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 in computational biology

Annabelle Collin[1,2,3], Mélanie Prague[1,4,5], 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

Estimating dynamical systems — in particular identifying theirs parameters — involved in computational biology – for instance in pharmacology, in virology or in epidemiology – is fundamental to put in accordance the model trajectory with the measurements at hand. Unfortunately, when the sampling of data is very scarce or the data are corrupted by noise, parameters mean and variance priors must be chosen very adequately to balance our measurement distrust. Otherwise the identification procedure fails. A circumvention consists in using repeated measurements collected in configurations that share common priors – for instance with multiple population subjects in a clinical study or clusters in an epidemiology investigation. This common information is of benefit and is typically modeled in statistics by nonlinear mixed-effect models. In this paper, we introduce a data assimilation methodology compatible with such mixed-effect strategy without being strangled by the potential resulting curse of dimensionality. We define population-based estimators through maximum likelihood estimation. Then, from filtering theory, we set-up an equivalent robust large population sequential estimator that integrates the data as they are collected. Finally, we limit the computational complexity by defining a reduced-order version of this population Kalman filter clustering subpopulations of common observation background. The resulting algorithm performances are evaluated on classical pharmacokinetics benchmark. The versatility of the proposed method is finally fully challenged in a real-data epidemiology study of COVID spread in regions and departments of France.

1 Introduction

Mathematical models – as for example systems of ordinary differential equations (ODE) or stochastic differential equation (SDE) to account for additional uncertainties – can be used to understand the evolution of biological processes. For instance, in pharmacology, they can describe the drug pharmacokinetics (PK) in plasma [50], or in virology, the viral kinetics during HIV treatment [36], or in epidemiology the recent evolution of the COVID epidemics [27, 45].

When employing these models in practical applications, a great difficulty consists in dealing with the many uncertain quantities that must be prescribed for making the models really predictive. These quantities include model initial conditions or parameters, which can be difficult to measure directly. Fortunately, additional information is provided by collected data which can be used to circumvent the uncertainties associated with the dynamical system definition.

Estimation strategies for these models can be seen as deterministic inverse methods, in which we want to recover the individual parameters that have produced some observations. The resolution of this inverse problem then typically leads to solve an ODE constrained optimization problem. But it could also be understood as a statistical estimation procedure with a variety of techniques based on maximum likelihood optimization, for example but not restricted to Expectation-maximization type algorithms. In this landscape, either deterministic or stochastic sequential estimation such as Kalman based filters, which consist in correcting the original dynamics as the data are collected are particularly appealing when affordable in terms of computational complexity [17, 46, 4, 2, 28].

Unfortunately in many situations, the estimation procedure – independently of the chosen method – can lack of robustness when the measurement are too sparse or too noisy, without relying on strong model priors. A way to compensate the underlying observability weakness can be to use the fact that multiple independent measurements have been collected on configuration where the model variables describe a statistical unit of interest belonging to a larger population where the variability is constrained. This is typically the assumption made in nonlinear mixed-effect models [24, 49, 26] – where the use of available population data improve identifiability of individual parameters through a coupled likelihood function maximization. Many algorithms and related softwares have

been developed upon such principle. Among others, we can cite the First Order Conditional Estimation (FOCE) algorithm implemented in `NONMEM` [40], the Stochastic Approximation of Expectation-Maximization (SAEM) algorithm proposed in `MONOLIX` [23], penalized maximum likelihood approaches available in `NIMROD` [43] or even full Markov chain Monte Carlo (MCMC) methods implemented in `Winbugs` [52] or `jags` [13]. 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 [15, 41, 43, 30]. Increasing the size of the underlying dynamics without comprising too much the algorithm complexity cost remains, however, a challenge.

In this work, we propose an alternative strategy inspired from data assimilation method generally developed for environmental sciences [2] and often confronted with the complexity burden associated with large-dimensional systems. Indeed, our strategy is to gather all the population information in a unified maximum likelihood estimation procedure. Further assuming Gaussian disturbances up to a change of variable, we approximate the corresponding optimization problem using a Kalman-based filter. More precisely, we choose the Unscented Kalman Filter [20], but other Gaussian filters [2] could lead to similar results. The potential curse of dimensionality induced by the population framework is then limited by covariance reduction techniques as proposed in the Reduced-Order Unscented Kalman Filter [31]. Note that it is not the first time that Kalman-based approaches are used in mixed-effect strategies. This is the case in [34, 47, 22] where specific combination of the FOCE method with the extended Kalman filter has been suggested. Or in [11] where a coupling of the SAEM algorithm with the extended Kalman filter is performed. However in this literature, the extended Kalman filters are only used to approximate the individual probability distribution function, whereas in this article we also use the Kalman approach at the population level.

In order to evaluate our approach, we propose a collection of numerical tests starting from a classical toy problem in pharmacokinetics. A classical synthetic data benchmark is first considered to validate our implementation. Then, a validation on a real data set is performed to evaluate the algorithm performances when compared to the literature. We have found that the COVID pandemic presents a very general case of application when trying to aggregate with SIR-based model [51] multiple data collection from various space locations. More precisely, we assume that the data recorded in France at the regional (or departmental) level are independent realizations of the same underlying epidemics of COVID-19 driven by random parameters specific to each region (or department), requiring a *population approach* for estimation of parameters. This numerical experiment really challenges our algorithm capabilities hence illustrates well its versatility.

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, the numerical section provides assessment of the performances before a final conclusion in Section 5.

2 Problem statement

2.1 Model formulation

Let us consider a population of N_P subjects. For each subject, we denote by $x^i \in \mathcal{X} \simeq \mathbb{R}^{N_x}$, $1 \leq i \leq N_P$, the system *state – i.e.* a vector which contains all the time-dependent variables of the system – and by θ^i, $1 \leq i \leq N_P$ the vector containing all the time-independent *parameters* of the system. The stochastic dynamics over the time window $[0, T]$ is modeled by a vector field $f \in C^1(\mathcal{X})$ representing a deterministic law of evolution such that for each subject $1 \leq i \leq N_P$,

$$\mathrm{d}x^i(t) = f(x^i(t), \theta^i, t)\mathrm{d}t + G(t)\mathrm{d}\omega^i(t), \quad t \in [0, T],$$

where ω^i is an additional Wiener process in $\mathcal{Q} \simeq \mathbb{R}^{N_\nu}$ and G a linear operator responsible for the stochastic contribution. Note that in a very general setting G could also depend of the state and parameters. Here, we restrict our presentation to more simple cases with $G(t)$ only. The parameters of interest $\theta^i \in \mathcal{P} \simeq \mathbb{R}^{N_\theta}$ 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

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

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

$$\begin{cases} \mathrm{d}z^i(t) = a(z^i(t), t)\mathrm{d}t + B(t)\mathrm{d}\omega^i(t), \quad \forall t \in [0, T] \\ z^i(0) = z_0^i = \begin{pmatrix} x^i(0) \\ \theta^i \end{pmatrix}, \end{cases} \tag{1}$$

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

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

Therefore, (1) is a general dynamics combining time-dependent and time-independent variables.

2.2 Uncertainties modeling

First, concerning the model stochasticity, we introduce the Wiener processes covariances Q^i, $1 \leq i \leq N_P$ such that as in [4, Chapter 7]

$$\mathbb{E}\big(\omega^i(t)\omega^{i\intercal}(s)\big) = \int_0^{\min(t,s)} Q^i(\tau)d\tau.$$

In addition to the stochasticity introduced in the dynamics, we consider that the initial state and the parameters of interest are subject to some stochasticity in the population considered. We have

$$z_0^i = z_0 + \xi^i$$

where z_0 is a deterministic known quantity and $\xi^i, 1 \leq i \leq N_P$, are random variables. Typically, in a population approach, the N_P members of the population – called subjects in this work – are assumed to be randomly sampled from a same distribution, here a Gaussian distribution

$$\xi^i \sim_{\text{i.i.d.}} \mathcal{N}(\xi^{\text{pop}}, P_0), \quad 1 \leq i \leq N_P,$$

where ξ^{pop} is the population mean value and P_0 is the associated covariance. We equivalently write

$$\xi^i = \xi^{\text{pop}} + \tilde{\xi}^i \text{ width } \tilde{\xi}^i \sim_{\text{i.i.d.}} \mathcal{N}(0, P_0), \quad 1 \leq i \leq N_P.$$

In mixed-effect models ξ^{pop} is called the population intercept whereas $\tilde{\xi}^i$ is the random effect. Typically ξ^{pop} is not known, neither P_0. However, we may assume that ξ^{pop} is bounded in the nondimensionalized norm defined by a given positive matrix M, namely

$$\langle \xi^{\text{pop}}, M\xi^{\text{pop}} \rangle = O(1).$$

Moreover, we have at our disposal some a priori $\hat{P}_0$ for P_0.

2.3 Observations

Our objective consists in estimating the unknown parts of the initial conditions in our population using the available data. In practice, we do not directly observe x^i but we have at our disposal time-sampled observations – or measurements – $y^i \in \mathcal{Y} \simeq \mathbb{R}^{N_{\text{obs}}}$. These observations y^i are related to x^i through the measurement procedure. This process is modeled by introducing an observation operator h_k, dependent of the time $t_k \in [0, T]$, so that

$$y_k^i = h_k^i\big(x^i(t_k^i)\big) + \chi_k^i, \quad 1 \leq i \leq N_P,\ 0 \leq k \leq N_{T_{\text{obs}}},$$

where χ_k^i represents an additive measurement error, typically

$$\chi_k^i \sim_{\text{i.i.d.}} \mathcal{N}(0, W_k^i), \quad 1 \leq i \leq N_P,\ 0 \leq k \leq N_{T_{\text{obs}}},$$

with W_k^i the observation error covariance matrix. Note that in order to define a discrete-time white noise that could be consistent with a potential continuous-time limit, we have the following scaling $\|W_k^i\| \propto \frac{1}{\Delta t_k^i}$ with $\Delta t_k^i = t_{k+1}^i - t_k^i$.

For the sake of simplicity when introducing the method formalism, the number of observation times $N_{T_{\text{obs}}} + 1$ will not be subject-dependent. However in the results section, we will explain how to practically consider subject-dependent measurement sampling.

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

$$y_k^i = c_k^i\big(z^i(t_k^i)\big) + \chi_k^i, \quad 1 \leq i \leq N_P,\ 0 \leq k \leq N_{T_{\text{obs}}},$$

where

$$c_k^i : \mathcal{X} \times \mathcal{P} \ni z = (x, \theta) \mapsto h_k^i(x) \in \mathcal{Y}.$$

2.4 Dynamics discretization

In practice, we will ultimately rely on a discretization of the dynamics (1). We here restrict our presentation to a forward-Euler time scheme, but it is rather straightforward to consider more general time-scheme. Choosing a time-step $\delta t_n = t_{n+1} - t_n$ small enough to satisfy the conditional stability condition, we should numerically solve

$$x_{n+1}^i = x_n^i + \delta t_n f(x_n^i, \theta^i, T_n) + G(T_n)\nu_n^i, \quad 0 \leq n \leq N_T - 1, \quad 1 \leq i \leq N_P, \tag{2}$$

with $T_n = \sum_{i=0}^n \delta t_i$, $T_{N_T} = T$ and the $(\nu_n^i)_{0 \leq n \leq N_T - 1} \sim_{\text{i.i.d}} \mathcal{N}(0, \delta t_n Q^i(t_n))$ are increment of the Wiener process during the interval δt_n. In the rest of the paper focusing on time-discrete formulation, we denote $Q_n^i = \delta t_n Q^i(t_n)$. From (2), there exists a discrete transition operator $\phi_{n+1|n}$ giving the state x_{n+1}^i knowing the state x_n^i, namely

$$x_{n+1}^i = \phi_{n+1|n}(x_n^i, \theta^i) + G_n \nu_n^i, \quad 0 \leq n < N_T - 1, \quad 1 \leq i \leq N_P, \tag{3}$$

with here $\phi_{n+1|n}(x_n^i, \theta^i) = (x_n^i + \delta t_n f(x_n^i, \theta^i, T_n))$ and $G_n = G(T_n)$. Note that more complex time-schemes will lead to other definitions of the transition operator, however (3) is a rather general time-discrete dynamics. As presented before, we define from $\phi_{n+1|n}$ a joint state-parameter transition operator $\psi_{n+1|n}$ and a joint operator B_n such that

$$\begin{cases} z_{n+1}^i = \psi_{n+1|n}(z_n^i) + B_n \nu_n^i, \quad 0 \leq n < N_T - 1, \quad 1 \leq i \leq N_P. \\ z_0^i = z_0 + \xi^i \end{cases} \tag{4}$$

We point-out that to simplify the presentation the model discretization and the observation discretization are consistant, namely $\max_i(\delta t_i)$ is small enough so that $N_T \geq N_{\text{obs}}$ and there exists a sequence of intergers $(n_k)_{0 \leq k \leq N_{\text{obs}}}$ such that for all observation time t_k, $1 \leq k \leq N_{\text{obs}}$, there exists $T_{n_k} = t_k$.

Note that when combining the augmented dynamics and the observations, we find that there exist non-linear mapping Γ such that

$$y_k^i = \Gamma(\xi^{\text{pop}}, \xi^i, (\nu_n^i)_{0\leq n<N_T-1},\, t_k) + \chi_k^i, \quad 1 \leq i \leq N_P, \quad 0 \leq k \leq N_{T_{\text{obs}}} \tag{5}$$

where χ_k^i is Gaussian noise, ξ^{pop} is a the fixed effect and ξ^i and $(\nu_n^i)_{0\leq n\leq N_T-1}$ are random effects as in classical mixed-effect models [26]. Moreover, even if the model dynamics is linear and the observation operator is linear, the dependency of the dynamics with respect to the parameters ultimately induces that Γ is a non-linear operator with respect to z_0^{pop} and ξ^i.

To summarize the problem statement, our objective is therefore to estimate $\xi^{\text{pop}}, \xi^i, (\nu_n^i)_{0\leq n\leq N_T-1}$ and $\hat{P}_0$ from $(y_k^i)_{1\leq i\leq N_P, 0\leq k\leq N_{T_{\text{obs}}}}$ given the relation (5) while benefiting – with a data assimilation point of view – from the dynamical structure behind Γ.

3 Methods

3.1 A population-based likelihood objective function for Gaussian disturbances

Let us first aggregate all the subject uncertainties in a vector of dimension of $N_z \times N_P$

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

We similarly denote $(\boldsymbol{\nu}_n)_{0\leq n\leq N_T-1} = ((\nu_n^1 \cdots \nu_n^{N_P})^\intercal)_{0\leq n\leq N_T-1}$. More generally in the rest of the presentation, bold notation will indicate quantities gathering all the subjects together.

In a classical Maximum Likelihood Estimation (MLE) strategy with Gaussian disturbances, the MLE should be obtained by minimizing a negative log-likelihood of the form

$$\min_{\substack{\xi^{\text{pop}}\in\mathcal{X}, \boldsymbol{\xi}\in(\mathcal{X}\times\mathcal{P})^{N_P}, \\ (\boldsymbol{\nu}_n)_{0\leq n\leq N_T-1}\in\mathcal{Q}^{N_P\times N_T}}} \Bigg\{ \mathscr{L}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\langle \xi^{\text{pop}}, M\xi^{\text{pop}}\rangle$$
$$+ \sum_{i=1}^{N_P}\Bigg[\frac{1}{2}\big\langle (\xi^i-\xi^{\text{pop}}), \hat{P}_0^{-1}\,(\xi^i-\xi^{\text{pop}})\big\rangle + \frac{1}{2}\sum_{n=0}^{N_T-1}\big\langle \nu_n^i, (Q_n^i)^{-1}\nu_n^i\big\rangle$$
$$+ \frac{1}{2}\sum_{k=0}^{N_{T_{\text{obs}}}}\big\langle (y_k^i - c_k^i(z_{n_k}^i)), (W_k^i)^{-1}(y_k^i - c_k^i(z_{n_k}^i))\big\rangle\Bigg]\Bigg\}, \tag{6}$$

where all $(z_n^i)_{0\leq n\leq N_T}$, $1\leq i\leq N_P$, are constrained to follow the dynamics (4). Here, this minimization includes the population intercept ξ^{pop}. To simplify the minimization problem, we remark that the population intercept represents the expected mean of all the subject uncertainties. We thus propose to model ξ^{pop} as an empirical mean over all the subjects namely

$$\xi^{\text{pop}} = \mathbb{E}_{N_P}(\boldsymbol{\xi}) \overset{\text{def}}{=} \frac{1}{N_P}\sum_{i=1}^{N_P}\xi^i.$$

Therefore, the minimization problem becomes

$$\min_{\substack{\boldsymbol{\xi}\in(\mathcal{X}\times\mathcal{P})^{N_P}, \\ (\boldsymbol{\nu}_n)_{0\leq n\leq N_T-1}\in\mathcal{Q}^{N_P\times N_T}}} \Bigg\{ \mathscr{L}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\big\langle \mathbb{E}_{N_P}(\boldsymbol{\xi}), M\mathbb{E}_{N_P}(\boldsymbol{\xi})\big\rangle$$
$$+ \sum_{i=1}^{N_P}\Bigg[\frac{1}{2}\big\langle (\xi^i-\mathbb{E}_{N_P}(\boldsymbol{\xi})), \hat{P}_0^{-1}\,(\xi^i-\mathbb{E}_{N_P}(\boldsymbol{\xi}))\big\rangle + \frac{1}{2}\sum_{n=0}^{N_T-1}\big\langle \nu_n^i, (Q_n^i)^{-1}\nu_n^i\big\rangle$$
$$+ \frac{1}{2}\sum_{k=0}^{N_{T_{\text{obs}}}}\big\langle (y_k^i - c_k^i(z_{n_k}^i)), (W_k^i)^{-1}(y_k^i - c_k^i(z_{n_k}^i))\big\rangle\Bigg]\Bigg\}, \tag{7}$$

where the unknowns of our population optimization procedure reduce to ξ^i, $1\leq i\leq N_P$ and ν_n^i, $1\leq i\leq N_P, 0\leq n\leq N_T-1$. In fact, the functional in (7) depends on fewer unknowns than in (6), 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 objective function can be rearranged by developing the first two terms into

$$\big\langle \mathbb{E}_{N_P}(\boldsymbol{\xi}), M\,\mathbb{E}_{N_P}(\boldsymbol{\xi})\big\rangle + \sum_{i=1}^{N_P}\big\langle (\xi^i-\mathbb{E}_{N_P}(\boldsymbol{\xi})), \hat{P}_0^{-1}\,(\xi^i-\mathbb{E}_{N_P}(\boldsymbol{\xi}))\big\rangle = \sum_{i,j=1}^{N_P}\Big\langle \xi^i, \Big[\frac{1}{{N_P}^2}M + \Big(\delta_{ij}-\frac{1}{N_P}\Big)(\hat{P}_0)^{-1}\Big]\xi^j\Big\rangle.$$

We then introduce the matrix $\boldsymbol{M}_0$ of dimension $(N_P \times N_z)^2$ defined by blocks such that for $1 \leq i, j \leq N_P$, the block $(\boldsymbol{M}_0)_{i,j}$ of indexes in $[(i-1)\,N_z + 1, i\,N_z] \times \left[(j-1)\,N_z + 1, j\,N_z\right]$ is given by

$$(\boldsymbol{M}_0)_{i,j} = \frac{1}{{N_P}^2} M + \Big(\delta_{ij} - \frac{1}{N_P}\Big)\hat{P}_0^{-1}. \tag{8}$$

Using Kronechker product $\otimes$, $\vec{1}_{N_P} = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}^\intercal \in \mathbb{R}^{N_P}$ and $\mathbb{I}_{N_P}$ denoting the identity matrix in $\mathbb{R}^{N_P}$, the matrix $\boldsymbol{M}_0$ can be rewritten as

$$\boldsymbol{M}_0 = \frac{1}{{N_P}^2} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal \otimes M + \left[\mathbb{I}_{N_P} - \frac{1}{N_P} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal\right] \otimes \hat{P}_0^{-1}.$$

Both matrices $\vec{1}_{N_P} \vec{1}_{N_P}^\intercal \otimes M$ and $(\mathbb{I}_{N_P} - \frac{1}{N_P} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal) \otimes \hat{P}_0^{-1}$ are positive semidefinite. Moreover the second one, $(\mathbb{I}_{N_P} - \frac{1}{N_P} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal) \otimes \hat{P}_0^{-1}$, is definite on $\mathrm{Ker}(\vec{1}_{N_P} \vec{1}_{N_P}^\intercal)^\perp \otimes \mathbb{R}^{N_x}$, whereas the first one,$\vec{1}_{N_P} \vec{1}_{N_P}^\intercal \otimes M$, is definite on $\mathrm{Ker}(\vec{1}_{N_P} \vec{1}_{N_P}^\intercal) \otimes \mathbb{R}^{N_x}$. This enforces the invertibility of the matrix $\boldsymbol{M}_0$. Note that the matrix ${N_P}^{-2} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal \otimes M$ can be small – in the sense of its Frobenius norm $\mathrm{Tr}(M)$ – hence interpretable as a regularization term with respect to the second matrix. In fact, when $\mathrm{Tr}(M)$ is small with respect to $Tr(\hat{P}_0^{-1})$, our criterion imposes that the estimation is strongly supported with a shared population variable ξ^{pop} and small variations $\tilde{\xi}^i$ around it. Conversely, $\mathrm{Tr}(M)$ large with respect to $\mathrm{Tr}(\hat{P}_0^{-1})$ in a large population $N_p \gg 1$ leads to an estimation where all members of the population are decoupled.

Ultimately introducing

$$\hat{\boldsymbol{P}}_0 = \boldsymbol{M}_0^{-1} = \left(\frac{1}{{N_P}^2} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal \otimes M + \left[\mathbb{I}_{N_P} - \frac{1}{N_P} \vec{1}_{N_P} \vec{1}_{N_P}^\intercal\right] \otimes \hat{P}_0^{-1}\right)^{-1}, \tag{9}$$

our estimator is defined as the minimizer of the least-square objective function (7) rewritten in the general form

$$\mathscr{L}_T(\boldsymbol{\xi}, \boldsymbol{\nu}) = \frac{1}{2}\big\langle \boldsymbol{\xi}, \hat{\boldsymbol{P}}_0^{-1}\, \boldsymbol{\xi}\big\rangle + \frac{1}{2} \sum_{n=0}^{N_T - 1} \big\langle \boldsymbol{\nu}_n, \boldsymbol{Q}_n^{-1} \boldsymbol{\nu}_n\big\rangle + \frac{1}{2} \sum_{k=0}^{N_{T_{\mathrm{obs}}}} \big\langle \boldsymbol{y}_k - \boldsymbol{c}_k(\boldsymbol{z}_k), \boldsymbol{W}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{c}_k(\boldsymbol{z}_{n_k}))\big\rangle. \tag{10}$$

We have gathered all the individual states in $\boldsymbol{z} = (z^1 \cdots z^{N_P})^\intercal$ constrained to follow the time-discrete "population" dynamics

$$\begin{cases} \boldsymbol{z}_{n+1} = \boldsymbol{\psi}_{n+1|n}(\boldsymbol{z}_n) + \boldsymbol{B}_n \boldsymbol{\nu}_n, \quad \forall t \in [0, T] \\ \boldsymbol{z}_0 = \vec{1}_{N_P} \otimes z_0 + \boldsymbol{\xi}, \end{cases} \tag{11}$$

with

$$\boldsymbol{\psi}_{n+1|n}(\boldsymbol{z}_n) = \begin{pmatrix} \psi_{n+1|n}(z_n^1) \\ \vdots \\ \psi_{n+1|n}(z_n^{N_P}) \end{pmatrix} \text{ and } \boldsymbol{B}_n = \mathbb{I}_{N_P} \otimes B_n,$$

and measurements are concatenated into $\boldsymbol{y} = (y^1 \cdots y^{N_P})^\intercal$ so that

$$\boldsymbol{y}_k = \boldsymbol{c}_k(\boldsymbol{z}_{n_k}) + \boldsymbol{\chi}_k, \quad 1 \leq k \leq N_{T_{\mathrm{obs}}},$$

with an observation operator

$$\boldsymbol{c}_k(\boldsymbol{z}_{n_k}) = \begin{pmatrix} c_k^1(z_{n_k}^1) \\ \vdots \\ c_k^{N_P}(z_{n_k}^{N_P}) \end{pmatrix},$$

and a measurement noise $\boldsymbol{\chi}_k = (\boldsymbol{\chi}_k^1 \cdots \boldsymbol{\chi}_k^{N_P})^\intercal$ of covariance

$$\boldsymbol{W}_k = \begin{pmatrix} W_k^1 & & 0 \\ & \ddots & \\ 0 & & W_k^{N_P} \end{pmatrix}.$$

Minimizing (10) under the constraint imposed by the dynamics (11) becomes a classical data assimilation problem [2, 28]. The key of our uncertainty modeling is that $\hat{\boldsymbol{P}}_0$ – built from (8) – couples the population members since, indeed,

$$\hat{\boldsymbol{P}}_0 \neq \hat{\boldsymbol{P}}_0^{\mathrm{un}} \stackrel{\mathrm{def}}{=} \mathbb{I}_{N_P} \otimes \hat{P}_0. \tag{12}$$

In the upcoming numerical section, the impact of considering $\hat{\boldsymbol{P}}_0$ rather than $\hat{\boldsymbol{P}}_0^{\mathrm{un}}$ will be illustrated by comparing our population estimation strategy with a completely decoupled approach where each subject dynamics is estimated separately.

3.2 Corresponding sequential estimator

Minimizing the functional (10) under the constraint of a dynamics of the form (11) can be naturally performed using variational data assimilation methods [5, 8, 2]. We denote by $\bar{\boldsymbol{z}}_{n|N_T}$, $0 \leq n \leq N_T$, the trajectory associated with the minimization of $\mathscr{L}_T$. 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"– namely for all n, $\boldsymbol{\psi}_{n+1|n}$, $\boldsymbol{c}_n$ and $\boldsymbol{B}_n$ being linear [21, 46, 4, 28]. With a nonlinear model operator or a nonlinear observation operator, it is classical to rely on approximate optimal sequential estimator based on the generalization of the Kalman filter to nonlinear operators [46, 28]. Here, we have found convenient to rely on the Unscented-Kalman Filter (UKF) where the mean and covariance operators are computed from empirical mean and empirical covariance based on N_σ sampling points – the so-called sigma points [20]. We define, for a given set of positive coefficient $\boldsymbol{\alpha} = (\alpha_j)_{1\leq j\leq N_\sigma}$ and a given set of particles $(v^{(j)})_{1\leq j\leq N_\sigma}$, the empirical mean and covariance by

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

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

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

where δ is a scaling factor. The UKF algorithm reads

$$\begin{cases}
\text{Initialization:} & \\
\hat{\boldsymbol{z}}_0^- = \vec{1}_{N_P} \otimes z_0, \quad \hat{\boldsymbol{P}}_0^- = \hat{\boldsymbol{P}}_0, & \\
\text{Sampling:} & \\
\hat{\boldsymbol{z}}_n^{(j)-} = \hat{\boldsymbol{z}}_n^- + \sqrt{\delta\hat{\boldsymbol{P}}_n^-}\boldsymbol{e}^{(j)}, & 0\leq n\leq N_T,\ 1\leq j\leq N_\sigma, \\
\text{Correction, when } T_n \in \{t_0,\cdots,t_{N_{\text{obs}}}\}: & \\
\tilde{\boldsymbol{y}}_n^{(j)} = \boldsymbol{c}_n\big(\hat{\boldsymbol{z}}_n^{(j)-}\big) - \boldsymbol{y}_n, & 0\leq n\leq N_T,\ 1\leq j\leq N_\sigma, \\
\boldsymbol{K}_n = \mathbb{C}\text{ov}_{\boldsymbol{\alpha}}\big(\hat{\boldsymbol{z}}_n^{(\cdot)-}, \tilde{\boldsymbol{y}}_n^{(\cdot)}\big)[\mathbb{C}\text{ov}_{\boldsymbol{\alpha},N_\sigma}\big(\tilde{\boldsymbol{y}}_n^{(\cdot)}\big) + \boldsymbol{W}_n]^{-1}, & 0\leq n\leq N_T, \\
\hat{\boldsymbol{z}}_n^+ = \hat{\boldsymbol{z}}_n^- + \boldsymbol{K}_n\big(\boldsymbol{y}_n - \boldsymbol{c}_n(\hat{\boldsymbol{z}}_n^-)\big), & 0\leq n\leq N_T, \\
\hat{\boldsymbol{P}}_n^+ = \hat{\boldsymbol{P}}_n^- - \boldsymbol{K}_n[\mathbb{C}\text{ov}_{\boldsymbol{\alpha},N_\sigma}\big(\tilde{\boldsymbol{y}}_n^{(\cdot)}\big) + \boldsymbol{W}_n]\boldsymbol{K}_n^\intercal, & 0\leq n\leq N_T, \\
\text{Sampling:} & \\
\hat{\boldsymbol{z}}_n^{(j)+} = \hat{\boldsymbol{z}}_n^+ + \sqrt{\delta\hat{\boldsymbol{P}}_n^+}\boldsymbol{e}^{(j)}, & 0\leq n\leq N_T-1,\ 1\leq j\leq N_\sigma, \\
\text{Prediction:} & \\
\hat{\boldsymbol{z}}_{n+1}^{(j)-} = \boldsymbol{\psi}_{n+1|n}\big(\hat{\boldsymbol{z}}_n^{(j)+}\big), & 0\leq n\leq N_T-1,\ 1\leq j\leq N_\sigma, \\
\hat{\boldsymbol{z}}_{n+1}^- = \mathbb{E}_{\boldsymbol{\alpha},N_\sigma}\big(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\big), & 0\leq n\leq N_T-1, \\
\hat{\boldsymbol{P}}_{n+1}^- = \mathbb{C}\text{ov}_{\boldsymbol{\alpha},N_\sigma}\big(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\big) + \boldsymbol{B}_n\boldsymbol{Q}_n\boldsymbol{B}_n^\intercal, & 0\leq n\leq N_T-1,
\end{cases} \tag{13}$$

and reduces to the classical Kalman Filter when all operators – $\boldsymbol{\psi}_{n+1|n}$, $\boldsymbol{c}_n$ and $\boldsymbol{B}_n$ – are linear. Note here that each sigma point is a combination of individual contributions in the population.

We now recall – see [17, 46, 4] for in-depth presentations – the resulting computed state vector $\hat{\boldsymbol{z}}_n^+$ approximates the conditional expectation

$$\hat{\boldsymbol{z}}_n^+ \simeq \mathbb{E}(\boldsymbol{z}_n|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n),$$

and the positive definite matrix $\hat{\boldsymbol{P}}_n^+$ is the approximation of the corresponding estimation error covariance

$$\hat{\boldsymbol{P}}_n^+ \simeq \mathbb{C}\text{ov}(\boldsymbol{z}_n - \hat{\boldsymbol{z}}_n^+|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n).$$

As a consequence concerning the estimated parameters, we can extract from each member of the population a covariance of the parameter estimation error

$$(\hat{\boldsymbol{P}}_n^+)_{i,i}^\theta \simeq \mathbb{C}\text{ov}(\theta_n^i - \hat{\theta}_n^{i+}|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n).$$

where again the index i selects the block of indexes in $[(i-1)\,N_z+1, i\,N_z]\times[(i-1)\,N_z+1, i\,N_z]$. At final time, averaging over the members of the population and taking the square root ${N_P}^{-1}\sum_{i=1}^{N_P}(\hat{\boldsymbol{P}}_{N_T}^+)_{i,i}^\theta$ gives an estimation of P_0 knowing the available data. Then, taking the square root of the diagonal entries of this matrix is an estimation the Standard Deviation of the Random Effects for each estimated parameter.

Additionally, as $\xi_n^{\text{pop}} = {N_P}^{-1}\sum_{i=1}^{N_P}\xi_n^i$, we can say that $({N_P}^{-2}\sum_{i,j=1}^{N_P}(\hat{\boldsymbol{P}}_{N_T}^+)_{i,j}^\theta)^{-1}$ can be seen as an update of M^θ, the parameter part of the introduced norm M to be used for a subsequent estimation procedure.

Remark 1 (Initial condition estimation through smoothing)**.** *One classical drawback of Kalman estimation is that a quantity of the form* $\mathbb{E}(\boldsymbol{z}_0|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n)$ *is not directly available. When considering the parameters, we circumvent this limitation by reusing the dynamics which imposes that* $\theta_n^i=\theta_0^i$*. However, this is not the case of* $\mathbb{E}(x_0^i|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n)$ *as the dynamics is not necessarily solvable backward in time. However, such quantities can be computed by considering a Kalman smoothing instead of the Kalman filter [46]. In practice, this can be achieved by augmenting the subject state* $z^i=(x^i\ \theta^i)^\intercal$ *with the quantity* $x_0^i=x^i(t=0)$ *which is, by construction, a time-independent variable. Therefore, we now have for all* $1\leq n\leq N_T$*,* $z_n^i=(x_n^i\ \theta^i\ x_0^i)^\intercal$*. The corresponding initial covariance should be modified accordingly from* $\hat{P}_0=\left(\begin{smallmatrix}\hat{P}_0^{xx}\hat{P}_0^{x\theta}\\ \hat{P}_0^{\theta x}\hat{P}_0^{\theta\theta}\end{smallmatrix}\right)$ *into*

$$\hat{P}_0=\begin{pmatrix}\hat{P}_0^{xx} & \hat{P}_0^{x\theta} & \hat{P}_0^{x\theta}\\ \hat{P}_0^{\theta x} & \hat{P}_0^{\theta\theta} & \hat{P}_0^{\theta x}\\ \hat{P}_0^{xx} & \hat{P}_0^{x\theta} & \hat{P}_0^{xx}\end{pmatrix},$$

where two lines are identical since the two variables are fully correlated at initial time. During the UKF algorithm, we can extract this time from the last diagonal bloc of $\hat{\boldsymbol{P}}_n^+$*, an estimation of* $\mathbb{E}(x_0^i|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n)$*.*

Remark 2 (Fading memory)**.** *It is common to rely on a so-called fading memory effect when using a sequential estimator in order to give greater emphasis to more recent data and, by contrast, limiting the risk of being struggle by the history prediction [46, Section 7.4]. In this respect, we introduce a forgetting factor* $0\leq\rho\leq 1$*, and modify the covariance prediction into*

$$\hat{\boldsymbol{P}}_{n+1}^-=\frac{1}{\rho}\mathbb{C}ov_{\boldsymbol{\alpha},N_\sigma}\left(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\right)+\boldsymbol{B}_n\boldsymbol{Q}_n\boldsymbol{B}_n^\intercal,\ 0\leq n\leq N_T. \tag{14}$$

In fact, such modification resonates as the result of changing the functional into

$$\mathscr{L}_T(\boldsymbol{\xi},\boldsymbol{\nu})=\frac{1}{2}\left\langle\boldsymbol{\xi},(\hat{\boldsymbol{P}}_0)^{-1}\,\boldsymbol{\xi}\right\rangle+\frac{1}{2}\sum_{n=0}^{N_T-1}\rho^{-2n}\left\langle\boldsymbol{\nu}_n,\boldsymbol{Q}_n^{-1}\boldsymbol{\nu}_n\right\rangle+\frac{1}{2}\sum_{k=0}^{N_{T_{obs}}}\rho^{-2n_k}\left\langle\boldsymbol{y}_k-\boldsymbol{c}_k(\boldsymbol{z}_{n_k},(\boldsymbol{W}_k)^{-1}(\boldsymbol{y}_k-\boldsymbol{c}_k(\boldsymbol{z}_{n_k}))\right\rangle.$$

Remark 3 (Non-linear transformation of state variables)**.** *In many cases, the state variable* x_n^i *which is modeled for each subject* i *at iteration* n *does not follow a Gaussian distribution. However, they can be close to Gaussian distribution after a non-linear transformation as classically envision in non-linear mixed effect models [33, 26]. This will be the case when considering non-negative values that can be transformed using a log-transform or bounded values that can be transformed using a logit transform. Then, it is not difficult to extend the UKF filter to manage variables which follow Gaussian distribution up to a given transformation. Introducing the general mapping* $\boldsymbol{\varphi}$ *that transform the augmented state into Gaussian variables, we modify the classical UKF algorithm into*

$$\begin{cases}
\textit{Initialization:}\\
\hat{\boldsymbol{z}}_0^-=\vec{1}_{N_P}\otimes z_0,\quad \hat{\boldsymbol{P}}_0^-=\hat{\boldsymbol{P}}_0,\\
\textit{Sampling:}\\
\boldsymbol{\zeta}_n^{(j)-}=\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_n^-)+\sqrt{\delta\hat{\boldsymbol{P}}_n^-}\boldsymbol{e}^{(j)}, & 0\leq n\leq N_T, 1\leq j\leq N_\sigma,\\
\textit{Correction, when } T_n\in\{t_0,\cdots,t_{N_{obs}}\}\textit{:}\\
\tilde{\boldsymbol{y}}_n^{(j)}=\boldsymbol{c}_n\left(\boldsymbol{\varphi}^{-1}(\boldsymbol{\zeta}_n^{(j)-})\right)-\boldsymbol{y}_n, & 0\leq n\leq N_T, 1\leq j\leq N_\sigma,\\
\boldsymbol{K}_n=\mathbb{C}ov_{\boldsymbol{\alpha}}\left(\boldsymbol{\zeta}_n^{(\cdot)-},\tilde{\boldsymbol{y}}_n^{(\cdot)}\right)[\mathbb{C}ov_{\boldsymbol{\alpha},N_\sigma}\left(\tilde{\boldsymbol{y}}_n^{(\cdot)}\right)+\boldsymbol{W}_n]^{-1}, & 0\leq n\leq N_T,\\
\hat{\boldsymbol{z}}_n^+=\boldsymbol{\varphi}^{-1}\Big(\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_n^-)+\boldsymbol{K}_n\left(\boldsymbol{y}_n-\boldsymbol{c}_n(\hat{\boldsymbol{z}}_n^-)\right)\Big), & 0\leq n\leq N_T,\\
\hat{\boldsymbol{P}}_n^+=\hat{\boldsymbol{P}}_n^--\boldsymbol{K}_n[\mathbb{C}ov_{\boldsymbol{\alpha},N_\sigma}\left(\tilde{\boldsymbol{y}}_n^{(\cdot)}\right)+\boldsymbol{W}_n]\boldsymbol{K}_n^\intercal, & 0\leq n\leq N_T,\\
\textit{Sampling:}\\
\boldsymbol{\zeta}_n^{(j)+}=\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_n^+)+\sqrt{\delta\hat{\boldsymbol{P}}_n^+}\boldsymbol{e}^{(j)}, & 0\leq n\leq N_T-1, 1\leq j\leq N_\sigma,\\
\textit{Prediction:}\\
\hat{\boldsymbol{z}}_{n+1}^{(j)-}=\boldsymbol{\psi}_{n+1|n}\left(\boldsymbol{\varphi}^{-1}(\boldsymbol{\zeta}_n^{(j)+})\right), & 0\leq n\leq N_T-1, 1\leq j\leq N_\sigma,\\
\hat{\boldsymbol{z}}_{n+1}^-=\boldsymbol{\varphi}^{-1}\Big(\mathbb{E}_{\boldsymbol{\alpha},N_\sigma}\left(\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-})\right)\Big), & 0\leq n\leq N_T-1,\\
\hat{\boldsymbol{P}}_{n+1}^-=\frac{1}{\rho}\mathbb{C}ov_{\boldsymbol{\alpha},N_\sigma}\left(\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-})\right)+\boldsymbol{B}_n\boldsymbol{Q}_n\boldsymbol{B}_n^\intercal, & 0\leq n\leq N_T-1.
\end{cases} \tag{15}$$

Note that these transformations also apply to the model parameters. Moreover, when $\boldsymbol{B}_n\neq 0$*, this approach is valid only if the modeling error acts on the transformed variable* $\boldsymbol{\varphi}(\hat{\boldsymbol{z}})$ *and not* $\hat{\boldsymbol{z}}$*. It may imply that the column of* $\boldsymbol{B}_n$ *are transformed accordingly. We refer to Section 4.2 for an enlightening example. Further, such transformation could also be used with to transform the measurement space, in order to change how observation errors are perceived by the UKF, for instance from a Gaussian distribution to a Poisson distribution using the Anscombe transform. Finally we want to emphasize that transforming the variables before each Kalman analysis step has common grounds with other transformed-based Kalman-like algorithms such as for instance [3].*

Remark 4 (Iterative Kalman procedure)**.** *For strong non-linearity, we know that the Kalman filter is only an approximation of the MLE that is often overcome in the literature by requiring a small number of iteration of the Kalman filter [17, Section 6.1]. In a case where we estimate parameters and initial condition through smoothing procedure, an iterative procedure can be easily set-up by considering multiple loops where new* a-priori *comes from the final* a-posteriori *of the previous estimation. Such procedure is in fact quite similar in spirit to the Bayesian approach described in [42].*

3.3 A reduced-order version of the sequential optimal estimator

The computation of the full covariance matrices $\hat{\boldsymbol{P}}_n^+$ and then $\hat{\boldsymbol{P}}_{n+1}^-$ can be prohibitive when the number of unknowns, parameters, and subjects increase. In this section, we therefore propose a strategy of uncertainty reduction. Assuming that $\hat{\boldsymbol{P}}_n^-$ is of reduced rank – typically much smaller than the dimension of the space $N_P \times N_z$ – the main idea in the reduced-order filtering strategies [39, 37, 32] is to be able to maintain covariance matrices in their factorized form

$$\hat{\boldsymbol{P}}_n^+ = \boldsymbol{L}_n (\boldsymbol{U}_n^+)^{-1} \boldsymbol{L}_n^\intercal \text{ follow by } \hat{\boldsymbol{P}}_{n+1}^- = \boldsymbol{L}_{n+1} (\boldsymbol{U}_{n+1}^-)^{-1} \boldsymbol{L}_{n+1}^\intercal \tag{16}$$

with $\boldsymbol{U}_n^+$ and $\boldsymbol{U}_{n+1}^-$ invertible matrices of small size which condense the system uncertainties and $\boldsymbol{L}_n$ an extension operator. What is crucial here is to be able to perform all computations on $\boldsymbol{L}_n$ and $\boldsymbol{U}_n^\pm$ without storing $\hat{\boldsymbol{P}}_n^\pm$. In this work, we follow the path of our Unscented Kalman strategy and use the reduced-order Unscented-Kalman Filter (RoUKF) introduced in [32]. In fact, we only need to define the initial reduced covariance matrix $\boldsymbol{U}_0$ and the initial extension matrix $\boldsymbol{L}_0$ from the initial covariance matrix $\hat{\boldsymbol{P}}_0$.

In this work, we proposed one possible strategy – among others see Remark 5 – of defining $\boldsymbol{U}_0$ and $\boldsymbol{L}_0$. We here choose 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

$$\mathbb{Y} = \begin{pmatrix} (y_0^1)_1, \cdots, (y_{N_{T_{\text{obs}}}}^1)_1, \ \cdots, (y_0^1)_{N_{\text{obs}}}, \cdots, (y_{N_{T_{\text{obs}}}}^1)_{N_{\text{obs}}} \\ \vdots \\ (y_0^{N_P})_1, \cdots, (y_{N_{T_{\text{obs}}}}^{N_P})_1, \ \cdots, (y_0^{N_P})_{N_{\text{obs}}}, \cdots, (y_{N_{T_{\text{obs}}}}^{N_P})_{N_{\text{obs}}} \end{pmatrix},$$

with $(y_k^i)_l$ corresponding 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 subjects 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 individuals belonging to the s^{th}-cluster.

We then define the reduced initial covariance matrix $\boldsymbol{U}_0^{-1}$ of dimension $(N_C \times N_z)^2$ computed by blocks. For $1 \leq r, s \leq N_C$, the block $(\boldsymbol{U}_0^{-1})_{r,s}$ 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

$$(\boldsymbol{U}_0^{-1})_{r,s} = \frac{1}{N_{c_r} N_{c_s}} \sum_{i \in c_r} \sum_{j \in c_s} (\hat{\boldsymbol{P}}_0)_{i,j}. \tag{17}$$

Concerning the extension operator $\boldsymbol{L}_0$ of dimension $(N_P \times N_z) \times (N_C \times N_z)$, the block $(\boldsymbol{L}_0)_{i,r}$ for $1 \leq i \leq N_P, 1 \leq r \leq N_C$ of indexes in $[(i-1)\,N_z + 1, i\,N_z] \times \left[(r-1)\,N_z + 1, r\,N_z\right]$ could be first given by

$$(\boldsymbol{L}_0)_{i,r} = \begin{cases} \mathbb{I}_{N_z}, & \text{if } i \in c_r, \\ \mathbb{0}_{N_z}, & \text{otherwise.} \end{cases} \tag{18}$$

However, such definition is not acceptable as it constraints the dynamics of all the subjects of a given cluster to be the same. As the *k-means* algorithm is based on the squared Euclidean distance metric, we can compute how much each subject $i \in [1, N_P]$ belongs to one cluster c_s with $s \in [1, N_C]$ by computing the following weight

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

Therefore, we prefer to replace (18) by

$$(\boldsymbol{L}_0)_{i,r} = \beta_{i,r} \mathbb{I}_{N_z}. \tag{19}$$

Remark 5 (Other reduction techniques)**.** *We here propose a practical definition of $\boldsymbol{L}_0$, $\boldsymbol{U}_0$ based on the* k-means*, but alternative strategies could have been followed, for instance hierarchical clustering or a principal component analysis of the $\mathbb{Y}\mathbb{Y}^\intercal$ as performed in the Proper-Orthogonal-Decomposition approach presented in [9].*

We conclude this section by giving our reduced estimator algorithm based on the reduced-order Unscented-Kalman Filter (RoUKF) [32]. We emphasized that in this algorithm, the number of sigma-points $N_\sigma^r \ll N_\sigma$ used in the full UKF approach (13). Therefore, the associated unitary sigma-points $\boldsymbol{e}_r^{(j)}$ generate a subspace of small dimension N_σ^r. Moreover, we here complement [32] by several features. First, we integrate the non-linear transformation presented in Remark 3. Second, we add the impact of the model error operator, that need to be projected in the reduced subspace generated by the column of $\boldsymbol{L}_n$ so that the decomposition (16) is maintained over the time iteration. In this respect, we follow [38, 37] using the orthogonal projector $\boldsymbol{\Pi}_n = \boldsymbol{L}_n (\boldsymbol{L}_n^\intercal \boldsymbol{L}_n)^{-1} \boldsymbol{L}_n^\intercal$ – i.e.

$\boldsymbol{\Pi}_n^2 = \boldsymbol{\Pi}_n$ and $\boldsymbol{\Pi}_n^\intercal = \boldsymbol{\Pi}_n$. Third, the forgetting factor $0 \leq \rho \leq 1$ that we have introduced in the UKF algorithm is preserved in the RoUKF algorithm by analogy with [38, 37]. We finally get

$$
\begin{cases}
\text{Initialization:} & \\
\hat{\boldsymbol{z}}_0^- = \vec{1}_{N_P} \otimes z_0, \quad \boldsymbol{L}_0 = \boldsymbol{L}_0, \quad \boldsymbol{U}_0 = \boldsymbol{U}_0, & \\
\text{Sampling:} & \\
\boldsymbol{\zeta}_n^{(j)-} = \boldsymbol{\varphi}(\hat{\boldsymbol{z}}_n^-) + \delta \boldsymbol{L}_n \sqrt{(\boldsymbol{U}_n^-)^{-1}} \boldsymbol{e}^{(j)}, & 0 \leq n \leq N_T, 1 \leq i \leq N_\sigma^r, \\
\text{Correction, when } T_n \in \{t_0, \cdots, t_{N_{\text{obs}}}\}: & \\
\tilde{\boldsymbol{y}}_n^{(j)} = \boldsymbol{c}_n\big(\boldsymbol{\varphi}^{-1}(\boldsymbol{\zeta}_n^{(j)-})\big) - \boldsymbol{y}_n, & 0 \leq n \leq N_T, 1 \leq i \leq N_\sigma^r, \\
\boldsymbol{\Lambda}_n = \mathbb{C}\text{ov}_{\boldsymbol{\alpha}, N_\sigma^r}\big(\boldsymbol{e}^{(\cdot)}, \tilde{\boldsymbol{y}}_n^{(\cdot)}\big)^T \boldsymbol{W}_n^{-1}, & 0 \leq n \leq N_T, \\
\boldsymbol{U}_n^+ = \boldsymbol{U}_n^- + \boldsymbol{\Lambda}_n \boldsymbol{W}_n \boldsymbol{\Lambda}_n, & 0 \leq n \leq N_T, \\
\boldsymbol{K}_n = \boldsymbol{L}_n (\boldsymbol{U}_n^+)^{-1} \boldsymbol{\Lambda}_n, & 0 \leq n \leq N_T, \\
\hat{\boldsymbol{z}}_n^+ = \boldsymbol{\varphi}^{-1}\Big(\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_n^-) + \boldsymbol{K}_n\big(\boldsymbol{y}_n - \boldsymbol{c}_n(\hat{\boldsymbol{z}}_n^-)\big)\Big), & 0 \leq n \leq N_T, \\
\text{Sampling:} & \\
\boldsymbol{\zeta}_n^{(j)+} = \boldsymbol{\varphi}(\hat{\boldsymbol{z}}_n^+) + \delta \boldsymbol{L}_n \sqrt{(\boldsymbol{U}_n^+)^{-1}} \boldsymbol{e}^{(j)}, & 0 \leq n \leq N_T - 1, 1 \leq i \leq N_\sigma^r, \\
\text{Prediction:} & \\
\hat{\boldsymbol{z}}_{n+1}^{(j)-} = \boldsymbol{\psi}_{n+1|n}\big(\boldsymbol{\varphi}^{-1}(\boldsymbol{\zeta}_n^{(j)+})\big), & 0 \leq n \leq N_T - 1, 1 \leq i \leq N_\sigma^r, \\
\hat{\boldsymbol{z}}_{n+1}^- = \boldsymbol{\varphi}^{-1}\Big(\mathbb{E}_{\boldsymbol{\alpha}, N_\sigma}\big(\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-})\big)\Big), & 0 \leq n \leq N_T - 1, \\
\boldsymbol{L}_{n+1} = \mathbb{C}\text{ov}_{\boldsymbol{\alpha}, N_\sigma^r}\big(\boldsymbol{e}^{(\cdot)}, \boldsymbol{\varphi}(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-})\big), & 0 \leq n \leq N_T - 1 \\
\boldsymbol{U}_{n+1}^- = \Big[\rho \mathbb{C}\text{ov}_{\boldsymbol{\alpha}, N_\sigma}(\boldsymbol{e}^{(\cdot)}) + \boldsymbol{\Xi}_{n+1} \boldsymbol{B}_n \boldsymbol{Q}_n \boldsymbol{B}_n^\intercal \boldsymbol{\Xi}_{n+1}^\intercal\Big]^{-1}, & 0 \leq n \leq N_T - 1,
\end{cases}
\tag{20}
$$

where $\boldsymbol{\Xi}_{n+1} = (\boldsymbol{L}_{n+1}^\intercal \boldsymbol{L}_{n+1})^{-1} \boldsymbol{L}_{n+1}^\intercal$ is such that $\boldsymbol{\Pi}_{n+1} = \boldsymbol{L}_{n+1} \boldsymbol{\Xi}_{n+1}$.

Remark 6 (Fisher information matrix)**.** *One additional benefit of the reduced-order formulation is that it is based on a square-root formulation of the Kalman filtering, which allows to easily compute a Fisher information matrix. Indeed, $\boldsymbol{\Lambda}_n$ is the sensitivity matrix of the observation errors with respect to the state. Therefore, computing*

$$
\boldsymbol{\Upsilon}_n = \sum_{n=0}^{N_T} \boldsymbol{\Lambda}_n \boldsymbol{W}_n \boldsymbol{\Lambda}_n,
$$

then taking the parameter block, and finally averaging over the members of the population gives

$$
\boldsymbol{I}^\theta = \frac{1}{N_P} \sum_{i=1}^{N_P} (\boldsymbol{\Upsilon}_n)_{i,i}^\theta,
$$

which is an estimation of the Fisher information matrix as computed in other Unscented Kalman Filter works [6].

4 Numerical results

In this section, we validate and illustrate our population based-Kalman filter on examples of different natures. In Section 4.1, we consider a toy problem corresponding to a pharmacokinetics one-compartment model with first-order absorption and elimination. We first validate our strategy using synthetic data by evaluating the performance of Algorithm 13 using the coupled covariance matrix $\hat{\boldsymbol{P}}_0$. Then, we look at the performance of Algorithm (20) by applying the clustering algorithm to the coupled covariance matrix $\hat{\boldsymbol{P}}_0$ (reduced-order population estimation). We conclude our investigation on this toy problem by considering a real data estimation.

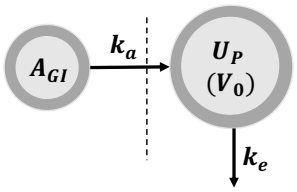
In Section 4.2, we will consider a more complex configuration from an actual estimation problem of the COVID-19 epidemic early evolution at 2 geographical resolutions – in regions ($N_P = 12$) and in departments ($N_P = 94$) – of metropolitan non-islandic France. The estimation objective is to reconstruct the lockdown health policy impact to the transmission rate in the early month of the virus propagation using an extended SIR type model [51] (from March 1$^{\text{st}}$ and May 11$^{\text{th}}$, 2020). We find this set-up emblematic as it requires the use of all the ingredients proposed in our methodology.

4.1 A Pharmacokinetics toy problem

4.1.1 Model and observations

We 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

(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)

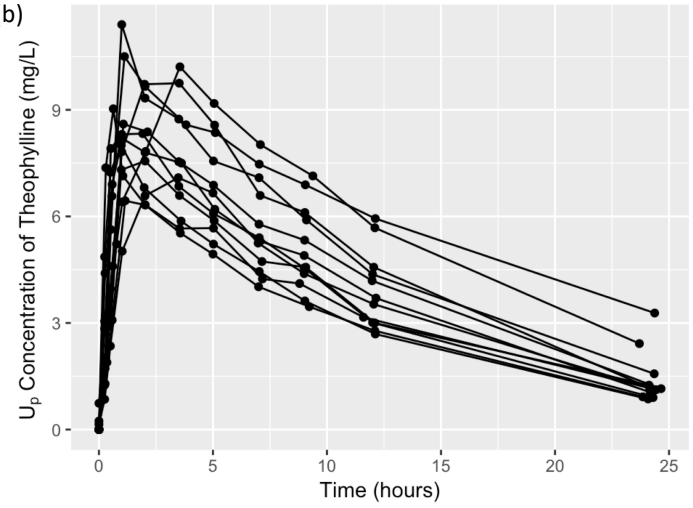


Figure 1: One-compartment PK model with oral absorption: (a) Representation of the mathematical model, (b) A benchmark dataset of the anti-asthmatic drug theophylline for 12 patients.

plasma (in mg). The interactions between A_{GI} and A_P are described in the following system:

$$\left\{ \begin{array}{rcl} \dot{A}_{GI} & = & -k_a A_{GI}, \\ \dot{A}_P & = & k_a A_{GI} - k_e A_P, \end{array} \right. \tag{21}$$

where k_a corresponds to the absorption rate (in h^{-1}) and k_e to the elimination rate (in 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 (21) can be rewritten as follows:

$$\left\{ \begin{array}{rcl} \dot{A}_{GI} & = & -k_a A_{GI}, \\ \dot{U}_P & = & \dfrac{k_a}{V_0} A_{GI} - k_e U_P, \end{array} \right. \tag{22}$$

see Figure 1(a) for an illustration.

Remark 7. *Obviously, the analytical solution of System* (22) *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,

$$z^i = \begin{pmatrix} A_{GI}^i \\ U_P^i \\ \log(k_a^i) \\ \log(k_e^i) \\ \log(V_0^i) \end{pmatrix}, \; a(z^i(t), t) = \begin{pmatrix} -k_a A_{GI} \\ \dfrac{k_a}{V_0} A_{GI} - k_e U_P \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad B(t) = 0, \;\; 1 \leq i \leq N_P.$$

Note that this first example is therefore based on a fully deterministic model as it is often the case in data assimilation [28, Section 2.1].

Concerning the measurements, the observation operator is given by

$$c_k^i(z^i(t_k^i)) = U_P^i(t_k)^{0.25}, \quad 1 \leq i \leq N_P, \; 0 \leq k \leq N_{T_{\mathrm{obs}}}. \tag{23}$$

The measurements are subject to stochastic noise with constant variance $W_k^i = w$ for each subject $1 \leq i \leq N_P$ and for each observation $0 \leq k \leq N_{T_{\mathrm{obs}}}$.

Another important aspect is the ability to handle the data time-sampling which, in practice, is imposed and can be rather different from the model time-step. In fact in general, the data time-sampling could also differ 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. To simplify our implementation, we have decided to interpolate the observations. More precisely, using the time discretization notations of the system introduced in Section 3, we first reconstruct the measurements for $1 \leq i \leq N_P, 0 \leq n < N_T$

$$y_n^i = \frac{t_{k+1} - t_n}{t_{k+1} - t_k} y^i(t_k) + \frac{t_k - t_n}{t_{k+1} - t_k} y^i(t_{k+1}), \quad \text{if } t_n \in [t_k, t_{k+1}).$$

These measurements are consistant with an observation operator c_k^n of the form of (23), with additional rescaling of $W_n^i \propto \frac{t_{k+1}-t_k}{t_{n+1}-t_n} w$. Finally, in our algorithms, we set $\rho = 1$, $\delta = 1$.

4.1.2 Validation using synthetic data

First, we are going to evaluate the performance of Algorithm 13 using the coupled covariance matrix $\hat{\boldsymbol{P}}_0$ (see Eq. (9)) (population estimation) with respect to using the diagonal covariance matrix $\hat{\boldsymbol{P}}_0^{\text{un}}$, see Eq. (12) (individual estimation). Then, we will evaluate the performance of the reduced approach (20) by applying the clustering algorithm to the coupled covariance matrix $\hat{\boldsymbol{P}}_0$ (reduced-order population estimation).

Data and statistical validation criteria – We consider a sequence 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

$$z_0 = \begin{pmatrix} (A_{GI})_0 \\ (U_P)_0 \\ \log(k_a)_0 \\ \log(k_e)_0 \\ \log(V_0)_0 \end{pmatrix} = \begin{pmatrix} 500 \\ 0 \\ -4.6 \\ -5.56 \\ -4.19 \end{pmatrix} \text{ and } \hat{P}_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2^2 & 0 & 0 \\ 0 & 0 & 0 & 0.25^2 & 0 \\ 0 & 0 & 0 & 0 & 0.1^2 \end{pmatrix}, \ 1 \leq i \leq N_P.$$

as mean values and covariances of the initial state conditions and parameters. The covariances are diagonal because the parameters are assumed to be independent variables. For the estimation process, we set the following *a-priori* values and covariances:

$$\hat{z}_0 = \begin{pmatrix} (A_{GI})_0 \\ (U_P)_0 \\ \log(k_a)_0 \\ \log(k_e)_0 \\ \log(V_0)_0 \end{pmatrix} = \begin{pmatrix} 500 \\ 0 \\ -3 \\ -5 \\ -3 \end{pmatrix} \text{ and } \hat{P}_0 = \begin{pmatrix} 0.001^2 & 0 & 0 & 0 & 0 \\ 0 & 0.001^2 & 0 & 0 & 0 \\ 0 & 0 & 3^2 & 0 & 0 \\ 0 & 0 & 0 & 3^2 & 0 \\ 0 & 0 & 0 & 0 & 3^2 \end{pmatrix}, \ 1 \leq i \leq N_P,$$

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

To evaluate the estimation performance, various statistical validation criteria can be considered. For a parameter θ, we define by θ_t (resp. θ_e) the matrix of size $N_R \times N_P$ containing in each line r the target (resp. estimated) values of θ for all the patients of the replicate r. We further compute The relative bias RBIAS

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

and the mean squared error MSE

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

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 the estimates. Using $\hat{P}_n^i$, we compute σ_e^θ the matrix of size $N_R \times N_P$ containing on each line r the estimated standard deviation of θ for all the patients of the replicate r. We additionally complement the RBIAS and MSE by 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 the estimates defined as the standard deviation of the estimates at convergence (ESTD)

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

Finally, we can compute the relative bias on standard deviation for random effects called BMIXED and defined by

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

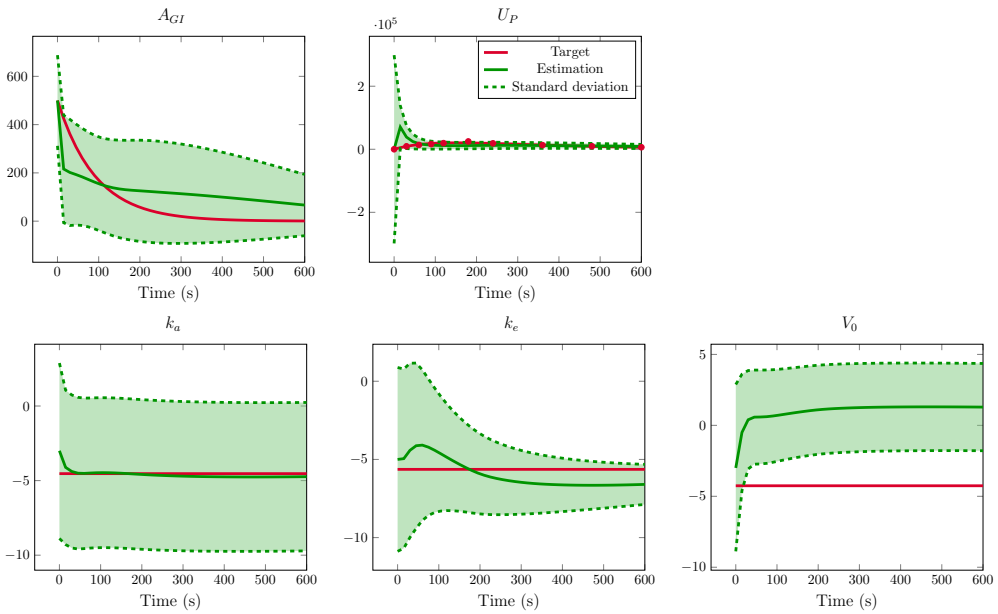


Figure 2: Time evolution of states A_{GI} and U_P and parameters k_a, k_e and V_0 for one patient using individual estimation with N_P=20 and $w_t = 0.3$. The target states and target values are in red. The red dots on the graph of U_P corresponds to the noisy observed data. The states and parameters values with individual estimation are in green. The dashed lines correspond to the 95% confidence interval.

Meas. Error	Sample Size	Individual				Population			
		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.

where $\sigma^{\theta}_{e,m}$ is the matrix of size $N_R \times N_P$ containing on each line r the estimated mixed standard deviation of θ for all the patients of the replicate r, hence built using $\tilde{\hat{P}}^i_n$. This indicator is completed with the coverage (COV) defined as the percentage of time when 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 when using an individual approach, for all sample sizes, estimates are similar in terms of precision and accuracy. We note that the individual Kalman filter do not properly estimate the parameters as underlined in Figure 2 where the time evolution of the state and parameters is plotted for one patient. The poor estimation performance is in fact mainly due to the poor parameters *a-priori* which is not counterbalanced in the individual approach.

By contrast, as shown in Table 1 the population approach really strengthens the estimation as the RBIAS and the MSE decrease with sample size and are dramatically smaller than for the individual (for $N = 100$ RBIAIS goes from 0.374 for individual approach to -0.007 for population approach). This is also confirmed by the resulting individual trajectories as illustrated in Figure 3 (the first column) with the time evolution of the state and parameters of one patient. We notice the STD is highly bigger than the ESTD. It denotes an overestimation of the standard deviation of estimates due to the fact that the initial estimation covariances where chosen very roughly (3^2 instead of $0.2^2, 0.25^2$ and 0.1^2 for respectively k_a, k_e and V_0). Then, in order to obtain a more accurate covariance for final estimate, we iterate the estimation procedure following Remark 4. Each run will be called a loop. The population procedure is evaluated for $w_t = 0.3$ and N_P=20 with Table 2 summarizing the results for each parameter of the

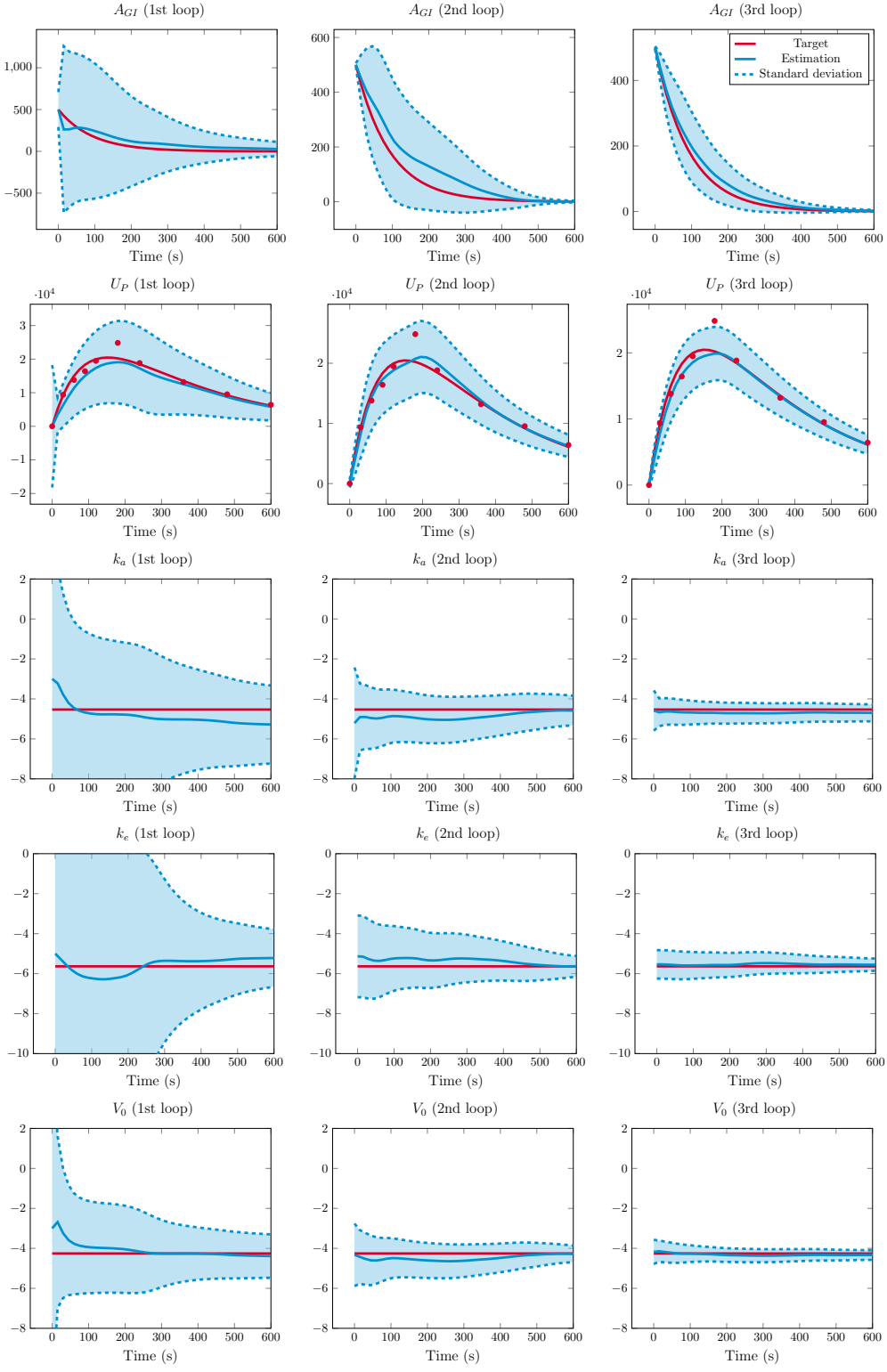


Figure 3: Time evolution of states A_{GI} and U_P and parameters k_a, k_e and V_0 for one patient using population estimation with N_P=20 and $w_t = 0.3$. Each column correspond to one loop of the algorithm. The target states and target values of parameters are in red. The red dots on the graph of U_P corresponds to the noisy observed data. The states and parameters values with population estimation are in blue. The dashed lines correspond to the 95% confidence interval.

		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.

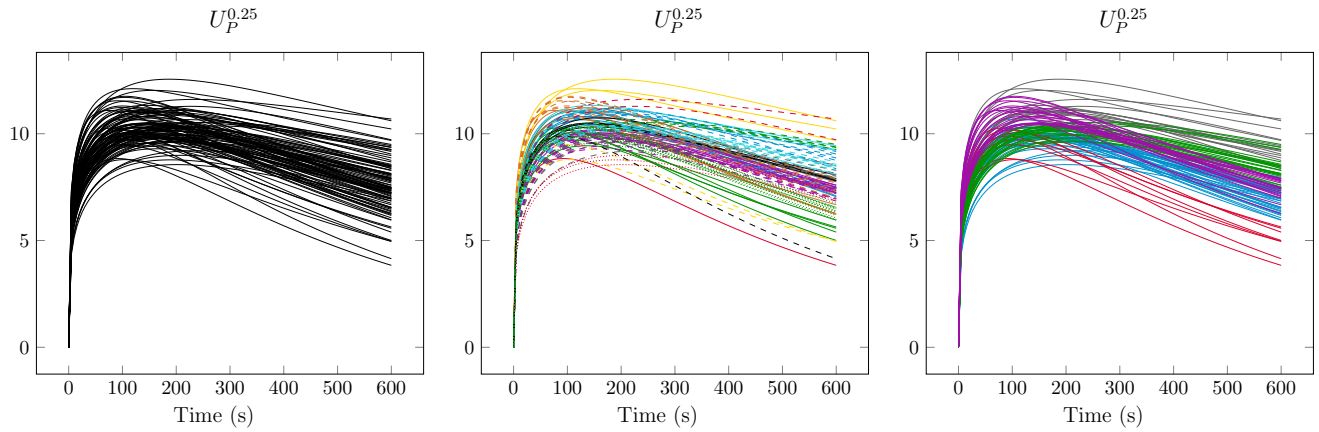


Figure 4: Illustration of the clustering algorithm (*k-means* algorithm) on the observation variable $U_P^{0.25}$ for a synthetic data set of 100 patients. Left: full data set. Middle: results of *k-means* algorithm with 20 clusters. Right: results of *k-means* algorithm with 5 clusters. Each color and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

model. As expected, the 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%. In fact, we naturally recommend to perform multiple estimation loops when starting from poor covariance *a-priori*. 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 (6) containing the population intercept but the functional (7) 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 complexity can still be prohibitive, hence the interest of the reduced-order version of the population Kalman filter that we are going to illustrate. We 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. Table 3 presents the results which are given for each parameter and also aggregated over the 3 parameters. These estimation performances are completed by computational times obtained with a 2,3 GHz Quad-Core Intel Core i7 computer and an implementation in `MATLAB-2020-b`. This gives an order of magnitude of the algorithm numerical complexity, although with a not fully optimized implementation. If we focus on the MSE, we obtain equivalent results for the full covariance and the reduced covariance approach. Concerning the RBIAS, the results obtained with the reduced version remain acceptable. As expected, 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. Using the reduced-order filter, the COV gets closer to their nominal value of 95% and finally, it also improves the BMIXED. This improvements can be explained by the fact that the reduced-order version of the population Kalman filter increases the coupling between the subjects. 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 obtained with the full population Kalman filter (top) and its reduced-order version with $N_C = 20$ (middle) and $N_C = 5$ (bottom). Typically, we see 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 (in the acceptable range) for V_0 – which is knwon to be the most difficult parameter of

		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 $\sigma_m = 0.3$.

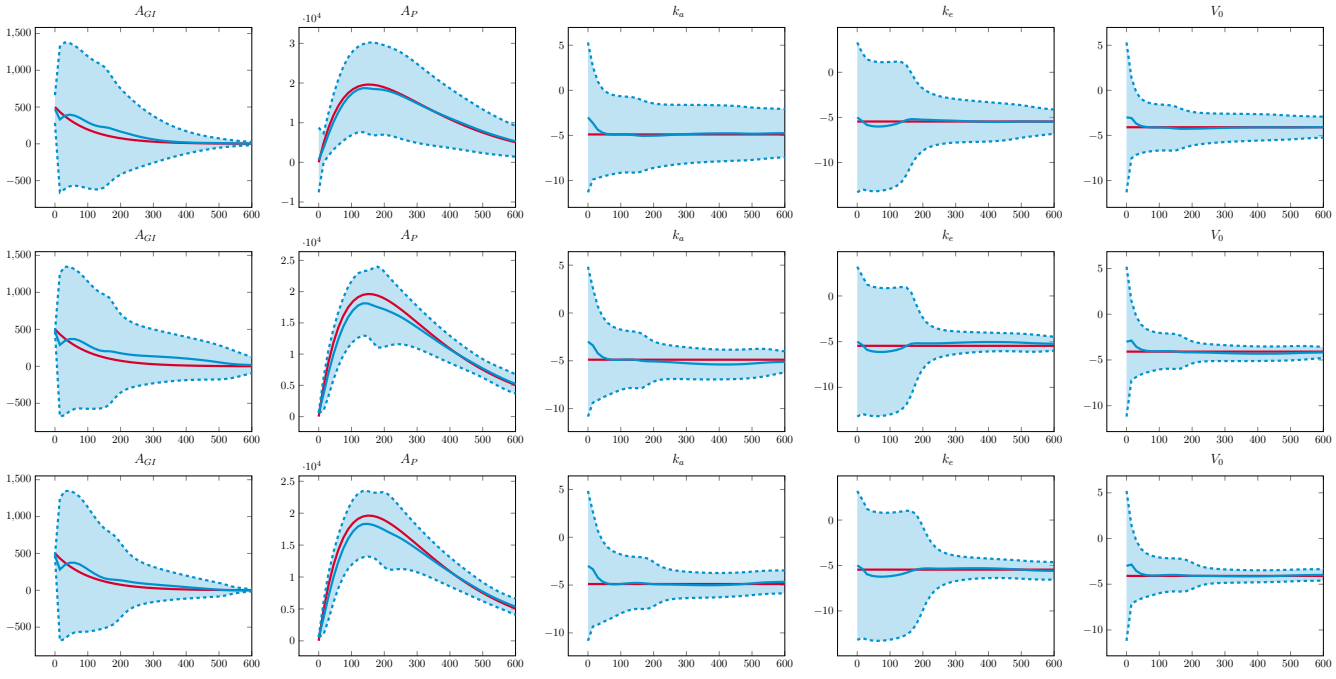


Figure 5: Time evolution of states A_{GI} and U_P and parameters k_a, k_e and V_0 for one patient using the full version of the algorithm (top) and the reduced-order versions (middle: $N_C = 20$, bottom: $N_C = 5$) with N_P=100 and $\sigma_m = 0.3$. The target states and parameters are in red. The estimated states and parameters are in blue. The dashed lines correspond to the 95% confidence interval.

the system to estimate. Figure 5 shows that increasing the coupling by reducing the number of clusters reduces the standard deviations removing the need of multiple loops. Finally with respect to 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 $N_C = 5$.

4.1.3 Experiment with real data

	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
σ_{k_a}	-	-	0.403	0.179	-	-	0.136	0.133
σ_{k_e}	-	-	0.015	0.008	-	-	0.083	0.010
σ_{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.

To conclude this toy problem study, we propose a validation on a real data set. 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 [48]. 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 7).

Individual nonlinear estimation is implemented using the `R`-package *nls* and population NLME model is implemented using the `R`-package *saemix* [10]. Our population-based Kalman filter gives very similar results to the population NLME, see Table 4. It is also interesting to note that the individual Kalman filter gives better results than the individual nonlinear strategy, in particular concerning the estimation of V_0 which is the most difficult parameter to estimate because it determines the slope of U_p and is very sensitive to the observation noise. This shows that the individual Kalman filter can be efficient when the *a-priori* are better initialized, a typical advantage brought by the population approach.

4.2 COVID example

We model the evolution of the COVID epidemic using a SIR type model [51] inferred with hospitalization data. The estimation objective is to infer the impact of the various government interventions, the weather and the variants on the transmission rate from March 1$^{\text{st}}$ 2020 and March 30$^{\text{th}}$ 2021.

4.2.1 SEIRAH model

More precisely, we use a SEIRAH model adapted from [51] where the population of size N is divided into 6 compartments: susceptible S, latent E, ascertained infectious I, unascertained infectious A, hospitalized H, and removed R (i.e. both recovered and deceased). The dynamics of such model is given by

$$\begin{cases} \dfrac{\mathrm{d}S}{\mathrm{d}t} = -\dfrac{b\,S(I+\alpha A)}{N} \\ \dfrac{\mathrm{d}E}{\mathrm{d}t} = \dfrac{b\,S(I+\alpha A)}{N} - \dfrac{E}{D_E} \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \dfrac{r_E}{D_E}E - \dfrac{1-r_I}{D_q}I - \dfrac{r_I}{D_I}I, \\ \dfrac{\mathrm{d}R}{\mathrm{d}t} = \dfrac{r_I I + A}{D_I} + \dfrac{H}{D_H} \\ \dfrac{\mathrm{d}A}{\mathrm{d}t} = \dfrac{1-r_E}{D_E}E - \dfrac{A}{D_I} \\ \dfrac{\mathrm{d}H}{\mathrm{d}t} = \dfrac{1-r_I}{D_q}I - \dfrac{H}{D_H}. \end{cases} \tag{24}$$

where $\alpha, r_E, D_E, r_I, D_I, D_q, D_H$ are time-independent parameters described in Table 5 while $t \mapsto b(t)$ is a function of time modeling the disease transmission over time resulting from the health policy followed in France during the first wave of the COVID-19 epidemic.

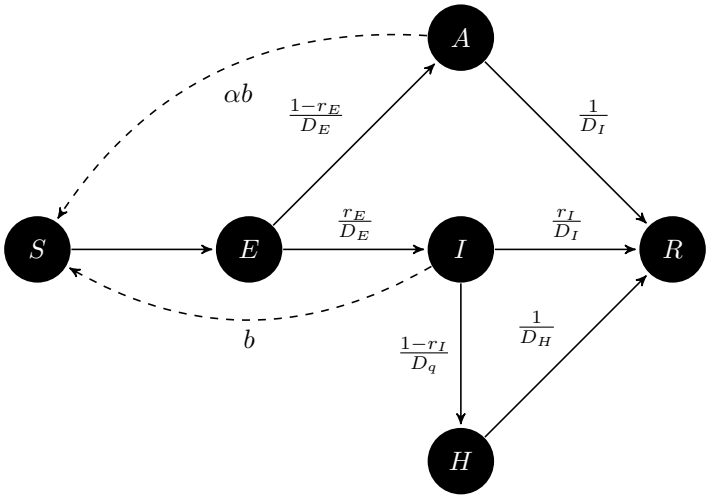


Figure 6: SEIRAH model representation – adapted from [51]

Parameter	Interpretation	Value
b	Transmission rate of ascertained cases	Region/Department Specific - **Estimated**
r_E	Ascertainment rate	0.844 [18]
r_I	Non hospitalized rate	0.966 [1]
α	Ratio of transmission between A and I	0.55 [29]
D_E	Latent (incubation) period (days)	5.1 [25]
D_I	Infectious period (days)	5 [7]
D_q	Duration from I onset to H (days)	11-D_E [12]
D_H	Hospitalization period (days)	17.3*
N	Population size	Region/Department Specific

Table 5: Model parameters for the SEIRAH model, and associated values. *Computed using the correlation between the data when considering region data, see Appendix A.

In this respect, System (24) should be completed in each region or department by a dynamics for b. We consider a dynamics of the form

$$\mathrm{d}b = g(t)\mathrm{d}t + \mathrm{d}\omega(t)$$

where g is a function describing the health policy effect to the transmission rate b and ω consists in an additional Wiener process. For instance, a transmission rate decreases from b_{init} to b_{end} due to a lockdown policy centered around the time t_ℓ with efficiency time-rate $\frac{1}{\tau}$ can typically be modeled with a logistic function

$$b(t) = \ell(t) \stackrel{\text{def}}{=} b_{\text{init}} - \frac{b_{\text{init}} - b_{\text{end}}}{1 + e^{-\frac{t-t_\ell}{\tau}}} \Rightarrow g(t) = \ell'(t).$$

Furthermore, multiple lockdown dynamics can be modeled with a combination of such g function during intervals, whereas unlockdown can also be represented by such g function with opposite sign. Note that if g is an accurate model of the lockdown effect, then ω is the Wiener process representing a modeling error. However, choosing $g \equiv 0$ traduces the absence of any *a priori* on the lockdown effect, forcing the data assimilation procedure to estimate the lockdown through the estimation of the process ω realization.

Following our methodology, we discretize the SEIRAH model using a forward Euler time-scheme with a small-enough time-step δt. The resulting discrete-time dynamical system involves the state variable $x = (E, I, R, A, H)^\intercal \in \mathbb{R}^5$ coupled to the dynamics of b and parameters θ to be estimated

$$\begin{pmatrix} x^i_{n+1} \\ b^i_{n+1} \\ \theta^i_{n+1} \end{pmatrix} = \begin{pmatrix} x^i_n + \delta t\, f(x^i_n, b^i_n, \theta^i_n) \\ b^i_n + \delta t\, g^i(t_n, \theta^i_n) \\ \theta^i_n \end{pmatrix} + \begin{pmatrix} \mathbb{0}_5 \\ 1 \\ \mathbb{0}_{N_p} \end{pmatrix} \nu_n, \quad \text{for each region/department } i$$

where f accounts for the dynamics in (24) and S is reconstructed afterwards by using $S = N - (E + I + R + A + H)$ in each region/department. Recalling that $\theta \in \mathbb{R}^{N_p}$ gathers all the unknown time-independent parameters that need to be estimated, here b_{end} and b_{init} characterizing the g^i functions.

As all state variables E, I, R, A, H are positive and bounded by N the total population size of the department, following our Remark 3, they are going to be mapped with a scaling and *logit* transform:

$$\text{logit}\Big(\frac{E}{N}\Big), \text{logit}\Big(\frac{I}{N}\Big), \text{logit}\Big(\frac{R}{N}\Big), \text{logit}\Big(\frac{A}{N}\Big), \text{logit}\Big(\frac{H}{N}\Big).$$

Concerning the variable b, we all use a weighted *logit* function: $\max_b \text{logit}(b)$, where $\max_b$ has been fixed to 1.5 after preliminary investigation. This allows to estimate b positive and inferior to $\max_b$.

When neglecting the deaths, the proportion of infected individuals among the population in each region/department at a given date is given by

$$\frac{E+I+R+A+H}{N}.$$

This quantity – which can be computed using other approaches – will be used to validate our strategy.

4.2.2 Available data

To proceed to the estimation of these uncertain quantities, we have at our disposal the hospitalization and the daily incident number data – denoted respectively H_{data} and $H_{in,data}$ – collected in the the SI-VIC database that recording COVID-19 patients hospitalizations in French hospitals. Those two quantities are related to the solutions of System (24) respectively as $H_{data} = H$ and $H_{in,data} = \frac{(1-r_I)}{D_q} I$.

4.2.3 Estimation in the 12 regions

Par.	IDF	CVL	BFC	Nor.	HDF	GE	PL	Bret.	NA	Occ.	AURA	PACA	mean
b_{init}	0.795	0.720	0.843	0.752	0.794	0.875	0.729	0.790	0.754	0.804	0.794	0.787	0.786
b_{end}	0.147	0.163	0.133	0.138	0.151	0.125	0.143	0.138	0.133	0.110	0.134	0.152	0.139
$100\left(1-\frac{b_{\text{end}}}{b_{\text{init}}}\right)\%$	81.6%	77.3%	84.2%	81.6%	81%	85.8%	80.4%	82.6%	82.3%	86.3%	83.1%	80.7%	82.2%

Table 6: Estimation of the parameters b_{init} and b_{end} at the regional level. The third line gives the corresponding decrease rate: $100\left(1-\frac{b_{\text{end}}}{b_{\text{init}}}\right)\%$.

We develop the following strategy allowing to avoid any *a priori* about the shape of b:

- Step 1 – Estimate the parameter b_{init} (resp. b_{end}) using the data obtained before the lockdown (resp. 10 days after the lockdown). The values are given in Table 6. One can see that the values are quite similar in all the regions. In this first step, the transmission rate b is assumed to a piecewise constant function equal to b_{init} before the lockdown and b_{end} after.
- Step 2 – Estimate the shape of b without the logistic function *a priori* (*i.e.* $g(t) = 0$, $\forall t$) but setting b_{init} at the value estimated in Step 1. The results are given in Figure 7-top-left for all the regions. A logistic function shape appears which confirms our modeling choice. As there is no information on b, the model error is very important and lead to an overfitting of the data. Oscillations of 0 mean and of period closed to 7 days are visible for many regions.
- Step 3 – In order to evaluate if these weekly plateaus are just a compensation of the lack of information about b, we propose to fix g as a logistic function (b_{init} and b_{end} resulting from Step 1, $\tau = 1$ and t_l being the 8^{th} day of lockdown). The results are given in Figure 7-top-right. In Figure 7-bottom-left, we compute the difference between the *a priori* logistic function b and its estimation. One can remark that in the regions with a high hospitalization rate: Île-de-France (IDF), Grand-Est and Bourgogne Franche Comté (BFC), we have $b_{estim}(t) > b_{apriori}(t)$.

When individuals are homogeneous and mix uniformly, the effective reproductive ratio $R_{\text{eff}}(t)$ is defined as the mean number of infections generated during the infectious period of a single infectious case at time t. In this model as recalled in Appendix B, the effective reproductive ratio can be written as a function of model parameters as:

$$R_{\text{eff}}(t) = b(t)\frac{S(t)}{N}\left(D_i\alpha(1-r_e) + \frac{D_iD_qr_e}{(1-r_i)D_i + r_iD_q}\right).$$

Figure 7-(bottom-right) compiles the estimated effective reproductive ratios. Before the lockdown the values of the effective reproductive ratios are around 3.5 a value comparable to the ones found in the literature of the basic reproductive ratio, see [44, 14]. The effective reproductive ratios at the end of the lockdown are below 0.7 showing the lockdown efficiency.

Proportion of infected individuals among the population in each region at the end of the first lockwdown (May 11^{th}, 2020) is given in Figure 8-top-left. These results are very similar to the ones given in [44] which allow to validate our strategy.

4.2.4 Estimation in the 94 departments

Let us now move on a more refined population and considering departments. Here, we only apply the final Step 3 using mean estimated values of b_{init} and b_{end} obtained during the 12-Regions estimation. This gives us the proportion of infected individuals among the population at the department level as reported in Figure 8-top-left. These results obtained with the full population Kalman filter are then compared to the results obtained using a reduced-order version with $N_C = 24$. Figure 8-bottom-right shows that the estimated infected individuals are very close to the

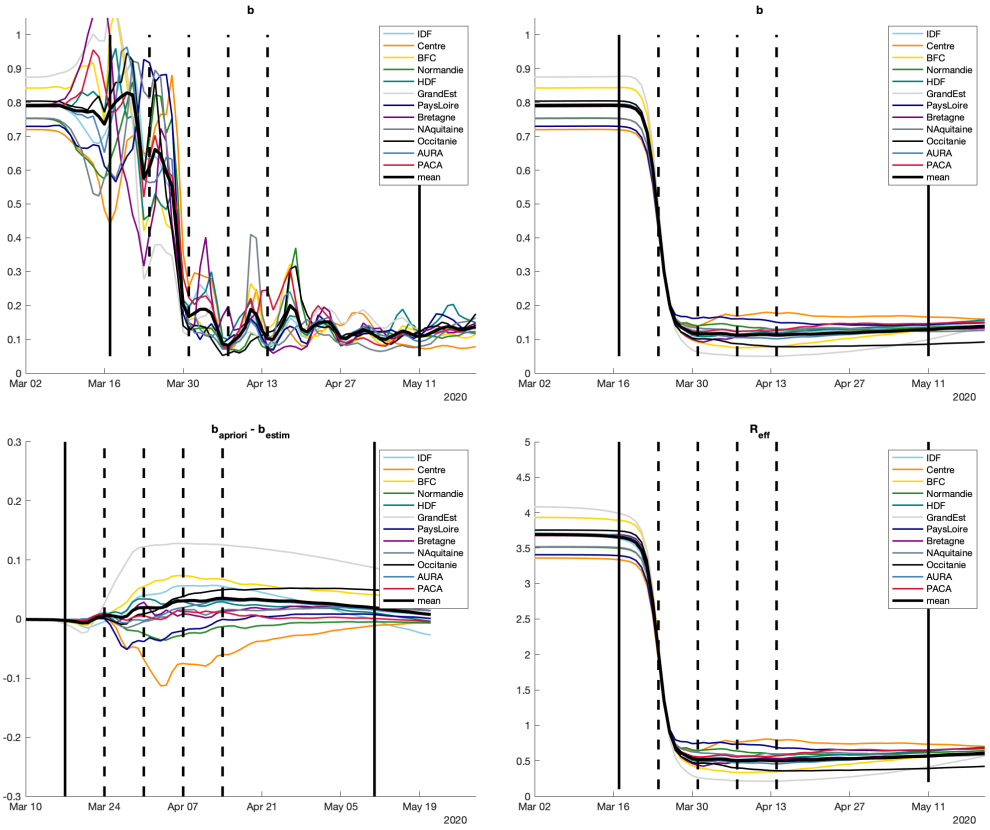


Figure 7: Left-Top: Time evolutions of b in the 12 french regions with $g = 0$ (no *apriori*). Right-Top: Time evolutions of b in the 12 french regions when g corresponds to a logistic function (with $\nu \neq 0$). Left-Bottom: Difference between g and b when g corresponds to a logistic function. Right-Bottom: Time evolutions of R_{eff} in the 12 french regions when g corresponds to a logistic function (with $\nu \neq 0$). The first (resp. last) vertical black line corresponds to the first (resp. last) day of lockdown. The other dashed lines are spaced every 7 days.

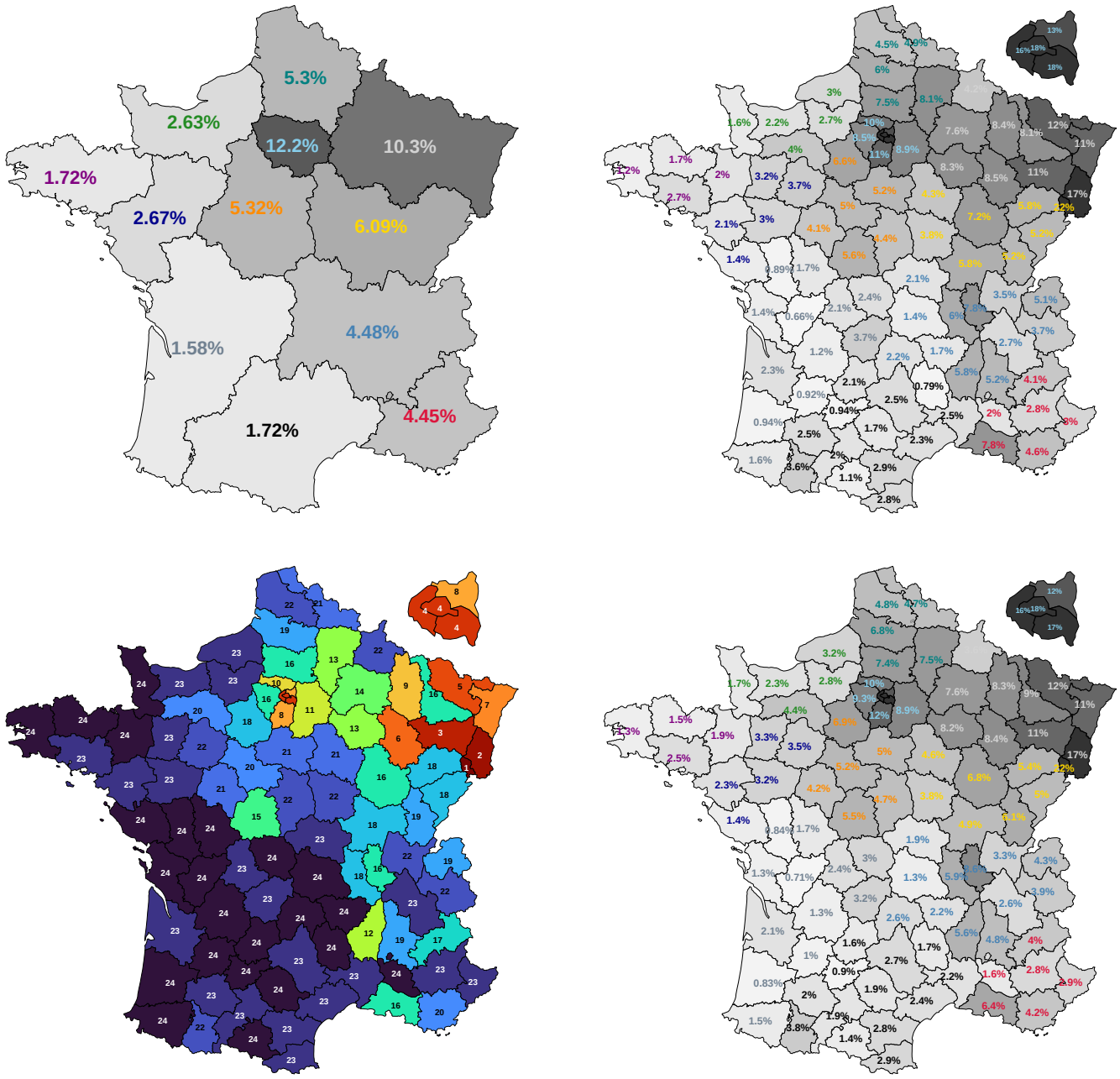


Figure 8: Model estimation for the proportion of Immunized individual in the population (deaths not taken into account) on May 11th, 2020. Top-Left: In the 12 regions. Top-Right: In the 94 departments. Bottom-Left: Clustering illustration. Bottom-Right: In the 94 departments using the reduced-order version of our population Kalman-based filter.

values obtained with the full population Kalman filter whereas the computational time is divided by 5: 40 min for the full filter against 8 min for its reduced order version on a 2,3 GHz Quad-Core Intel Core i7 computer with a sequential implementation of the algorithm using `MATLAB-2020-b`. We also remark that the cluster – see Figure 8-bottom-left – is not fully aligned with the Region partition, as the epidemic in France follow North-Est through South-West gradient. Therefore, our reduced-order strategy allows to obtain better results at the Department scale than the estimation performed at the Region scale, but for a comparable computational complexity.

All these results are going to be discussed in greater details in a forthcoming work dedicated to the epidemic. However, we believe that they already illustrate how versatile is our approach in terms of model complexity with appealing estimation performance and controlled computational complexity.

5 Conclusion

In this work, we have proposed a new strategy allowing to combine data assimilation principle to population configuration classically taking into account in mixed effect formulation. In this respect, assuming Gaussian disturbances up to some variable transformation, we have defined a log-likelihood functional which elegantly couples the subjects. We approximate the resulting minimization using a dynamic programing approach, here the Unscented Kalman Filter [20]. To decrease the computational times, we complete our approach with a reduced-order version of our population filter following a clustering of the measurements. Note that our choice of UKF could be replaced by an Extended Kalman Filter [46] or an Ensemble Kalman Filter [16] as soon as reduced order version are available [39, 37].

We have 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. Then we moved to a on a more complex configuration, with stronger nonlinearities and modeling errors.

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 obtained computational time pave the way of using our strategy with large-dimensional ODE models or possibly partial differential (PDE) systems.

Appendices

Appendix A Estimation of D_H using the correlation between the hospitalization data H and the daily incident number of hospitalization $H_{\text{in,data}}$

The relation between H_{data} and $H_{\text{in,data}}$ is governed by

$$\frac{\mathrm{d}}{\mathrm{d}t}H_{\text{data}} = -\frac{H_{\text{data}}}{D_H} + H_{\text{in,data}}. \tag{25}$$

The colleration betwen H_{data} and $H_{\text{in,data}}$ at the regional level allows to compute D_H in each region by using a mean squared estimation. We fix D_H at the mean value. The obtained values are given in Table 7. Concerning the data at the department level, the hospitalization data can not be trusted. Indeed, each region is composed of several departments and the sum between the hospitalization in all the departments of a region is not compatible with the hospitalization in the region (while this is the case for the daily incident number data). Furthemore, if we compute D_H at the departmental level, we obtain non realistic values with a minimal value at 10 days and a maximal at 30.5 days. This is in part due to a poor report in patient transferts between hospitals from different departments. For this reason, when considering departments, we only use the daily incident number data and we will fix D_H at the mean value estimated using the data at the regional level (given in Table 7).

Par.	IDF	CVL	BFC	Nor.	HDF	GE	PL	Bret.	NA	Occ.	AURA	PACA	mean
D_H	18.3	19.4	17.4	20.7	18.6	17.5	15.2	19.1	17.4	16.5	16.8	17.7	17.9

Table 7: Estimation of the parameter D_H at the regional level

Appendix B Computation of the effective reproductive ratio R_{eff}

To compute the reproductive ratio R_{eff} of our the SEIRAH model (24), we apply the *Next Generation Matrix* approach – see for instance [19]. The principle consists in focusing on three categories:

i) latent E, ii) ascertained infectious I and iii) unascertained infectious A with the following dynamics

$$\begin{cases} \dfrac{\mathrm{d}E}{\mathrm{d}t} = \dfrac{b\,S(I+\alpha A)}{N} - \dfrac{E}{D_E} \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \dfrac{r_E}{D_E}E - \dfrac{1-r_I}{D_q}I - \dfrac{r_I}{D_I}I, \\ \dfrac{\mathrm{d}A}{\mathrm{d}t} = \dfrac{1-r_E}{D_E}E - \dfrac{A}{D_I} \end{cases}$$

Then, we build two matrices corresponding to: (1) V following the arrivals and departures from one other category and (2) F following the arrivals from another compartment exterior to the three categories. We have

$$V = \begin{pmatrix} \frac{1}{D_E} & 0 & 0 \\ -\frac{r_E}{D_E} & \dfrac{1-r_I}{D_q} + \dfrac{r_I}{D_I} & 0 \\ -\dfrac{1-r_E}{D_E} & 0 & \frac{1}{D_I} \end{pmatrix} \text{ and } F = \begin{pmatrix} 0 & b\frac{S}{N} & \alpha b\frac{S}{N} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is then well known – see for instance –[35] for a proof –that

$$R_{\text{eff}} = \rho(FV^{-1}),$$

where $\rho(FV^{-1})$ is the spectral radius of the Next Generation Matrix FV^{-1}. Here, we have

$$FV^{-1} = \begin{pmatrix} b\dfrac{S}{N}\left(D_I\alpha(1-r_E) + \dfrac{D_I D_q r_E}{(1-r_I)D_I + r_I D_q}\right) & b\dfrac{S}{N}\dfrac{D_I D_q}{(1-r_I)D_I + r_I D_q} & b\dfrac{S}{N}D_I\alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with

$$V^{-1} = \begin{pmatrix} D_E & 0 & 0 \\ \frac{D_I D_q r_E}{(1-r_I)D_I + D_q r_I} & \frac{D_I D_q}{(1-r_I)D_I + D_q r_I} & 0 \\ (1-r_E)D_I & 0 & D_I \end{pmatrix}.$$

We therefore obtain

$$R_{\text{eff}}(t) = b(t)\frac{S(t)}{N}\left(D_I\alpha(1-r_E) + \frac{D_I D_q r_E}{(1-r_I)D_I + r_I D_q}\right).$$

References

[1] F. J. Angulo, L. Finelli, and D. L. Swerdlow. Estimation of US SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths Using Seroprevalence Surveys. *JAMA network open*, 4(1):e2033706–e2033706, 2021.

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

[3] A. Barrau and S. Bonnabel. The Invariant Extended Kalman Filter as a Stable Observer. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, 62(4):1797–1812, 2017.

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

[5] J. Blum, F.-X. Le Dimet, and I. 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.

[6] A. Caiazzo, F. Caforio, G. Montecinos, L. O. Muller, P. J. Blanco, and E. F. Toro. Assessment of reduced-order unscented Kalman filter for parameter identification in 1-dimensional blood flow models using experimental data. *International Journal for Numerical Methods in Biomedical Engineering*, 33(8):e2843, Aug. 2017.

[7] M. Cevik, M. Tate, O. Lloyd, A. E. Maraolo, J. Schafers, and A. Ho. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, 2020.

[8] 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.

[9] D. Chapelle, A. Gariah, P. Moireau, and J. Sainte-Marie. A Galerkin strategy with Proper Orthogonal Decomposition for parameter-dependent problems – Analysis, assessments and applications to parameter estimation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(6):1821–1843, 2013.

[10] 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.

[11] 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.

[12] J.-F. Delfraissy, L. Atlani Duault, D. Benamouzig, L. Bouadma, S. Cauchemez, F. Chauvin, A. Fontanet, A. Hoang, D. Malvy, and Y. Yazdanpanah. Une deuxième vague entrainant une situation sanitaire critique. 2020.

[13] 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.

[14] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza. Impact of lockdown on covid-19 epidemic in île-de-france and possible exit strategies. *BMC medicine*, 18(1):1–13, 2020.

[15] 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.

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

[17] A. Gelb, J. F. Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland. *Applied Optimal Estimation*. The MIT Press, 1974.

[18] J. He, Y. Guo, R. Mao, and J. Zhang. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of medical virology*, 93(2):820–830, 2021.

[19] J. M. Heffernan, R. J. Smith, and L. M. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.

[20] S. Julier and J. 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.

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

[22] 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.

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

[24] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[25] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582, 2020.

[26] M. Lavielle. *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014.

[27] M. Lavielle, M. Faron, j. lefevre, and J.-D. Zeitoun. Extension of a sir model for modelling the propagation of covid-19 in several countries. *medRxiv*, 2020.

[28] K. Law, A. Stuart, and K. Zygalakis. *Data assimilation: A mathematical introduction*, volume 62 of *Texts in Applied Mathematics*. Springer, Cham, 2015.

[29] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.

[30] 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.

[31] 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.

[32] 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.

[33] A. Oberg and M. Davidian. Estimating data transformations in nonlinear mixed effects models. *Biometrics*, 56(1):65–72, 2000.

[34] 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.

[35] A. Perasso. An introduction to the basic reproduction number in mathematical epidemiology. *ESAIM: Proceedings and Surveys*, 62:123–138, 2018.

[36] 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.

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

[38] D. T. Pham, J. Verron, and L. Gourdeau. Filtres de Kalman singuliers évolutifs pour l'assimilation de données en océanographie. *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, 326(4):255–260, 1998.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] H. Salje, C. T. Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, et al. Estimating the burden of sars-cov-2 in france. *Science*, 369(6500):208–211, 2020.

[45] F. L. Schumacher, C. S. Ferreira, M. O. Prates, A. Lachos, and V. H. Lachos. A robust nonlinear mixed-effects model for covid-19 deaths data, 2020.

[46] D. Simon. *Optimal State Estimation: Kalman, H^∞, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[47] 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.

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

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

[50] 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.

[51] C. Wang, L. Liu, X. Hao, H. Guo, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, A. Pan, et al. Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in wuhan, china. *MedRxiv*, 2020.

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