# Estimation for dynamical systems using a population-based Kalman filter – Applications in computational biology

Annabelle Collin, Mélanie Prague, Philippe Moireau

# Estimation for dynamical systems using a population-based Kalman filter – Applications to pharmacokinetics models

Annabelle Collin[1,2,3], Mélanie Prague[4,5], and Philippe Moireau[6,7]

[1]Inria, Inria Bordeaux - Sud-Ouest, Talence, France
[2]Bordeaux INP, IMB UMR 5251, Talence, France
[3]IMB UMR 5251, Université Bordeaux, Talence, France
[4]INSERM, U1219 Bordeaux Public Health, Talence, France
[5]Vaccine Research institute, Créteil, France
[6]Inria, Inria Saclay-Ile-de-France, France
[7]LMS, CNRS UMR 7649, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

**Abstract**

Many methods exist to identify parameters of dynamical systems. Unfortunately, in addition to the classical measurement noise and under-sampling drawbacks, mean and variance priors of the estimated parameters can be very vague. These difficulties can lead the estimation procedure to underfitting. In clinical studies, a circumvention consists in using the fact that multiple independent patients are observed as proposed by nonlinear mixed-effect models. However, these very effective approaches can turn to be time-consuming or even intractable when the model complexity increases. Here, we propose an alternative strategy of controlled complexity. We first formulate a population least square estimator and its associated a Kalman based filter, hence defining a robust large population sequential estimator. Then, to reduce and control the computational complexity, we propose a reduced-order version of this population Kalman filter based on a clustering technique applied to the observations. Using simulated pharmacokinetics data and the theophylline pharmacokinetics data, we compare the proposed approach with literature methods. We show that using the population filter improves the estimation performance compared to the classical and fast patient-by-patient Kalman filter and leads to estimation results comparable to state-of-the-art population-based approaches. Then, the reduced-order version allows to drastically reduce the computational time for equivalent estimation and prediction.

## 1 Introduction

Mathematical models – as for example systems of ordinary differential equations (ODE) – can be used to understand the evolution of biological processes. They allow to represent dynamic interactions in time between various biomarkers of interest. For instance, in pharmacology, they can describe the drug pharmacokinetics (PK) in plasma [29] or, in virology, the viral kinetics during HIV treatment [18].

When employing these models in practical applications, a great difficulty consists in dealing with the many uncertain quantities that must be prescribed for running model simulations. These quantities include initial conditions and physical parameters of the model, which can be difficult to measure. Fortunately, additional information is provided by available measurements which can be related to the unknowns by linear or nonlinear operators and can be used to circumvent the uncertainties associated with the dynamical model definition.

Estimation methods for these models can be seen as inverse methods, in which we want to recover the individual parameters that have produced some observations. The resolution of the inverse problem leads to solve an ODE constrained optimization problem. It exists many methods to solve this kind of

problem as for example as a non-exhaustive list: gradient descent based methods ; stochastic strategies as for example Monte-Carlo based strategies ; expectation-maximization algorithm or Kalman based filters (also known as linear-quadratic estimation (LQE)) that consist in correcting the original dynamic at each time by a feedback control [2, 1].

Unfortunately, most of the time, data are measured with errors and observations are sampled so that data can be very sparse in time. Even if some of these methods it possible to incorporate prior information, it is sometimes difficult to specify and results in very vague priors. These difficulties can lead the estimation procedure to failure. A way to compensate the weakness of the observations is to use the fact that multiple independent patients are observed and it is possible to build a hierarchical model which assumes that every individual patient has a different value of parameters but the variability in the population is constrained. One example of this modeling strategy is called nonlinear mixed-effect model [28] – and it proposes to use the availability of population data to improve identifiability of individual parameters by maximizing a coupled likelihood function. Many algorithms and software have been developed to perform the optimization. Among others we can cite NONMEM [21] implementing a First Order Conditional Estimation (FOCE) algorithm, MONOLIX [13] implementing a Stochastic Approximation of Expectation-Maximization algorithm, NIMROD [24] implementing a penalized maximum likelihood approach or full-MCMC techniques implemented in Winbugs [30] or jags [7]. A specific combination of the FOCE method with the extended Kalman filter has also been suggested [17, 26, 12]. Also, a coupling of the SAEM algorithm with the extended Kalman filter has been proposed [6]. In these articles, Kalman-based filters are used to approximate the individual probability distribution function. Most of these estimation algorithms and software have been compared in former works and proved to have similar performance results in identifiable medium-size models, such as in models in pharmacokinetics [8, 22, 24, 14].

Nevertheless, when the ODE system gets bigger in terms of the number of unknowns and parameters, these strategies can turn inaccurate, time-consuming or intractable. Here we propose an alternative method in this article to circumvent the complexity increase associated with large systems. We emphasize that our strategy could even be applied to partial differential (PDE) systems. In this respect, we define a population likelihood functional where all disturbances are Gaussian and we optimize it using a Kalman-based filter. More precisely we rely on the Unscented Kalman Filter [10], but other Gaussian filters could lead to similar results [1]. Then, to overcome the potential curse of dimensionality associated with such approaches, we also propose a reduced-order version of this population-based Kalman Filter based on a clustering of the observations. This reduced-order version allows to drastically reduce the computational complexity.

The remainder of this paper is organized as follows. Section 2 formalizes the estimation problem and introduces notation. Section 3 presents the population-based Kalman filter and its reduced-order version. In Section 4, a numerical section provides assessment of the performances of both population-based filters in simulations using a one-compartment pharmacokinetics model. An illustration on the benchmark dataset of Theophilline is proposed and the results are compared to a classical nonlinear mixed effect approach.

# 2 Problem setting

## 2.1 Model formulation

Let us consider a population of $N_P$ subjects. For each subject, we denote by $x^i$, $1 \leq i \leq N_P$, the system state – *i.e.* a vector which contains all the time-varying variables of the system – and by $\theta^i$, $1 \leq i \leq N_P$ the vector containing all the parameters of the system. The state dynamics over the time window $[0, T]$ is modeled by an operator $F$ – eventually nonlinear – such that for all patients $1 \leq i \leq N_P$,

$$\dot{x}^i(t) = F(x^i(t), \theta^i, t) + G(t)\nu^i(t), \quad t \in [0, T],$$

where $\dot{x}^i$ is the time derivative of the state $x^i \in \mathcal{X} \simeq \mathbb{R}^{N_x}$, $\nu^i \in \mathcal{Q} \simeq \mathbb{R}^{N_\nu}$ is a possible unknown error and $G$ the model error operator. The parameters $\theta^i \in \mathcal{P} \simeq \mathbb{R}^{N_\theta}$ can be treated as the state variable by augmenting the state dimension and recalling that $\dot{\theta}^i = 0$. We hence denote by $z^i$ the augmented state, such that

$$z^i = \begin{pmatrix} x^i \\ \theta^i \end{pmatrix} \in \mathcal{X} \times \mathcal{P} \simeq \mathbb{R}^{N_x + N_\theta},$$

and for $1 \leq i \leq N_P$, the dynamics reads

$$\begin{cases} \dot{z}^i(t) = A(z^i(t), t) + B(t)\nu^i(t), \quad \forall t \in [0, T] \\ z^i(0) = z_0^i, \end{cases} \tag{1}$$

with for all $t \in [0, T]$

$$A(z^i(t), t) = \begin{pmatrix} F(x^i(t), \theta^i, t) \\ 0 \end{pmatrix}, \quad B(t) = \begin{pmatrix} G(t) \\ 0 \end{pmatrix}.$$

## 2.2 Uncertainties modeling

The initial state, the values of the parameters and the model errors are not perfectly known for each patient. We assume that we can decompose the initial augmented state as follows

$$z_0^i = z_\diamond + \xi^i, \quad 1 \leq i \leq N_P,$$

where $z_\diamond$ corresponds to the *a-priori* part and $\xi^i$ to the unknown part. Note that this decomposition also models the parameter uncertainties as it is written on the augmented state. Concerning the state part $x_\diamond$ of the *a-priori*, we either assume that $x_\diamond$ is built from the experiment initial conditions or is modeled as an equilibrium state reached before the experiment beginning. In this case, $x_\diamond$ can be analytically or numerically computed by solving the equation $\dot{x}^i(t) = 0$. Additionally, the initial value of the parameters $\theta_\diamond$ is deduced from the experiment conditions and previous biological measurements from past studies. We quantify the level of uncertainty on each $\xi^i$ by considering that $\xi^i$ can be modeled as a Gaussian disturbance of zero mean and $P_\diamond^i$ covariance. Then, in the spirit of mixed-effect methods, we assume that $\xi^i$ is decomposed into two parts, namely

$$\xi^i(t) = \xi^0 + \tilde{\xi}^i,$$

where $\xi^0$ corresponds to the population intercept and $\tilde{\xi}^i$ to the individual effect. We then assume that $\xi^0$ and $\tilde{\xi}^i$, $1 \leq i \leq N_P$ are independent Gaussian disturbances, hence the covariances $(P_\diamond^i - \tilde{P}_\diamond^i)$, $1 \leq i \leq N_P$ are equal to the same matrix denoted by $P_\diamond^0$. Concerning the model error, we also introduce covariances $Q^i$, $1 \leq i \leq N_P$ and assume $\nu^i$ to be a continuous-time white noise (*i.e.* $\mathbb{E}\big(\nu^i(t)\nu^{i\intercal}(s)\big) = Q^i\delta(t-s), \quad 0 \leq t, s \leq T$, with $\delta$ the 1D Dirac delta function [25]).

**Remark 1.** Covariates effects – Our formalism is compatible with models in which parameters are also functions of covariates $w^i$ possibly depending on time such as

$$\theta^i(t) = \theta^0 + \beta w^i(t) + \tilde{\theta}^i,$$

as described in the literature[24]. In this configuration, the parameters are now time-dependent and $\beta$ is a parameter shared by the population. Then, the proposed formalism is compatible with two situations:

- If $w^i(t) = \mathbb{1}_{t>t_s}$, by considering two different periods of estimation $t < t_s$ and $t > t_s$ with $\beta$ incorporated directly into $\theta$.
- If $w^i(t)$ regular enough, by using the following joint dynamics

$$\begin{cases} \dot{x}^i &= F(x^i, \theta^i, \beta^i, t) \\ \dot{\theta}^i &= \beta^i \dot{w}^i(t) \\ \dot{\beta}^i &= 0 \\ x^i(0) &= x_\diamond + \xi_x^i \\ \theta^i(0) &= \theta_\diamond + \xi_\theta^i \\ \beta^i(0) &= \beta_\diamond + \xi_\beta^i \end{cases} \tag{2}$$

  which can be rewritten in the formalism of System (1) with

$$z^i = \begin{pmatrix} x \\ \theta \\ \beta \end{pmatrix} \text{ and } A(z^i(t), t) = \begin{pmatrix} F(x^i, \theta^i, \beta^i, t) \\ \beta^i \dot{w}^i(t) \\ 0 \end{pmatrix}.$$

We define the vector of uncertainties as

$$\xi^i = \begin{pmatrix} \xi^i_x \\ \xi^i_\theta \\ \xi^i_\beta \end{pmatrix} \text{ with again } \xi^i = \xi^0 + \tilde{\xi}^i.$$

Here, $\beta$ varies for each member of the population, but we can force its uniformity by imposing $\tilde{\xi}^i_\beta$ to remain small, namely by scaling the covariance $\|(\tilde{P}^i_\diamond)_\beta\| = (\mathrm{Tr}((\tilde{P}^i_\diamond)^2_\beta))^{\frac{1}{2}} = O(\varepsilon)$.

### 2.3 Observations

Our objective consists in estimating the unknown part $\xi^i$ using the available medical data. In practice, we do not directly observe $x^i$ but we have at our disposal time-sampled observations $y^i$ indirectly related to $x^i$ through the measurement procedure. We model this process by introducing, at each measurement time $t_k \in [0, T]$, an observation operator $H_k$ so that

$$\mathcal{Y} \ni y^i_k = H_k\big(x^i(t_k)\big) + \chi^i_k, \quad 1 \le i \le N_P,\, 0 \le k \le N_{T,\mathrm{obs}},$$

where $\chi^i_k$ represents additive measurement errors. For the sake of simplicity when introducing the method formalism, the number of observation times $N_{T,\mathrm{obs}}$ does not depend on the patient. However in the numerical results section, we will explain how to consider patient-dependent observation times. We denote by $N_{\mathrm{obs}}$ the number of observations corresponding to the size of the vector $y^i_k$, $1 \le i \le N_P$, $0 \le k \le N_{T,\mathrm{obs}}$.

Here again, we can rewrite the system using the augmented state, namely

$$y^i_k = C_k\big(z^i(t_k)\big) + \chi^i_k, \quad 1 \le i \le N_P,\, 0 \le k \le N_{T,\mathrm{obs}},$$

where

$$C_k : \mathcal{X} \times \mathcal{P} \ni z = (x, \theta) \mapsto H_k(x) \in \mathcal{Y}.$$

In terms of uncertainties quantification, we again introduce covariances matrices $W^i_k$, $1 \le i \le N_P$, $0 \le k \le N_{T,\mathrm{obs}}$. And in order to define a consistent discrete-time white noise, we choose $\|(W^i_k)\| \propto \frac{1}{\Delta t_k}$ with $\Delta t_k = t_{k+1} - t_k$.

## 3 A population-based Kalman filter

### 3.1 A least-square functional

We gather all the patient uncertainties in a vector of dimension of $N_z \times N_P$

$$\boldsymbol{\xi} = \begin{pmatrix} \xi^1 \\ \vdots \\ \xi^{N_P} \end{pmatrix} \in (\mathcal{X} \times \mathcal{P})^{N_P} \simeq \mathbb{R}^{N_z \times N_P},$$

with bold notation indicating, in the rest of the presentation, quantities gathering all the patients together. Considering our uncertainty modeling, a strategy would be to rely on a maximum likelihood estimation where all disturbances are Gaussian. We will then minimize

$$\min_{\boldsymbol{\xi},\boldsymbol{\nu}} \left\{ \mathscr{J}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\big\langle \xi^0, (P^0_\diamond)^{-1}\,\xi^0 \big\rangle + \sum_{i=1}^{N_P} \left[ \frac{1}{2}\big\langle (\xi^i - \xi^0), (\tilde{P}^i_\diamond)^{-1}\,(\xi^i - \xi^0) \big\rangle + \frac{1}{2}\int_0^T \big\langle \nu^i(t), Q^i(t)^{-1}\nu^i(t) \big\rangle \,\mathrm{d}t \right.\right.$$
$$\left.\left. + \frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}} \big\langle y^i_k - C\big(z^i(t_k)\big), (W^i_k)^{-1}\big(y^i_k - C\big(z^i(t_k)\big)\big) \big\rangle \right] \right\}, \quad (3)$$

in order to estimate $\boldsymbol{\xi}$ and eventually the errors $\boldsymbol{\nu} = (\nu^i)_{1 \le i \le N_P}$. However, this minimization necessitates to know (or to estimate) the population intercept $\xi^0$ and the corresponding covariance $\tilde{P}^i_\diamond$. To overcome this constraint, we recall that the population intercept represents a weighted expected value of all the

patient uncertainties. We thus propose to model $\xi^0$ as a weighted empirical mean over all the patients. We set

$$\xi^0 = \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi}) \overset{\text{def}}{=} \sum_{i=1}^{N_P} \alpha^i \xi^i \text{ with } \sum_{i=1}^{N_P} \alpha^i = 1,$$

where $\boldsymbol{\alpha} = (\alpha^i)_{1\leq i\leq N_P}$ are weight coefficients mitigating the impact of each patient in the population characteristics. Therefore, our minimization problem in (3) becomes:

$$\begin{aligned}\min_{\boldsymbol{\xi},\boldsymbol{\nu}} \Bigg\{ \mathscr{J}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\big\langle \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi}), (P_\diamond^0)^{-1}\, \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi})\big\rangle \\ + \sum_{i=1}^{N_P}\Bigg[\frac{1}{2}\big\langle \xi^i - \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi}), (\tilde{P}_\diamond^i)^{-1}\,(\xi^i - \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi}))\big\rangle + \frac{1}{2}\int_0^T \big\langle \nu^i(t), Q^i(t)^{-1}\nu^i(t)\big\rangle\,\mathrm{d}t \\ + \frac{1}{2}\sum_{k=0}^{N_{T,\text{obs}}} \big\langle y_k^i - C\big(z^i(t_k)\big), (W_k^i)^{-1}(y_k^i - C\big(z^i(t_k)\big)\big\rangle\Bigg]\Bigg\}. \quad (4)\end{aligned}$$

We stress here that the functional in (4) is function of fewer unknowns than in (3), meaning that our coupling decreases the minimization space which is already rather large, namely $\mathbb{R}^{N_P\times N_z}\times L^2\big([0,T],\mathbb{R}^{N_P\times N_z}\big)$. The unknowns of our population formulation are given by $\xi^i$, $1\leq i\leq N_P$. As a matter of fact, the two first terms of the least square criterion can be developed into

$$\begin{aligned}&\big\langle \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi}), (P_\diamond^0)^{-1}\, \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi})\big\rangle + \sum_{i=1}^{N_P}\big\langle (\xi^i - \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi})), (\tilde{P}_\diamond^i)^{-1}\,(\xi^i - \mathbb{E}_{\boldsymbol{\alpha},N_P}(\boldsymbol{\xi}))\big\rangle \\ &\qquad = \sum_{i,j=1}^{N_P}\Big\langle \xi^i, \Big[\alpha^i\alpha^j(P_\diamond^0)^{-1} + \delta_{ij}(\tilde{P}_\diamond^i)^{-1} - (\alpha^i(\tilde{P}_\diamond^j)^{-1} + \alpha^j(\tilde{P}_\diamond^i)^{-1}) + \alpha^i\alpha^j\sum_{\ell=1}^{N_P}(\tilde{P}_\diamond^\ell)^{-1}\Big]\xi^j\Big\rangle.\end{aligned}$$

We thus introduce the matrix $\boldsymbol{P}_\diamond$ of dimension $(N_P\times N_z)^2$ defined by blocks such that for $1\leq i,j\leq N_P$, the block $(\boldsymbol{P}_\diamond^{-1})_{i,j}$ of indexes in $[(i-1)\,N_z+1, i\,N_z]\times\big[(j-1)\,N_z+1, j\,N_z\big]$ is given by

$$(\boldsymbol{P}_\diamond^{-1})_{i,j} = \alpha^i\alpha^j(P_\diamond^0)^{-1} + \delta_{ij}(\tilde{P}_\diamond^i)^{-1} - (\alpha^i(\tilde{P}_\diamond^j)^{-1} + \alpha^j(\tilde{P}_\diamond^i)^{-1}) + \alpha^i\alpha^j\sum_{\ell=1}^{N_P}(\tilde{P}_\diamond^\ell)^{-1}. \quad (5)$$

Furthermore, the assumption $\xi^0 = \sum_{i=1}^{N_P}\alpha^i\xi^i$ imposes from a statistical point of view that, for $1\leq i\leq N_P$, $P_\diamond^0 = \sum_{i=1}^{N_P}(\alpha^i)^2P_\diamond^i$ and $\tilde{P}_\diamond^i = P_\diamond^i - P_\diamond^0$. Thus, we can reformulate the whole estimation problem from a population point of view gathering all the patient states in $\boldsymbol{z} = (z^1\cdots z^{N_P})^\intercal$. On the one hand, the dynamics becomes

$$\begin{cases}\dot{\boldsymbol{z}}(t) = \boldsymbol{A}(\boldsymbol{z}(t),t) + \boldsymbol{B}(t)\boldsymbol{\nu}(t), \quad \forall t\in[0,T]\\ \boldsymbol{z}(0) = \boldsymbol{z}_0,\end{cases} \quad (6)$$

with

$$\boldsymbol{A}(\boldsymbol{z}(t),t) = \begin{pmatrix} A(z^1(t),t)\\ \vdots\\ A(z^{N_P}(t),t)\end{pmatrix} \text{ and } \boldsymbol{z}_0 = z_\diamond\times\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} + \boldsymbol{\xi},$$

and a particular weight matrix $\boldsymbol{P}_\diamond$ coupling all the patients together. On the other hand, the measurements are $\boldsymbol{y} = (y^1\cdots y^{N_P})^\intercal$ so that

$$\boldsymbol{y}_k = \boldsymbol{C}_k(\boldsymbol{z}(t_k)) + \chi_k, \quad 1\leq k\leq N_{T,\text{obs}},$$

with an observation operator

$$\boldsymbol{C}_k(\boldsymbol{z}(t_k)) = \begin{pmatrix} C_k(z^1)\\ \vdots\\ C_k(z^{N_P})\end{pmatrix} \text{ and a weight matrix } {}^\natural W = \begin{pmatrix} W^1 & & 0\\ & \ddots & \\ 0 & & W^{N_P}\end{pmatrix}.$$

In this context our estimator will be defined as the minimizer of the least-square functional (4) rewritten in the general form

$$\mathscr{J}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\big\langle \boldsymbol{\xi}, (\boldsymbol{P}_\diamond)^{-1}\,\boldsymbol{\xi}\big\rangle + \frac{1}{2}\int_0^T \big\langle \boldsymbol{\nu}(t), \boldsymbol{Q}(t)^{-1}\boldsymbol{\nu}(t)\big\rangle\,\mathrm{d}t + \frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}} \big\langle \boldsymbol{y}_k^i - \boldsymbol{C}_k(\boldsymbol{z}(t_k)), (\boldsymbol{W}_k)^{-1}(\boldsymbol{y}_k^i - \boldsymbol{C}_k(\boldsymbol{z}(t_k)))\big\rangle. \quad (7)$$

The key of our uncertainty modeling is that $\boldsymbol{P}_\diamond$ – given in (5) – couples the population members since indeed

$$\boldsymbol{P}_\diamond \neq \boldsymbol{P}_\diamond^{\mathrm{un}} = \mathrm{diag}(P_\diamond^1, \cdots, P_\diamond^{N_P}). \quad (8)$$

In the numerical part, we will compare our population strategy of parameters estimation with a non-population strategy which consists in considering instead of $\boldsymbol{P}_\diamond$, the uncoupled matrix $\boldsymbol{P}_\diamond^{\mathrm{un}}$.

## 3.2 Minimization of the functional and associated sequential optimal estimator

In practice, the minimization of (7) is performed after discretization of the dynamics (1). Considering the dynamics of interest in pharmacodynamic models, we rely on explicit time scheme, but our choice is not a restriction for the general presentation that we will make. Choosing a time step $\delta t$ small enough to satisfy the stability condition, we define as a discretization of (1) the scheme

$$\frac{x_{n+1}^i - x_n^i}{\delta t} = F(x_n^i, \theta^i, t_n) + G\nu^i(t_n), \quad 0 \leq n \leq N_T, \quad 1 \leq i \leq N_P,$$

with $t_n = n\delta t$, $N_T\,\delta t = T$. Therefore, there exists a discrete transition operator $\Phi_{n+1|n}$ giving the state $x_{n+1}^i$ knowing the state $x_n^i$, namely

$$x_{n+1}^i = \Phi_{n+1|n}(x_n^i, \theta^i) + G_{n+1}\nu_{n+1}^i, \quad 0 \leq n < N_T, \quad 1 \leq i \leq N_P, \quad (9)$$

with here $\Phi_{n+1|n}(x_n^i, \theta^i) = (x_n^i + \delta t F(x_n^i, \theta^i, t_n))$ and $G_n = \delta t G$. As presented before, we define from $\Phi_{n+1|n}$ a joint state-parameter transition operator $\Psi_{n+1|n}$ and the corresponding transition operator for the population $\boldsymbol{\Psi}_{n+1|n}$. Then, we discretize the criterion (7) using, for instance, an adequate left-centered quadrature rule and a model time-step compatible with the observation time step, namely $\frac{\Delta t}{\delta t} \in \mathbb{N}$. We get

$$\mathscr{J}_{N_T}(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\big\langle \boldsymbol{\xi}, (\boldsymbol{P}_\diamond)^{-1}\,\boldsymbol{\xi}\big\rangle + \frac{1}{2}\sum_{n=1}^{N_T} \big\langle \boldsymbol{\nu}_n, (\boldsymbol{Q}_n)^{-1}\boldsymbol{\nu}_n\big\rangle + \frac{1}{2}\sum_{k=0}^{N_{T,\mathrm{obs}}} \big\langle \boldsymbol{y}_k - \boldsymbol{C}_k(\boldsymbol{z}_{n_k}), (\boldsymbol{W}_k)^{-1}(\boldsymbol{y}_k^i - \boldsymbol{C}_k(\boldsymbol{z}_{n_k}))\big\rangle, \quad (10)$$

with typically $n_k = k\frac{\Delta t}{\delta t}$, and $\boldsymbol{Q}_n = \boldsymbol{Q}\delta t$. To simplify the notation between the model time discretization and the measurement time sampling, we now complement our observation at each model time step by considering

$$y_n = 0 \text{ and } H_n = 0 \text{ if } n\frac{\Delta t}{\delta t} \notin \mathbb{N}, \quad 0 \leq n \leq N_T.$$

Minimizing such functional under the constraint of a dynamics of the form (6) is very standard (actually known as the 4D-Var approach [3, 4, 1] in data assimilation). We denote by $\bar{\boldsymbol{z}}_{n|N_T}$, $1 \leq n \leq N_T$, the trajectory associated with the minimization of $\mathscr{J}_{N_T}$. It is also well known that a sequential estimation approach can equivalently give a recursive formulation of the so-called optimal sequential estimator defined by $\hat{\boldsymbol{z}}_n = \bar{\boldsymbol{z}}_{n|n}$. This is the Kalman estimator in the case where all operators are "linear"– more precisely: for all $n$, $\Phi_{n+1|n}$, $H_n$ and $G_n$ are linear [11]. With a nonlinear model operator and a nonlinear observation operator, the optimal sequential estimator can still be defined but to the price of an unaffordable computational complexity for large-dimensional systems [15]. Therefore, it is classical

to rely on approximate optimal sequential estimator based on the generalization of the Kalman filter to nonlinear operators [25]. Here, we find convenient to rely on the Unscented-Kalman Filter (UKF) where the mean and covariance operators are computed from empirical mean and covariance based on $N_\sigma$ sampling points – the so-called sigma points [10]. We define, for a given set of positive coefficient $\boldsymbol{\alpha} = (\omega_i)_{1\leq i\leq N_\sigma}$ and a given set of particles $(v^{(i)})_{1\leq i\leq N_\sigma}$, the empirical mean and covariance by

$$\mathbb{E}_{\boldsymbol{\alpha},N_\sigma}(v^{(\cdot)}) \stackrel{\text{def}}{=} \sum_{i=1}^{N_\sigma} \alpha^i v^{(i)}, \quad \mathbb{C}\text{ov}_{\boldsymbol{\omega},N_\sigma}(v^{(\cdot)}) \stackrel{\text{def}}{=} \sum_{i=1}^{N_\sigma} \omega^i \Big(v^{(i)} - \mathbb{E}_{\boldsymbol{\alpha}}(v^{(\cdot)})\Big)\Big(v^{(i)} - \mathbb{E}_{\boldsymbol{\alpha}}(v^{(\cdot)})\Big)^{\intercal}.$$

Let us now introduce a set of positive weight coefficients $\boldsymbol{\omega} = (\omega_i)_{1\leq i\leq N_\sigma}$ with $\sum_{i=1}^{N_\sigma} \omega_i = 1$, and a set of unitary sigma points $(\boldsymbol{e}^{(i)})_{1\leq i\leq N_\sigma} \in ((\mathcal{X}\times\mathcal{P})^{N_P})^{N_\sigma}$ with the following empirical mean and covariances

$$\mathbb{E}_{\boldsymbol{\omega},N_\sigma}(\boldsymbol{e}^{(\cdot)}) = \sum_{i=1}^{N_\sigma} \omega^i \boldsymbol{e}^{(i)} = 0 \text{ and } \mathbb{C}\text{ov}_{\boldsymbol{\omega},N_\sigma}(\boldsymbol{e}^{(\cdot)}) = \sum_{i=1}^{N_\sigma} \omega^i \boldsymbol{e}^{(i)}\boldsymbol{e}^{(i)\intercal} = \mathbb{1}.$$

The UKF algorithm reads

$$\begin{cases}
\text{Initialization:} & \\
\hat{\boldsymbol{z}}_0^- = x_\diamond \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \boldsymbol{P}_0^- = \boldsymbol{P}_\diamond, & \\
\text{Sampling:} & \\
\tilde{\boldsymbol{y}}_n^{(i)} = \boldsymbol{y}_n - \boldsymbol{C}_n\big(\hat{\boldsymbol{z}}_n^- + \sqrt{\lambda_i \boldsymbol{P}_n^-}\boldsymbol{e}^{(i)}\big), & 0\leq n\leq N_T, 1\leq i\leq N_\sigma, \\
\hat{\boldsymbol{z}}_n^{(i)-} = \hat{\boldsymbol{z}}_n^- + \sqrt{\lambda_i \boldsymbol{P}_n^-}\boldsymbol{e}^{(i)}, & 0\leq n\leq N_T, 1\leq i\leq N_\sigma, \\
\text{Correction:} & \\
\boldsymbol{K}_n = \mathbb{C}\text{ov}_{\boldsymbol{\omega}}\big(\hat{\boldsymbol{z}}_n^{(\cdot)-}, \tilde{y}_n^{(\cdot)}\big)[\mathbb{C}\text{ov}_{\boldsymbol{\omega},N_\sigma}\big(\tilde{y}_n^{(\cdot)}\big) + \boldsymbol{W}_n]^{-1}, & 0\leq n\leq N_T, \\
\hat{\boldsymbol{z}}_n^+ = \hat{\boldsymbol{z}}_n^- + \boldsymbol{K}_n\big(\boldsymbol{y}_n - \boldsymbol{C}_n(\hat{\boldsymbol{z}}_n^-)\big), & 0\leq n\leq N_T, \\
\boldsymbol{P}_n^+ = \boldsymbol{P}_n^- - \boldsymbol{K}_n[\mathbb{C}\text{ov}_{\boldsymbol{\omega},N_\sigma}\big(\boldsymbol{y}_n^{(\cdot)}\big) + \boldsymbol{W}_n]\boldsymbol{K}_n^{\intercal}, & 0\leq n\leq N_T, \\
\text{Sampling:} & \\
\hat{\boldsymbol{z}}_n^{(i)+} = \hat{\boldsymbol{z}}_n^+ + \sqrt{\lambda_i \boldsymbol{P}_n^+}\boldsymbol{e}^{(i)}, & 0\leq n\leq N_T, 1\leq i\leq N_\sigma, \\
\text{Prediction:} & \\
\hat{\boldsymbol{z}}_{n+1}^{(i)-} = \boldsymbol{\Psi}_{n+1|n}\big(\hat{\boldsymbol{z}}_n^{(i)+}\big) + \natural\beta_n, & 0\leq n\leq N_T, 1\leq i\leq N_\sigma, \\
\hat{\boldsymbol{z}}_{n+1}^- = \mathbb{E}_{\boldsymbol{\omega},N_\sigma}\big(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\big), & 0\leq n\leq N_T, \\
\boldsymbol{P}_{n+1}^- = \mathbb{C}\text{ov}_{\boldsymbol{\omega},N_\sigma}\big(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\big) + \boldsymbol{B}_{n+1}\boldsymbol{Q}_{n+1}\boldsymbol{B}_{n+1}^{\intercal}, & 0\leq n\leq N_T.
\end{cases} \tag{11}$$

Note here that the sigma points are of the same order of magnitude that $N_P$ but they do not correspond to each patient. Indeed in order to account for the coupling between patients, each sigma-point mixes the patient information.

Finally, we emphasize that in practice, we can extract from $\boldsymbol{P}_n^+$ the time evolutions of the covariance matrices $P_n^i$, $\tilde{P}_n^i$ and $P_n^0$ associated with the times evolutions of the uncertainties $\xi_n^i$, $\tilde{\xi}_n^i$ and $\xi_n^0 = \sum_{i=1}^{N_P} \alpha^i \xi_n^i$. Indeed, as the estimation process is applied to $\boldsymbol{\xi}$, each block $(\boldsymbol{P}_n^+)_{i,j}$, $1\leq i,j\leq N_P$ of indexes in $[(i-1)\,N_z+1, i\,N_z]\times[(j-1)\,N_z+1, j\,N_z]\Big]$ can be seen as $P_n^{ij}$, the covariance matrix between the patient $i$ and the patient $j$. In particular, we have $P_n^i = P_n^{ii}$. As $\xi_n^0 = \sum_{i=1}^{N_P} \alpha^i \xi_n^i$, the covariance matrix associated with $\xi_n^0$ equals to $P_n^0 = \sum_{i,j=1}^{N_P} \alpha^i\alpha^j P_n^{ij}$ and finally, this implies that

$$\tilde{P}_n^i = P_n^i - P_n^0.$$

This decomposition will be used in the numerical results to evaluate how our strategy estimates the so-called fixed and mixed effects of the parameters.

### 3.3 A reduced-order version of the sequential optimal estimator

The computation of the full covariance matrix $\boldsymbol{P}$ can be prohibitive when the number of unknowns, parameters, and patients increase. In this section, we therefore propose a strategy of cost reduction. Assuming that $\boldsymbol{P}$ is of reduced rank – typically much smaller than the dimension of the space $N_P \times N_z$ – the main idea in reduced-order filtering is to be able to manipulate covariance matrices in the factorized form

$$\boldsymbol{P} = \boldsymbol{L}\boldsymbol{U}^{-1}\boldsymbol{L}^{\intercal},$$

with $\boldsymbol{U}$ an invertible matrix of small size which represents the main uncertainties in the system and $\boldsymbol{L}$ an extension operator. What is crucial here is to be able to perform all computations on $\boldsymbol{L}$ and $\boldsymbol{U}$ without needing to compute $\boldsymbol{P}$. To do so, we will use the reduced-order Unscented-Kalman Filter (RoUKF) [16, 19]. The only missing step consists in defining the initial reduced covariance matrix $\boldsymbol{U}_\diamond$ and the initial extension matrix $\boldsymbol{L}_\diamond$ from the initial covariance matrix $\boldsymbol{P}_\diamond$.

Our strategy is based on a clustering approach applied to the observations sequence using the *k-means* algorithm. More precisely, the *k-means* algorithm is applied to the following matrix

$$\boldsymbol{Y} = \begin{pmatrix} y^1 \\ \vdots \\ y^{N_P} \end{pmatrix}, \text{ where for } i \in \{1, \cdots, N_P\},$$

$$y^i = ((y_1^i)_1, \cdots, (y_{N_{T,\text{obs}}}^i)_1, \ \cdots \ , (y_1^i)_{N_{\text{obs}}}, \cdots, (y_{N_{T,\text{obs}}}^i)_{N_{\text{obs}}}),$$

with $(y_k^i)_l$ corresponds to the $l^{-th}$ coordinate of $y_k^i$, $\quad 1 \leq i \leq N_P, 1 \leq k \leq N_{T,\text{obs}}$. This clustering allows to separate the $N_P$ patients in $N_C$ clusters. For $1 \leq s \leq N_C$, we define $c_s$ the $s^{th}$-cluster, $N_{c_s}$ the number of patients belonging to the $s^{th}$-cluster.

We then define the reduced initial covariance matrix $\boldsymbol{U}_\diamond$ of dimension $(N_C \times N_z)^2$ defined by blocks such that for $1 \leq r, s \leq N_C$, the block $(\boldsymbol{U}_\diamond)_{r,s}$ of indexes in $[(r-1)\,N_z + 1, r\,N_z] \times \left[(s-1)\,N_z + 1, s\,N_z\right]$ is given by

$$(\boldsymbol{U}_\diamond)_{r,s} = \frac{1}{N_{c_r} N_{c_s}} \sum_{i \in c_r} \sum_{j \in c_s} (\boldsymbol{P}_\diamond)_{i,j}. \tag{12}$$

If we consider for the extension operator $\boldsymbol{L}_\diamond$ the matrix of dimension $(N_P \times N_z) \times (N_C \times N_z)$ defined by blocks such that for $1 \leq i \leq N_P$ and for $1 \leq r \leq N_C$, the block $\boldsymbol{L}_{i,r}$ of indexes in $[(i-1)\,N_z + 1, i\,N_z] \times \left[(r-1)\,N_z + 1, r\,N_z\right]$ is given by

$$(\boldsymbol{L}_\diamond)_{i,r} = \begin{cases} \mathbb{1}_{N_z,N_z}, & \text{if } i \in c_r, \\ \mathbb{0}_{N_z,N_z}, & \text{else,} \end{cases}$$

the dynamics of all the patients belonging to the same cluster will be the same. This strong constraint cannot give acceptable results. As the *k-means* algorithm uses the squared Euclidean distance metric, we compute how much each patient $i \in [1, N_P]$ belongs to one cluster $c_s$ with $s \in [1, N_C]$ by computing the following weight

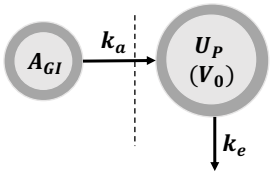
$$\beta_{i,s} = \begin{cases} \dfrac{\|y^i - \frac{1}{N_{c_s}} \sum_{j \in c_s} y^j\|_2^{-1}}{\sum_{r=1}^{N_C} \|y^i - \frac{1}{N_{c_r}} \sum_{j \in c_r} y^j\|_2^{-1}}, & \text{if } N_{c_s} > 1, \\ 1, & \text{else.} \end{cases}$$

Finally, we define the block $(\boldsymbol{L}_\diamond)_{i,r}$ of indexes in $[(i-1)\,N_z + 1, i\,N_z] \times [(r-1)\,N_z + 1, r\,N_z]$ by

$$(\boldsymbol{L}_\diamond)_{i,r} = \beta_{i,r} \mathbb{1}_{N_z,N_z}.$$

We conclude this section by giving our reduced estimator algorithm based on the reduced-order Unscented-Kalman Filter (RoUKF) [16]. We emphasized that in this algorithm, the number of sigma-points $N_\sigma^r \ll N_\sigma$ used in the full UKF approach (11). Therefore, the associated unitary sigma-points $\boldsymbol{e}_r^{(i)}$

(a)




| | | |
|---|---|---|
| $A_{GI}$ | Dose in Gastro-intestinal compartment | mg |
| $U_p$ | Concentration Plasma compartment | mg/L |
| $k_a$ | Drug absorption rate | $h^{-1}$ |
| $k_e$ | Drug elimination rate | $h^{-1}$ |
| $V_0$ | Volume of distribution | L |

(b)

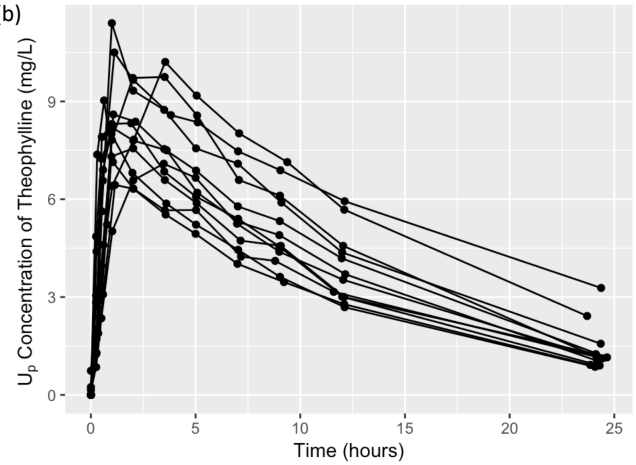



Figure 1: One-compartment PK model with oral absorption: (a) Representation of the mathematical model, (b) A benchmark dataset of the anti-asthmatic drug theophylline for 12 patients.

generate a subspace of small dimension $N_\sigma^r$:

$$
\begin{cases}
\text{Initialization:} & \\
\hat{\boldsymbol{z}}_0^- = x_\diamond \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \boldsymbol{L}_0 = \boldsymbol{L}_\diamond, \quad \boldsymbol{U}_0 = \boldsymbol{U}_\diamond, & \\
\text{Sampling:} & \\
\tilde{\boldsymbol{y}}_n^{(i)} = \boldsymbol{y}_n - \boldsymbol{C}_n\big(\hat{\boldsymbol{z}}_n + \boldsymbol{L}_n\sqrt{\boldsymbol{U}_n^{-1}}\boldsymbol{e}^{(i)}\big), & 0 \le n \le N_T, 1 \le i \le N_\sigma^r, \\
\text{Correction:} & \\
\boldsymbol{\Lambda}_{n+1} = \mathbb{C}\mathrm{ov}_{\boldsymbol{\omega},N_\sigma^r}\big(\boldsymbol{e}^{(\cdot)}, \tilde{\boldsymbol{y}}_n^{(\cdot)}\big)^T(\boldsymbol{W}_n)^{-1}, & 0 \le n \le N_T, \\
\boldsymbol{U}_{n+1} = \mathbb{1} + \boldsymbol{\Lambda}_{n+1}\mathbb{C}\mathrm{ov}_{\boldsymbol{\omega},N_\sigma^r}\big(\boldsymbol{e}^{(\cdot)}, \tilde{\boldsymbol{y}}_n^{(\cdot)}\big), & 0 \le n \le N_T, \\
\hat{\boldsymbol{z}}_n^+ = \hat{\boldsymbol{z}}_n^- + \boldsymbol{L}_n\boldsymbol{U}_{n+1}^{-1}\Lambda_{n+1}\big(\boldsymbol{y}_n - \boldsymbol{C}_n(\hat{\boldsymbol{z}}_n^-)\big), & 0 \le n \le N_T, \\
\text{Sampling:} & \\
\hat{\boldsymbol{z}}_n^{(i)+} = \hat{\boldsymbol{z}}_n + \boldsymbol{L}_n\sqrt{\boldsymbol{U}_n^{-1}}\boldsymbol{e}^{(i)}, & 0 \le n \le N_T, 1 \le i \le N_\sigma^r, \\
\text{Prediction:} & \\
\hat{\boldsymbol{z}}_{n+1}^{(i)-} = \boldsymbol{\Psi}_{n+1|n}\big(\hat{\boldsymbol{z}}_n^{(i)+}\big), & 0 \le n \le N_T, 1 \le i \le N_\sigma^r, \\
\hat{\boldsymbol{z}}_{n+1}^- = \mathbb{E}_{\boldsymbol{\omega},N_\sigma^r}\big(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\big), & 0 \le n \le N_T, \\
\boldsymbol{L}_{n+1} = \mathbb{C}\mathrm{ov}_{\boldsymbol{\omega},N_\sigma^r}\big(\boldsymbol{e}^{(\cdot)}, \hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\big), & 0 \le n \le N_T.
\end{cases}
\tag{13}
$$

# 4 Numerical results and discussion

## 4.1 Pharmacokinetics one-compartment model

To illustrate the proposed strategy, we will consider a pharmacokinetics one-compartment model with first-order absorption and elimination. We denote by $A_{GI}$, the amount of drug in the gastrointestinal compartment (in mg) and by $A_P$ the amount of drug in the plasma (in mg). The interactions between $A_{GI}$ and $A_P$ are described in the following system:

$$
\begin{cases}
\dot{A}_{GI} &= -k_a A_{GI}, \\
\dot{A}_P &= k_a A_{GI} - k_e A_P,
\end{cases}
\tag{14}
$$

where $k_a$ corresponds to the absorption rate (in $\mathrm{h}^{-1}$) and $k_e$ to the elimination rate (in $\mathrm{h}^{-1}$). From a clinical practice, only the plasma concentration $U_P = \frac{A_P}{V_0}$ (in $\mathrm{mg.L}^{-1}$) is supposed observed at $N_{T,\mathrm{obs}}$ observation times, where $V_0$ is the volume of the plasma compartment (in L). System (14) can be rewritten as follows:

$$\left\{ \begin{array}{rcl} \dot{A}_{GI} & = & -k_a A_{GI}, \\ \dot{U}_P & = & \frac{k_a}{V_0} A_{GI} - k_e U_P, \end{array} \right. \tag{15}$$

see Figure 1(a) for an illustration.

**Remark 2.** Obviously, the analytical solution of System (15) is given by:

$$A_{GI}(t) = A_{GI}(0)e^{-k_a t} \text{ and } U_P(t) = \frac{A_{GI}(0)k_a}{V_0(k_a - k_e)}(e^{-k_e t} - e^{-k_a t}), \forall t.$$

However, in this numerical part, we will solve it using the first order Euler time scheme as we want to validate Kalman based-filters which are by definition written on the time derivative part of ODE or PDE systems (or their discretizations).

Concerning the initial values, $A_{GI}(0)$ equals to a known initial dose given in $mg$ given by oral absorption and $U_P(0) = 0$. The parameters $k_a, k_e$ and $V_0$ are unknown and will be estimated using the available observations. These parameters are all positive which will be constrained by a log-transformation during the estimation procedure further on. Following the notations of Section 2 and considering a population of $N_P$ patients, we have in this context,

$$z^i = \begin{pmatrix} A^i_{GI} \\ U^i_P \\ \ln(k^i_a) \\ \ln(k^i_e) \\ \ln(V^i_0) \end{pmatrix}, \; A(z^i(t), t) = \begin{pmatrix} -k_a A_{GI} \\ \frac{k_a}{V_0} A_{GI} - k_e U_P \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad B(t) = 0, \;\; 1 \leq i \leq N_P,$$

and

$$C^i_k = \begin{pmatrix} 0 & U^i_P(t_k)^{0.25} & 0 & 0 & 0 \end{pmatrix}, 1 \leq i \leq N_P, \; 0 \leq k \leq N_{T,\mathrm{obs}}.$$

We recall that the variance of the error measurements is denoted by $W^i_k$ for each subject $i$ and for each observation $k$. For the considered observations, the matrices are of dimension $1 \times 1$ and will be considered as constant: $W^i_k = w, 1 \leq i \leq N_P, \; 0 \leq k \leq N_{T,\mathrm{obs}}$.

Another important aspect is the ability to handle the time-sampling of the data, which in practice is imposed and can be quite different from the time-step used in the model and different for each patient. In order to handle the data time-sampling, two strategies are conceivable, namely, we can either use the data only when they are available, or we can rely on some time-interpolation. As we also have to deal with the fact that the time-sampling of the observations depends on each patient, we decide to interpolate the observations. More precisely, using the time discretization notations of the system introduced in Section 3 we will consider, $1 \leq i \leq N_P$, $0 \leq n < N_T$,

$$C^i_n = \begin{pmatrix} 0 & ((1-\alpha_n)U^i_P(t_k) + \alpha_n U^i_P(t_{k+1}))^{0.25} & 0 & 0 & 0 \end{pmatrix},$$

with $\alpha_n = \frac{t_n - t_k}{t_n - t_{k+1}}$ and with $t_k, t_{k+1}$ are the data sampling times adjacent to the simulation time $t_n$.

## 4.2 Validation using synthetic data

**Data and statistical validation criteria** – We consider a gradient of simulations with different numbers of patients ($N_P$=2, 20 and 100) and with different measurement errors (standard deviation of an additive Gaussian noise $w_t = 0.3$ and 1). We generate 100 simulation replicates ($N_R = 100$) for each scenario. The observation times are 30, 60, 90, 120, 180, 240, 360, 480 and 600 minutes. The synthetic data (or simulation replicates) are generated using

$$z_\diamond = \begin{pmatrix} (A_{GI})_\diamond \\ (U_P)_\diamond \\ \ln(k_a)_\diamond \\ \ln(k_e)_\diamond \\ \ln(V_0)_\diamond \end{pmatrix} = \begin{pmatrix} 500 \\ 0 \\ -4.6 \\ -5.56 \\ -4.19 \end{pmatrix} \text{ and } P^i_\diamond = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2^2 & 0 & 0 \\ 0 & 0 & 0 & 0.25^2 & 0 \\ 0 & 0 & 0 & 0 & 0.1^2 \end{pmatrix}, \; 1 \leq i \leq N_P.$$

as mean values and covariances of the initial state conditions and of the parameters. The matrices $P_\diamond^i,\ 1 \leq i \leq N_P$ are diagonal because we assume that the parameters are independent. In the estimation process, we will consider the following *a-priori* values and covariances for the estimation:

$$z_\diamond = \begin{pmatrix} (A_{GI})_\diamond \\ (U_P)_\diamond \\ \ln(k_a)_\diamond \\ \ln(k_e)_\diamond \\ \ln(V_0)_\diamond \end{pmatrix} = \begin{pmatrix} 500 \\ 0 \\ -3 \\ -5 \\ -3 \end{pmatrix} \text{ and } P_\diamond^i = \begin{pmatrix} 0.001^2 & 0 & 0 & 0 & 0 \\ 0 & 0.001^2 & 0 & 0 & 0 \\ 0 & 0 & 3^2 & 0 & 0 \\ 0 & 0 & 0 & 3^2 & 0 \\ 0 & 0 & 0 & 0 & 3^2 \end{pmatrix},\ 1 \leq i \leq N_P,$$

and for the standard deviation of the measurements noise $w = 10$.

In that follows, we will evaluate the performances of Algorithm 11 using the coupled covariance matrix $\boldsymbol{P}_\diamond$ (see Eq. (5)) (population estimation) and the uncoupled ones $\boldsymbol{P}_\diamond^{\text{un}}$, see Eq. (8) (individual estimation) and the performances of Algorithm (13) by applying the clustering algorithm to the coupled covariance matrix $\boldsymbol{P}_\diamond$ (reduced-order population estimation).

To validate the strategies, various statistical validation criteria will be considered. For a parameter $\theta$, we define by $\theta_t$ (resp. $\theta_e$) the matrix of size $N_R \times N_P$ containing on each line $r$ the target (resp. estimated) values of $\theta$ for all the patients of the replicate $r$. The relative bias RBIAS and the mean squared error MSE

$$\text{RBIAS} = \frac{1}{N_R N_P} \sum_{r=1}^{N_R} \sum_{i=1}^{N_P} \frac{\theta_t(r,i) - \theta_e(r,i)}{\theta_t(r,i)} \text{ and MSE} = \frac{1}{N_R N_P} \sum_{r=1}^{N_R} \sum_{i=1}^{N_P} (\theta_t(r,i) - \theta_e(r,i))^2$$

will be computed. The RBIAS describes the deviation of the mean over the estimated parameters from their true values. The MSE summarizes both the bias and the variability in estimates. Using $P_n^i$, we compute $\sigma_e^\theta$ the matrix of size $N_R \times N_P$ containing on each line $r$ the estimated standard deviation of $\theta$ for all the patients of the replicate $r$. We will additionally report to the RBIAS and MSE, the mean estimated standard deviation (STD)

$$\text{STD} = \frac{1}{N_R N_P} \sum_{r=1}^{N_R} \sum_{i=1}^{N_P} \sigma_e^\theta(r,i),$$

and the empirical standard deviation of estimates defined as the standard deviation of the estimates at convergence (ESTD)

$$\text{ESTD} = \sqrt{\frac{1}{N_R N_P} \sum_{r=1}^{N_R} \sum_{i=1}^{N_P} \Big(\theta_e(r,i) - \frac{1}{N_R N_P} \sum_{n=1}^{N_R} \sum_{i=1}^{N_P} \theta_e(r,i)\Big)^2}.$$

For some situations, the relative bias on standard deviation for random effects called BMIXED and defined by

$$\text{BMIXED} = \frac{1}{N_R N_P} \sum_{r=1}^{N_R} \sum_{i=1}^{N_P} \Big(\sigma_{e,m}^\theta(r,i) - \sigma_t^\theta(r,i)\Big),$$

will be given, where $\sigma_{e,m}^\theta$ is the matrix of size $N_R \times N_P$ containing on each line $r$ the estimated mixed standard deviation of $\theta$ for all the patients of the replicate $r$ which is built using $\tilde{P}_n^i$. We also define the coverage (COV) defined as the percentage of time the true value is included in the 95% confidence interval of the estimator values.

**Population Kalman filter validation** − Table 1 summarizes the results for the 6 scenarios: $(N_P, w_t) \in \{(2, 0.3), (20, 0.3), (100, 0.3), (2, 1), (20, 1), (100, 1)\}$. For the sake of readability, the statistical criteria are aggregated over the 3 parameters. As expected, for all sample sizes, estimates are similar in terms of precision and accuracy when using an individual approach. We can note that the individual Kalman filter cannot allow to properly estimate the parameters. This is also represented in Figure 2 where the time evolution of states and parameters for one patient is plotted. The fail of the estimation is due to the fact that the *a-priori* of the parameters are really vague. As we will see with real data when the parameters *a-priori* are better determined, the individual Kalman filter gives acceptable results.

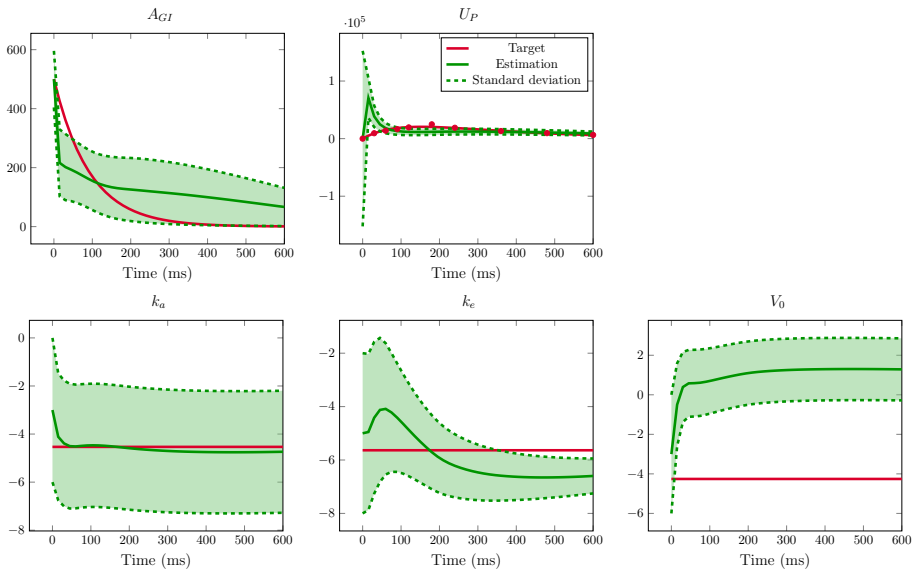



Figure 2: Time evolution of states $A_{GI}$ and $U_P$ and parameters $k_a, k_e$ and $V_0$ for one patient using individual estimation with $N_P$=20 and $w_t = 0.3$. The target states and target values are in red. The red dots on the graph of $U_P$ corresponds to the noisy observed data. The states and parameters values with individual estimation are in green. The dashed lines correspond to the standard deviations.

| Meas. | Sample | Individual | | | | Population | | | |
|---|---|---|---|---|---|---|---|---|---|
| Error | Size | RBIAS | MSE | STD | ESTD | RBIAS | MSE | STD | ESTD |
| $w_t = 0.3$ | $N_P$=2 | 0.375 | 10.471 | 2.534 | 0.180 | 0.664 | 18.629 | 1.352 | 0.549 |
| | $N_P$=20 | 0.376 | 10.406 | 1.578 | 0.259 | -0.009 | 0.085 | 1.676 | 0.195 |
| | $N_P$=100 | 0.377 | 10.446 | 1.578 | 0.259 | 0.0004 | 0.0632 | 0.9037 | 0.1837 |
| $w_t = 1.0$ | $N_P$=2 | 0.374 | 10.371 | 1.582 | 0.360 | 0.686 | 19.445 | 1.365 | 0.579 |
| | $N_P$=20 | 0.373 | 10.510 | 1.583 | 0.380 | -0.006 | 0.127 | 0.682 | 0.301 |
| | $N_P$=100 | 0.374 | 10.517 | 1.583 | 0.370 | -0.007 | 0.134 | 0.782 | 0.339 |

Table 1: Performances of the estimation algorithms evaluated over 100 replicates in 6 different scenarios. Summary statistics are the relative bias (RBIAS), the mean squared error (MSE), the mean estimated standard deviation (STD) and the empirical standard deviation of estimates (ESTD). Results are aggregated over all the parameters of the model.

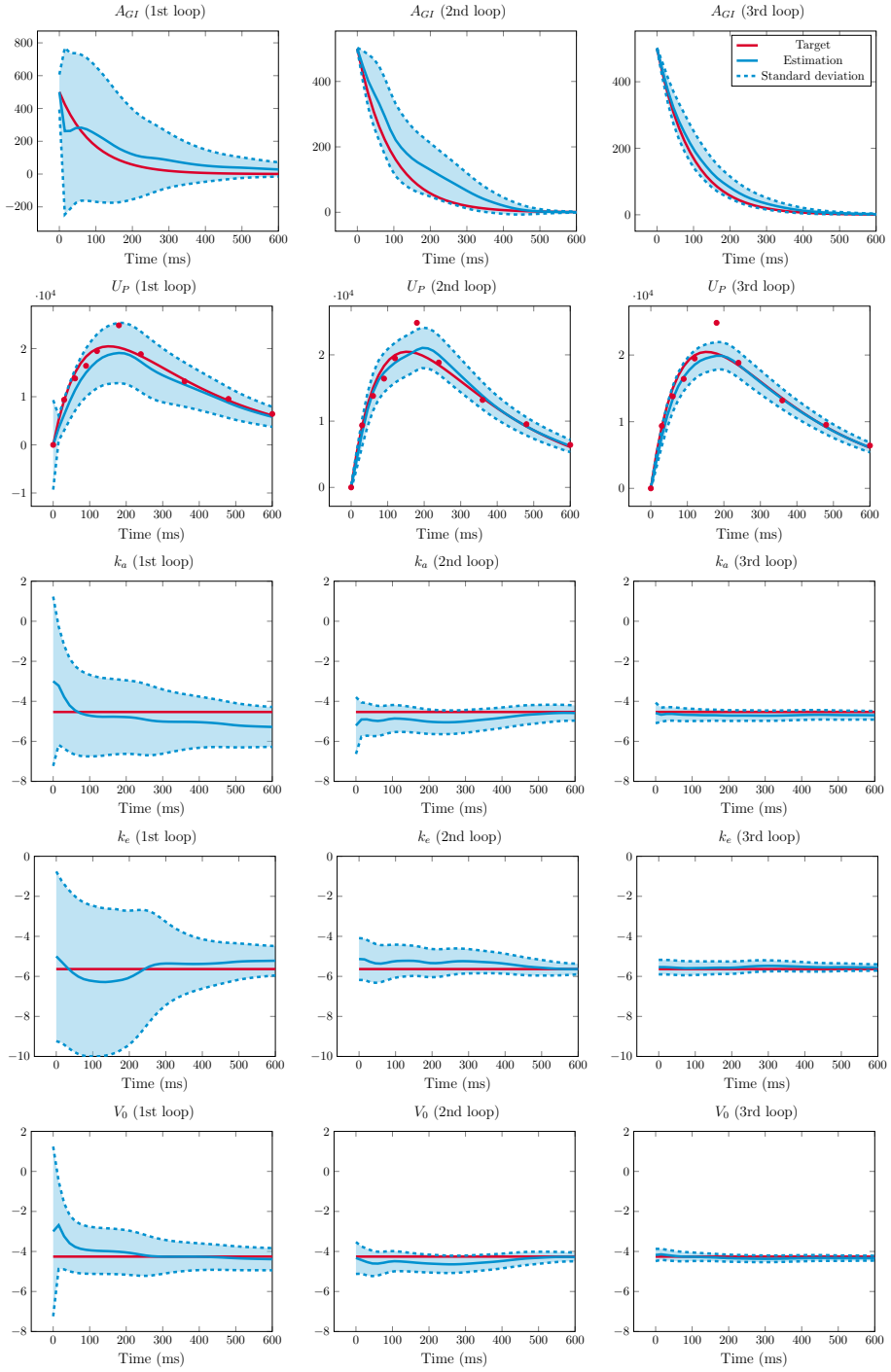



Figure 3: Time evolution of states $A_{GI}$ and $U_P$ and parameters $k_a, k_e$ and $V_0$ for one patient using population estimation with $N_P$=20 and $w_t = 0.3$. Each column correspond to one loop of the algorithm. The target states and target values of parameters are in red. The red dots on the graph of $U_P$ corresponds to the noisy observed data. The states and parameters values with population estimation are in blue. The dashed lines correspond to the standard deviations.

| | | Population | | | | | |
|---|---|---|---|---|---|---|---|
| Loop | Parameter | RBIAS | MSE | STD | ESTD | BMIXED | COV |
| 1 | $k_a$ | -0.064 | 0.169 | 0.816 | 0.233 | 0.816 | 100.0 |
| | $k_e$ | 0.030 | 0.063 | 0.686 | 0.192 | 0.686 | 100.0 |
| | $V_0$ | 0.007 | 0.025 | 0.524 | 0.161 | 0.524 | 100.0 |
| 2 | $k_a$ | -0.001 | 0.024 | 0.335 | 0.215 | 0.335 | 100.0 |
| | $k_e$ | -0.004 | 0.018 | 0.241 | 0.224 | 0.241 | 99.8 |
| | $V_0$ | 0.005 | 0.013 | 0.175 | 0.132 | 0.175 | 99.4 |
| 3 | $k_a$ | -0.008 | 0.023 | 0.120 | 0.154 | 0.200 | 99.0 |
| | $k_e$ | -0.001 | 0.017 | 0.143 | 0.173 | 0.1432 | 96.7 |
| | $V_0$ | 0.002 | 0.010 | 0.110 | 0.104 | 0.111 | 96.6 |

Table 2: Performances of the population estimation algorithm evaluated over 100 replicates for $N_P$=20 and $w_t = 0.3$ using three loops. Summary statistics are the relative bias for parameters (RBIAS), the mean squared error (MSE), the mean estimated standard deviation (STD), the empirical standard deviation of estimates (ESTD), the standard deviation for random effects (BMIXED) and the coverage (COV).

The population approach really strengthen the estimation as RBIAS and MSE decrease with sample size and are dramatically smaller than for an individual approach when sample size increases. We can see that increasing the noise increases the MSE (even if the RBIAS values are close) implying that it is more difficult to well estimate the parameters if the noise is too important. We notice that in this example, STD is highly bigger than ESTD. It denotes an overestimation of the standard deviation of estimates. This is due to the fact that the initial covariances of the parameters are really vague ($3^2$ instead of $0.2^2, 0.25^2$ and $0.1^2$ for respectively $k_a, k_e$ and $V_0$). The first column of Figure 3 corresponds to the time evolution of states and parameters for one patient. Thus, in order to estimate the most accurate covariance of final estimate we will have to use a sequential Bayesian approach as described in the literature[23]. This consists in running multiple times the estimation algorithm described in Section 3, by taking as new *a-priori* the *a-posteriori* of the previous estimation. Each run will be called a loop. The population procedure is evaluated for $w_t = 0.3$ and $N_P$=20. Table 2 summarizes the results for each parameter of the model. Additionally to Table 1, BMIXED and COV are also given. Results are more accurate as RBIAS and MSE decrease when multiple loops are used. The second (resp. third) column of Figure 3 shows the time evolution of states and parameters for one patient for the second (resp. third) loop. Moreover, the STD get closer to the ESTD and the COV get closer to their nominal value of 95%. Regarding the results, we recommend to perform multiple loops for the estimation when the *a-priori* are very vague, as illustrated on Figure 3. Estimation of the standard error of random effect is also less biased with 3 loops as shown by BMIXED but still important. Maybe this can be explained by the fact that we do not minimize the full functional (3) containing the population intercept but the functional (4) assuming the population intercept can be seen as an empirical mean over all the patients. It would be a perspective of this work to compare our results using the full functional.

**Reduced-order population Kalman filter validation** – The results obtained with the population Kalman filter are very encouraging but the computational time is still too important. The objective of this section is to numerically illustrate the reduced-order version of the population Kalman filter introduced in Section 3.3. We will focus on the scenario of $N_P$=100 and $\sigma_m = 0.3$. Figure 4 illustrates the clustering algorithm (for $N_C = 20$ and $N_C = 5$) by showing the observation variable $U_P^{0.25}$ for a replicate. Summary statistics are the relative bias for parameters (RBIAS), the mean squared error (MSE), the mean estimated standard deviation (STD), the empirical standard deviation of estimates (ESTD), the standard deviation for random effects (BMIXED) and the coverage (COV). Table 3 presents the results which are given for each parameter and also aggregated over the 3 parameters. The computational times are also given. If we focus on the MSE, we obtain equivalent results (even a little bit better with the reduced versions). Concerning the RBIAS, the results obtained with the reduced versions are less good but still acceptable. As predictable, both standard deviations criteria (STD and ESTD) decrease with the reduced-order filter but theirs values are also acceptable. Concerning the ESTD, it is even better.

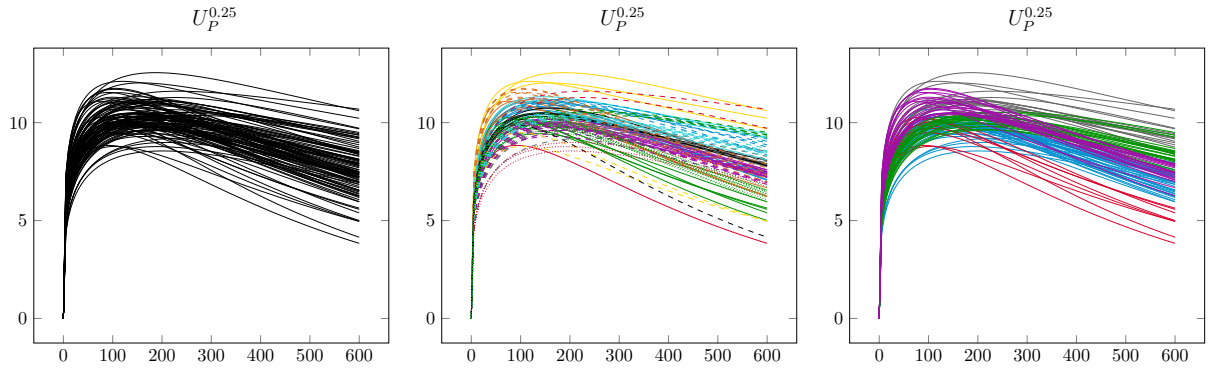



Figure 4: Illustration of the clustering algorithm (*k-means* algorithm) on the observation variable $U_P^{0.25}$ for a synthetic data set of 100 patients. Left: full data set. Middle: results of *k-means* algorithm with 20 clusters. Right: results of *k-means* algorithm with 5 clusters. Each color and style combination corresponds to one cluster.

| | | Population | | | | | |
|---|---|---|---|---|---|---|---|
| Version | Parameter | RBIAS | MSE | STD | ESTD | BMIXED | COV |
| Full | $k_a$ | -0.0452 | 0.114 | 1.4 | 0.211 | 1.08 | 100 |
| 13 min | $k_e$ | 0.0237 | 0.0509 | 0.697 | 0.201 | 0.25 | 100 |
| (laptop) | $V_0$ | 0.0228 | 0.0247 | 0.614 | 0.139 | 0.412 | 100 |
| | aggreg. | 0.0004 | 0.0632 | 0.9037 | 0.1837 | 0.5807 | 100 |
| Reduced | $k_a$ | -0.00974 | 0.0461 | 0.627 | 0.13 | -0.124 | 100 |
| ($N_P = 20$) | $k_e$ | -0.00347 | 0.0445 | 0.486 | 0.101 | -0.202 | 99.9 |
| 3 min | $V_0$ | 0.052 | 0.0596 | 0.306 | 0.0873 | -0.0595 | 99.6 |
| (laptop) | aggreg. | 0.0129 | 0.0501 | 0.4730 | 0.1061 | -0.1285 | 99.8 |
| Reduced | $k_a$ | -0.000443 | 0.046 | 0.42 | 0.115 | -0.14 | 99.9 |
| ($N_P = 5$) | $k_e$ | -0.0123 | 0.0461 | 0.333 | 0.0799 | -0.218 | 97.9 |
| $< 2$ min | $V_0$ | 0.0604 | 0.076 | 0.207 | 0.081 | -0.068 | 87.4 |
| (laptop) | aggreg. | 0.0159 | 0.0560 | 0.3200 | 0.0920 | -0.1420 | 95.1 |

Table 3: Performances of the population estimation algorithms (full and reduced versions) evaluated over 100 replicates for $N_P$=100 and $\sigma_m = 0.3$. Summary statistics are the relative bias for parameters (RBIAS), the mean squared error (MSE), the mean estimated standard deviation (STD), the empirical standard deviation of estimates (ESTD), the standard deviation for random effects (BMIXED) and the coverage (COV).

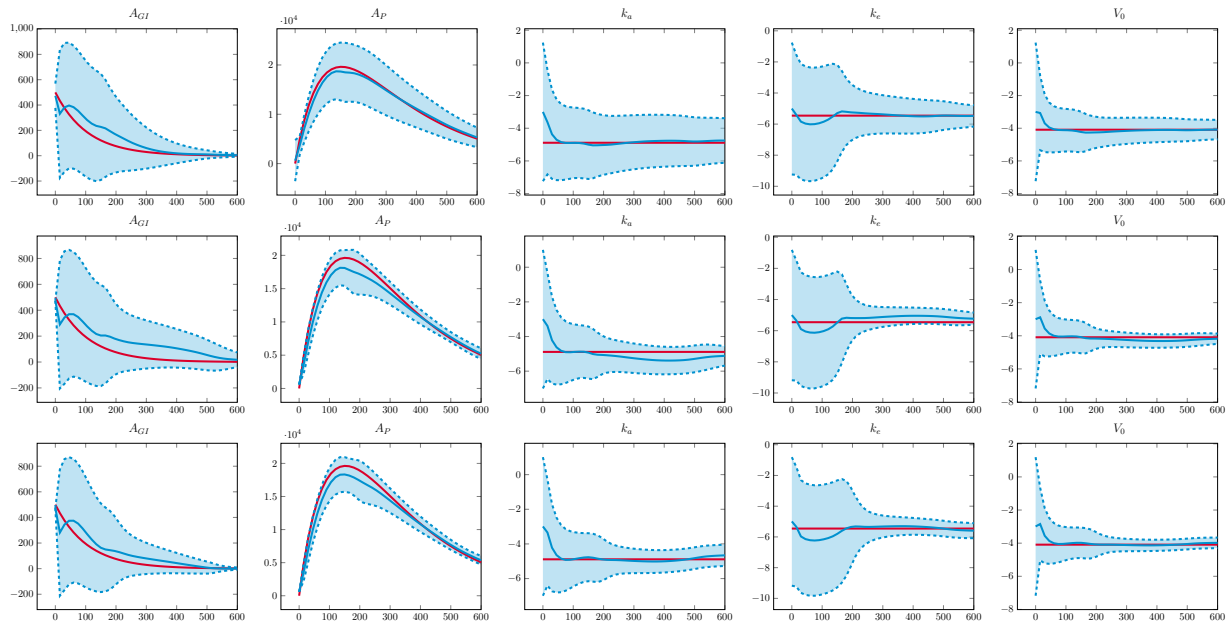



Figure 5: Time evolution of states $A_{GI}$ and $U_P$ and parameters $k_a, k_e$ and $V_0$ for one patient using the full version of the algorithm (top) and the reduced-order versions (middle: $N_C = 20$, bottom: $N_C = 5$) with $N_P$=100 and $\sigma_m = 0.3$. The target states and parameters are in red. The estimated states and parameters are in blue. The dashed lines correspond to the standard deviations.

Using the reduced-order filter, the COV gets closer to their nominal value of 95% and finally, it also improves the BMIXED. The improvements can be explained by the fact that the reduced-order version of the population Kalman filter increases the patient coupling. This also means that the number of clusters has to be chosen carefully in order to not to overly constrain the system. Figure 5 shows the time evolutions of the states and the parameters of one patient of one replicate obtained with the full population Kalman filter (top) and its reduced-order version with $N_C = 20$ (middle) and $N_C = 5$ (bottom). One can remark that the time evolutions have a similar behavior in the four cases. This means that the system is not too constrained even if the COV is low for the $V_0$ parameter (but still acceptable). Figure 5 shows that increasing the coupling by reducing the number of clusters reduces the standard deviations removing the need of multiple loops. Concerning the computational times, the reduced-order version of the population Kalman filter allows to drastically reduce it: a factor 7 for example between the full filter and the reduced-order version with $N_C = 5$.

## 4.3 Experiment with real data

To conclude this numerical section, a validation on a real data set is proposed. Theophylline is a methylxanthine drug used in therapy for respiratory diseases such as chronic obstructive pulmonary disease and asthma under a variety of brand names. The data were collected by Upton et al. in 12 subjects given a single oral dose of theophylline who then contributed 11 blood samples over a period of 25 hours [27]. The 12 subjects received $A_{GI}(0) = 320$mg of theophylline. A spaghetti plot representing the plasma concentration of the drug with time is available in Figure 1-(b). Considering the small number of patients, only the full population filter will be tested. For the sake of illustration, it is possible to compare our results with a simple nonlinear mixed effect (NLME) approach (based on the analytic solution given in Remark 2).

Individual nonlinear estimation is implemented using the package *nls* of the free software for statistical computing `R` and population NLME model is implemented using the package *saemix* [5]. We conclude that the population-based Kalman filter that we have proposed in this article gives similar results to the

| | Individual NL | | Population NLME | | Individual Kalman | | Population Kalman | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Mean | STD | Mean | STD | Mean | STD | Mean | STD |
| $k_a$ | 2.189 | 2.28 | 1.586 | 0.308 | 1.327 | 0.609 | 1.693 | 0.317 |
| $k_e$ | 0.089 | 0.01 | 0.088 | 0.006 | 0.076 | 0.008 | 0.074 | 0.006 |
| $V_0$ | 58.75 | 70.6 | 32.47 | 1.407 | 38.43 | 3.085 | 31.98 | 2.863 |
| $\sigma_{k_a}$ | - | - | 0.403 | 0.179 | - | - | 0.136 | 0.133 |
| $\sigma_{k_e}$ | - | - | 0.015 | 0.008 | - | - | 0.083 | 0.010 |
| $\sigma_{V0}$ | - | - | 0.022 | 0.055 | - | - | 0.047 | 0.032 |

Table 4: Comparison of four estimators for pharmacokinetics one-compartment model with first-order absorption and elimination on the Theophylline dataset: Individual nonlinear estimation and individual UKF, and population nonlinear estimation and population UKF. Estimates and estimated standard deviations are reported.

population NLME on real data, see Table 4. It is also interesting to note that the individual Kalman filter gives better results that the individual nonlinear strategy, in particular concerning the estimation of parameter $V_0$ which is the most difficult to estimate because it determines the slope of $U_p$ and its estimation is very sensitive to the observation noise. This shows that the individual Kalman filter can be efficient when the *a-priori* are better determined.

# 5 Concluding remarks and perspectives

We have proposed a new strategy to estimate, using available measurements, the state of a dynamical system and model parameters using a population approach. To compensate the weakness of the observations which are noisy and time-sampled, our population approach uses the fact that multiple independent patients are observed. We have defined a log-likelihood functional (where all disturbances are Gaussian) which couples all the patients and we have minimized it using the Unscented Kalman Filter [10]. To decrease the computational times, we have proposed a reduced-order version of our population filter based on an observation clustering. Note that our choice of UKF could be replaced by an Extended Kalman Filter [25] or an Ensemble Kalman Filter [9] as soon as reduced order version are available [19, 20].

We have then illustrated and assessed our proposed strategy on a simple pharmacokinetics model. Our approach has proven to be successful on synthetic data. Moreover on real data, our results are very comparable with a simple nonlinear mixed effect approach. We can now proceed to a more complete numerical evaluation on more complex systems, with potentially stronger nonlinearities.

As explained in the introduction, the objective of this work was to propose a strategy which stays accurate, tractable and in reasonable running time when the ODE system dimension increases. The obtained results are very encouraging and the computational time is very acceptable with less than 2 minutes for 100 patients on a laptop. This paves the way of using our strategy with large-dimensional models or possibly partial differential (PDE) systems.

# References


[1] M. Asch, M. Bocquet, and M. Nodet. *Data assimilation: methods, algorithms, and applications.* Fundamentals of Algorithms. SIAM, 2016.

[2] A. Bensoussan. *Estimation and Control of Dynamical Systems.* Interdisciplinary Applied Mathematics. Springer, 2018.

[3] J. Blum, F.-X. Le Dimet, and I.M. Navon. Data assimilation for geophysical fluids. In R. Temam and J. Tribbia, editors, *Handbook of Numerical Analysis: Computational Methods for the Atmosphere and the Oceans.* Elsevier, 2008.

[4] D. Chapelle, M. Fragu, V. Mallet, and P. Moireau. Fundamental principles of data assimilation underlying the Verdandi library: applications to biophysical model personalization within euHeart. *Medical & Biological Eng & Computing*, 51:1221–1233, 2013.

[5] E. Comets, A. Lavenu, and M. Lavielle. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software*, 2016.

[6] M. Delattre and M. Lavielle. Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and its interface*, 6(4):519–532, 2013.

[7] M. J. Denwood. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(9):1–25, 2016.

[8] S. B. Duffull, C. M. J. Kirkpatrick, B. Green, and N. H. G. Holford. Analysis of population pharmacokinetic data using NONMEM and WinBUGS. *Journal of Biopharmaceutical Statistics*, 15(1):53–73, 2004.

[9] G. Evensen. *Data Assimilation – The Ensemble Kalman Filter*. Springer Verlag, 2007.

[10] S.J. Julier and J.K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[11] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Trans. ASME J. Basic. Eng.*, 83:95—108, 1961.

[12] S. Klim, S. B. Mortensen, N. R. Kristensen, R. V. Overgaard, and H. Madsen. Population stochastic modelling (PSM)—an R package for mixed-effects models based on stochastic differential equations. *Computer methods and programs in biomedicine*, 94(3):279–289, 2009.

[13] E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038, 2005.

[14] X. Liu and Y. Wang. Comparing the performance of [FOCE] and different expectation-maximization methods in handling complex population physiologically-based pharmacokinetic models. *Journal of pharmacokinetics and pharmacodynamics*, 43(4):359–370, 2016.

[15] P. Moireau. A Discrete-time Optimal Filtering Approach for Non-linear Systems as a Stable Discretization of the Mortensen Observer. *ESAIM: Control, Optimisation and Calculus of Variations*, 2017.

[16] P. Moireau and D. Chapelle. Reduced-order unscented Kalman filtering with application to parameter identification in large-dimensional systems. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(2):380–405, 2011.

[17] R. V. Overgaard, N. Jonsson, C. W. Tornøe, and H. Madsen. Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *Journal of pharmacokinetics and pharmacodynamics*, 32(1):85–107, 2005.

[18] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271(5255):1582–1586, 1996.

[19] D. T. Pham, J. Verron, and C. M. Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine systems*, 16(3-4):323–340, 1998.

[20] D.T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207, 2001.

[21] J. C. Pinheiro and D. M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1):12–35, 1995.

[22] E. L. Plan, A. Maloney, F. Mentré, M. O. Karlsson, and J. Bertrand. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose–response models. *The AAPS journal*, 14(3):420–432, 2012.

[23] M. Prague. Use of dynamical models for treatment optimization in HIV infected patients: a sequential bayesian analysis approach. *Journal de la Société Française de Statistique*, 157(2):20, 2016.

[24] M. Prague, D. Commenges, J. Guedj, J. Drylewicz, and 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(2):447–458, 2013.

[25] D. Simon. *Optimal State Estimation: Kalman, $H^\infty$, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[26] C. W. Tornøe, R. V. Overgaard, H. Agersø, H. Nielsen, H. A .and Madsen, and E. N. Jonsson. Stochastic differential equations in nonmem®: implementation, application, and comparison with ordinary differential equations. *Pharmaceutical research*, 22(8):1247–1258, 2005.

[27] R. A. Upton. Pharmacokinetic interactions between theophylline and other medication (Part I). *Clinical pharmacokinetics*, 20(1):66–80, 1991.

[28] G. Verbeke. Linear mixed models for longitudinal data. In *Linear mixed models in practice*, pages 63–153. Springer, 1997.

[29] J. Wakefield and A. Racine-Poon. An application of bayesian population pharmacokinetic/pharmacodynamic models to dose recommendation. *Statistics in medicine*, 14(9):971–986, 1995.

[30] H. Wu. Statistical methods for HIV dynamic studies in aids clinical trials. *Statistical methods in medical research*, 14(2):171–192, 2005.