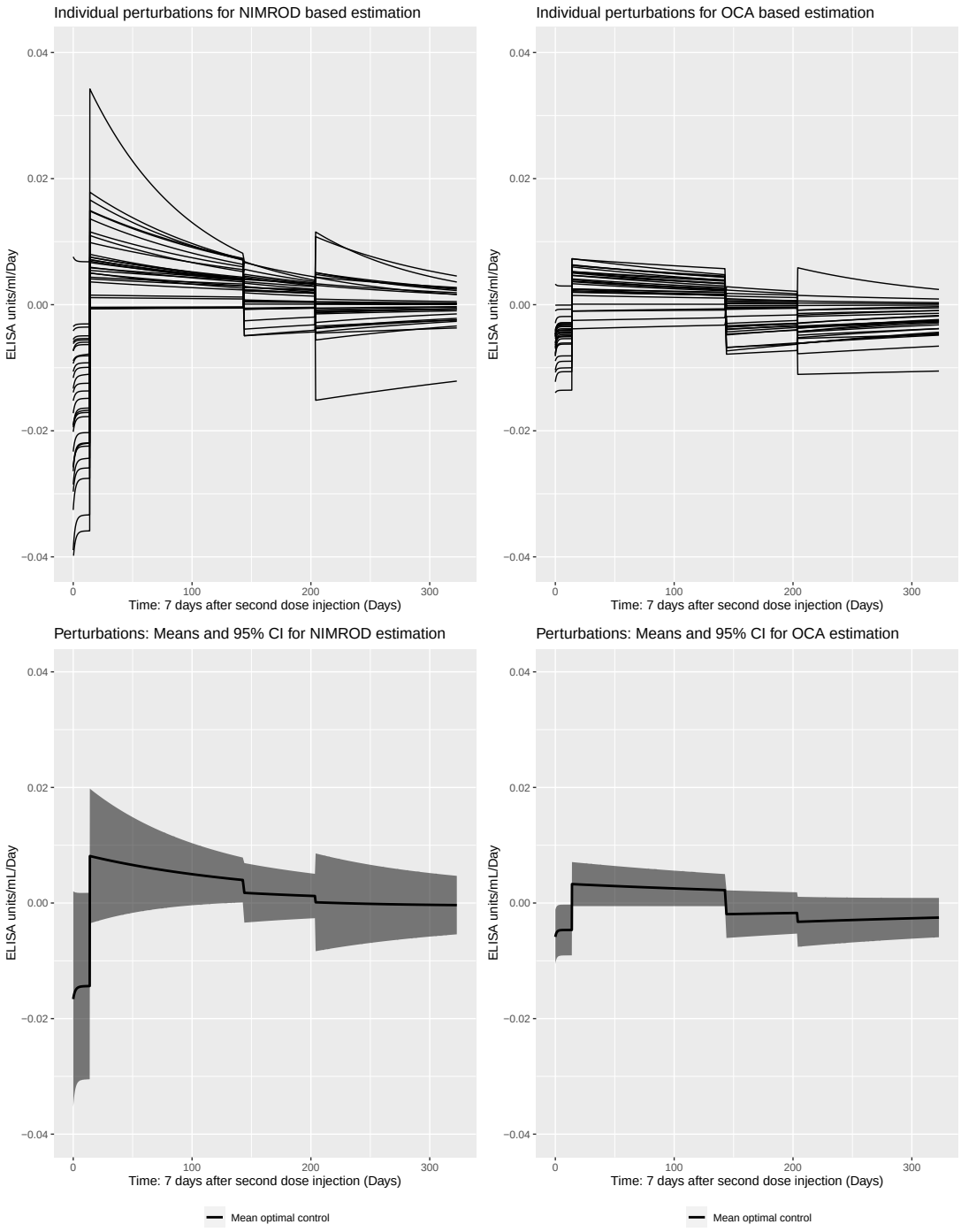

Figure 3: 1) Up: Estimated residual controls for each subject, 2) bottom: mean optimal control and 95% confidence interval for the optimal controls a) left: $\overline{u}_i^P$ obtained from parameter estimation in [31], b) right: $\overline{u}_{i,\widehat{\theta},b_i(\widehat{\theta})}$ obtained from our estimation.

to estimate initial conditions and their dramatic performance degradation when facing poorly identifiable parameters. We propose here a method based on control theory accounting for the presence of potential model uncertainty at the subject level and which can be easily profiled on the initial conditions. Simulations with both presence and absence of model errors illustrate the benefits of regularization techniques for estimating poorly identifiable parameters, subject specific parameters as well as their variances in NLME-ODEs. In addition, bypassing estimation of initial conditions represents a clear advantage for partially observed systems comparing to likelihood based approaches, as emphasized in simulations.

Still, this benefit in term of estimation accuracy comes with a computational price. On a server with the parallelization package Snow in R language, it takes approximately 10-15 minutes to obtain an estimation for the two-dimensional linear model, 30 minutes for the insulin model and 3-4 hour for the antibody concentration evolution one, whereas it was a matter of minutes for the other approaches. Nevertheless, the use of compiled languages and proper parallelization could reduce the computation time. Moreover, we have willingly separated the formal definition of the optimal control problem required by our method and the numerical procedure used to solve it, in case it may exists better suited approaches for this specific control problem. Right now, our current strategy allows us to profile on initial conditions, so looking for another numerical procedure is beyond the scope of this paper.

An under-exploited feature of the method so far is the obtained optimal controls. The qualitative based analysis exposed in section 6 can be made more rigorous. For example, to stay in a Bayesian setting, we can specify a prior distribution for the controls and then compare it with the obtained posterior once the inference is made. This would lead to a semi-parametric inference problem for which an optimal control based approach has already been proven useful (see [27]). This is a subject for further work.

Software

Our estimation method is implemented in R and a code reproducing the examples of Section 5 is available on a GitHub repository located here.

Acknowledgement

Experiments presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr/). This manuscript was developed under WP4 of EBOVAC3.

Funding: This work has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under projects EBOVAC1 and EBOVAC3 (respectively grant agreement No 115854 and No 800176). The IMI2 Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Association.

References

[1] A. Perelson, A. Neumann, M. Markowitz, J. Leonard, D. Ho, Hiv-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time, Science 271 (1996) 1582–1586.

[2] H. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller, P. Schuster, Inverse problems in systems biology, Inverse Problems 25 (12).

[3] L. Villain, D. Commenges, C. Pasin, M. Prague, R. Thiébaut, Adaptive protocols based on predictions from a mechanistic model of the effect of il7 on cd4 counts, Statistics in medicine 38 (2) (2019) 221–235.

[4] A. F. M. Lavielle, A. Samson, F. Mentre, Maximum likelihood estimation of long terms hiv dynamic models and antiviral response., Biometrics 67 (2011) 250–259.

[5] J. Guedj, R. Thiebaut, D. Commenges, Maximum likelihood estimation in dynamical models of hiv, Biometrics 63 (2007) 1198–206.

[6] J. Pinheiro, D. M. Bates, Approximations to the loglikelihood function in the nonlinear mixed effects model., Journal of the Computational and Graphical Statistics 4 (1994) 12–35.

[7] E. Comets, A. Lavenu, M. Lavielle, Parameter estimation in nonlinear mixed effect models using saemix, an r implementation of the saem algorithm, Journal of Statistical Software 80 (2017) 1–42.

[8] M. Lavielle, F. Mentré, Estimation of population pharmacokinetic parameters of saquinavir in hiv patients with the monolix software, Journal of Pharmacokinetics and Pharmacodynamics 34.

[9] D. Lunn, A.Thomas, N. Best, D. Spiegelhalter, Winbugs - a bayesian modelling framework: Concepts, structure and extensibility., Statistics and Computing 10 (2000) 325–337.

[10] Y. Huang, G. Dagne, A bayesian approach to joint mixed-effects models with a skew normal distribution and measurement errors in covariates, Biometrics 67 (2011) 260–269.

[11] A. Raftery, L. Bao, Estimating and projecting trends in hiv/aids generalized epidemics using incremental mixture importance sampling, Biometrics 66 (2010) 1162–1173.

[12] M. Prague, D. Commengues, J. Guedj, J. Drylewicz, R. Thiébaut, Nimrod: A program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations, Computer Methods and Programs in Biomedicine 111 (2013) 447–458.

[13] C. Tornoe, H. Agerso, E. N. Jonsson, H. Madsen, H. A. Nielsen, Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in nlme using differential equations., Computer Methods and Programs in Biomedicine 76 (2004) 31–41.

[14] S. Donnet, A. Samson, Estimation of parameters in incomplete data models defined by dynamical systems, Journal of Statistical Planning and Inference 137 (9) (2006) 2815–2831.

[15] L. Wang, J. Cao, J. Ramsay, D. Burger, C. Laporte, J. Rockstroh, Estimating mixed-effects differential equation models, Statistics and Computing 24 (2014) 111–121.

[16] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464.

[17] T. Kurtz, Strong approximation theorems for density dependent markov chains., Stochastic Processes and their Applications 6 (1978) 223–240.

[18] N. V. Kampen, Stochastic Process in Physics and Chemistry, Elsevier, 1992.

[19] J. Brynjarsdottir, A. O'Hagan, Learning about physical parameters: The importance of model discrepancy, Inverse Problems 30 (2014) 24.

[20] Y. Huang, T. Lu, Modeling long-term longitudinal hiv dynamics with application to an aids clinical study., Annal of Applied Statistics 2 (2008) 1348–1408.

[21] R. N. Gutenkunst, J. Waterfall, F. Casey, K. Brown, C. Myers, J. Sethna, Universally sloppy parameter sensitivities in systems biology models, Public Library of Science Computational Biology 3 (2007) e189.

[22] T. O. Leary, A. Sutton, E. Marder, Computational models in the age of large datasets, Current Opinion in Neurobiology 32 (2015) 87–94.

[23] D. Campbell, Bayesian collocation tempering and generalized profiling for estimation of parameters from differential equation models, Ph.D. thesis, McGill University Montreal,Quebec (2007).

[24] M. Lavielle, L. Aarons, What do we mean by identifiability in mixed effects models?, Journal of pharmacokinetics and pharmacodynamics.

[25] J. M. Varah, A spline least squares method for numerical parameter estimation in differential equations, SIAM J.sci. Stat. Comput. 3 (1) (1982) 28–46.

[26] J. Ramsay, G. Hooker, J. Cao, D. Campbell, Parameter estimation for differential equations: A generalized smoothing approach, Journal of the Royal Statistical Society (B) 69 (2007) 741–796.

[27] Q. Clairon, A regularization method for the parameter estimation problem in ordinary differential equations via discrete optimal control theory, Journal of Statistical Planning and Inference.

[28] R. Tuo, C. Wu, Efficient calibration for imperfect computer models, Annals of Statistics.

[29] F. Clarke, Functional Analysis, Calculus of Variations and Optimal Control, Graduate Texts in Mathematics, Springer-Verlag London, 2013.

[30] T. Cimen, S. Banks, Global optimal feedback control for general nonlinear systems with nonquadratic performance criteria, Systems and Control Letters 53 (2004) 327–346.

[31] C. Pasin, I. Balelli, T. Van Effelterre, V. Bockstal, L. Solforosi, M. Prague, M. Douoguih, R. Thiébaut, Dynamics of the humoral immune response to a prime-boost ebola vaccine: quantification and sources of variation, Journal of virology 93 (18) (2019) e00579–19.

[32] E. Sontag, Mathematical Control Theory: Deterministic finite-dimensional systems, Springer-Verlag (New-York), 1998.

[33] M. Aliyu, Nonlinear H-Infinity Control, Hamiltonian Systems and Hamilton-Jacobi Equations, CRC Press, 2011.

[34] M. Dashti, K. J. H. Law, A. Stuart, J. Voss, Map estimators and their consistency in bayesian nonparametric inverse problems, Inverse Problems 29.

[35] S. Murphy, A. V. der Vaart, On profile likelihood, Journal of American Statistical Association 95 (2000) 449–465.

[36] T. Cimen, S. Banks, Nonlinear optimal tracking control with application to super-tankers for autopilot design, Automatica 40 (2004) 1845–1863.

[37] T. Cimen, State-dependent riccati equation (sdre) control: A survey, IFAC Proceedings 41 (2008) 3761–3775.

[38] A. van der Vaart, Asymptotic Statistics, Cambridge Series in Statistical and Probabilities Mathematics, Cambridge University Press, 1998.

[39] N. Sartori, Modified profile likelihood in models with stratum nuisance parameters, Biometrika 90 (2003) 553–549.

[40] L. D. V. R. G. Hooker, S. P. Ellner, D. J. D. Earn, Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in ontario, Journal of the Royal Society 8 (2011) 961–974.

[41] J. C. Nash, Using and extending the optimr package.

[42] I. Balelli, C. Pasin, M. Prague, F. Crauste, T. Van Effelterre, V. Bockstal, L. Solforosi, R. Thiébaut, A model for establishment, maintenance and reactivation of the immune response after vaccination against ebola virus, Journal of Theoretical Biology (2020) 110254.

[43] G. Hooker, S. P. Ellner, et al., Goodness of fit in nonlinear dynamics: misspecified rates or misspecified states?, The Annals of Applied Statistics 9 (2) (2015) 754–776.