

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

Annabelle Collin, Mélanie Prague, Philippe Moireau

▶ **To cite this version:**

Annabelle Collin, Mélanie Prague, Philippe Moireau. Estimation for dynamical systems using a population-based Kalman filter – Applications in computational biology. MathematicS In Action, 2022. hal-02869347v5

HAL Id: hal-02869347

https://inria.hal.science/hal-02869347v5

Submitted on 11 Apr 2022 (v5), last revised 10 Oct 2022 (v6)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

April 11, 2022

Abstract

Estimation of dynamical systems - in particular, identification of their parameters - is fundamental in computational biology, e.g., pharmacology, virology, or epidemiology, to reconcile model runs with available measurements. Unfortunately, the mean and variance priorities of the parameters must be chosen very appropriately to balance our distrust of the measurements when the data are sparse or corrupted by noise. Otherwise, the identification procedure fails. One option is to use repeated measurements collected in configurations with common priorities - for example, with multiple subjects in a clinical trial or clusters in an epidemiological investigation. This shared information is beneficial and is typically modeled in statistics using nonlinear mixed-effects models. In this paper, we present a data assimilation method that is compatible with such a mixed-effects strategy without being compromised by the potential curse of dimensionality. We define population-based estimators through maximum likelihood estimation. We then develop an equivalent robust sequential estimator for large populations based on filtering theory that sequentially integrates data. Finally, we limit the computational complexity by defining a reduced-order version of this population-based Kalman filter that clusters subpopulations with common observational backgrounds. The performance of the resulting algorithm is evaluated against classical pharmacokinetics benchmarks. Finally, the versatility of the proposed method is tested in an epidemiological study using real data on the hospitalisation of COVID-19 patients in the regions and departments of France.

1 Introduction

Mathematical models – such as systems of ordinary differential equations (ODE) or even stochastic differential equations (SDE) to account for additional uncertainties – can be used to understand the evolution of biological processes. For example, in pharmacology they can describe the pharmacokinetics of drugs (PK) in plasma [53], or in virology the viral kinetics during HIV treatment [39], or in epidemiology the recent evolution of COVID-19 epidemics [29, 48].

When using these models in practical applications, a major difficulty is dealing with the many uncertain quantities that must be prescribed for the models to be truly predictive. These quantities include the initial conditions or parameters of the model, which are difficult to measure directly. Fortunately, collected data provide additional information that can be used to circumvent the uncertainties associated with the definition of the dynamical system.

Estimation strategies for these models can be viewed as deterministic inverse methods, where we wish to recover the individual parameters that led to particular observations. The solution to this inverse problem then typically leads to the solution of a ODE constrained optimization problem. However, it could also be thought of as a statistical estimation procedure with a variety of techniques based on maximum likelihood optimization, for example, but not limited 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 attractive when they are affordable in terms of computational complexity [19, 49, 4, 2, 30].

Unfortunately, whatever method is chosen, the estimation procedure may not be robust enough in many situations when measurements are too sparse or noisy without relying on strong priors of parameters. One way to compensate for the underlying weakness in observability may be to take advantage of the fact that multiple independent measurements have been collected for a configuration in which the model variables describe a statistical unit of interest that belongs to a larger population in which variability is constrained. This is typically the assumption in nonlinear mixed-effects models [26, 52, 28] –where the use of available population data improves the identifiability of individual parameters through coupled maximization of the likelihood function. Many algorithms and corresponding software have been developed using this principle. Among others, we can mention the First Order Conditional Estimation (FOCE) algorithm implemented in `NONMEM` [43], the Stochastic Approximation of Expectation-Maximization (SAEM) algorithm proposed in `MONOLIX` [25], penalized maximum likelihood approaches available in `NIMROD` [46], or even full Markov chain Monte Carlo (MCMC) methods implemented in `Winbugs` [54], `jags` [14], or `Stan` [7]. Most of these estimation algorithms and software have been compared in previous work and have been shown to achieve similar performance results for identifiable models of intermediate size, such as models in pharmacokinetics [16, 44, 46, 32]. However, increasing the size of the underlying dynamics without increasing the complexity cost of the algorithm too much remains a challenge.

In this work, we propose an alternative strategy inspired by the data assimilation method generally developed for environmental sciences [2] and often confronted with the complexity burden associated with large-scale systems. Our strategy is to capture all population information in a unified maximum likelihood estimation procedure. Assuming Gaussian disturbances up to a change of variables, we approximate the corresponding optimization problem with a Kalman-based filter. More specifically, we choose the Unscented Kalman Filter [22], but other Gaussian filters [2] could also lead to similar results. The potential curse of dimensionality caused by the population framework is then limited by covariance reduction techniques such as those proposed in the Reduced-Order Unscented Kalman Filter [33]. Note that this is not the first time that Kalman-based approaches have been used in mixed-effects strategies. This is the case in [36, 50, 24] where a specific combination of the FOCE method with the extended Kalman filter was proposed ; in [12] where a coupling of the SAEM algorithm with the extended Kalman filter is performed for ODE models and in [36] for SDE models. However, in this literature, the extended Kalman filters are used only to approximate the individual probability distribution function, whereas in this paper we use the Kalman approach also at the population level.

To evaluate our approach, we propose a series of numerical tests starting from a classical toy problem in pharmacokinetics. First, a classical synthetic data benchmark is considered to validate our implementation. Then, a validation is performed on a real data set to evaluate the performance of the algorithm against the literature. We found that the pandemic COVID-19 represents a very general use case when trying to aggregate multiple data collections from different locations in space using SIR -based models [37]. More specifically, we assume that the data collected in France at the regional (or departmental) level are independent realizations of the same underlying epidemic of COVID-19 driven by random parameters specific to each region (or department), which requires an *population approach* to estimate them. This numerical experiment is a real challenge for our algorithm and illustrates its versatility.

The paper is organized as follows. Section 2 formalizes the estimation problem and introduces the notation. Section 3 introduces the population-based Kalman filter and its reduced-order version. In section 4, we provide numerical experiments and an evaluation of the performances before a final conclusion is drawn in section 5.

2 Problem statement

2.1 Model formulation

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

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

The parameters of interest $\theta^i \in \mathcal{P} \simeq \mathbb{R}^{N_\theta}$ can be treated as state variables by expanding the state dimension and remembering that $\dot{\theta}^i = 0$. We therefore 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

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

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

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

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

In practice, we will ultimately rely on a discretization of dynamics (1). We restrict ourselves here to a forward Euler time scheme, but it is quite easy to consider more general time schemes. We should solve numerically for the choice of a time-step $\delta T_n = T_{n+1} - T_n$ that is small enough to satisfy the conditional stability constraint

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

with $T_n = \sum_{i=0}^n \delta T_i$, $T_{N_T} = T$. In the remainder of the paper, which focuses on the discrete-time formulation, we assume that there exists a discrete transition operator $\phi_{n+1|n}$ that yields the state x_{n+1}^i in knowledge of the state x_n^i, namely

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

where here $\phi_{n+1|n}(x_n^i, \theta^i) = (x_n^i + \delta T_n f(x_n^i, \theta^i, T_n))$. Note that more complex time schemes will lead to different definitions of the transition operator, but (5) is a fairly general discrete-time dynamics. As shown before, we define from $\phi_{n+1|n}$ a joint state- parameter transition operator $\psi_{n+1|n}$ such that

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

2.2 Uncertainties modeling

In the spirit of [30, Chapter 2], we introduce some stochasticity into the discrete-time formulation (3) and consequently into (4). In fact, we complete (3) by introducing an additive modeling error contribution with the introduction of an operator $G \in \mathcal{L}(\mathbb{R}^{N_x}, \mathbb{R}^{N_\nu})$ and for all $1 \leq i \leq N_P$, $\nu_n^i \sim_{\text{i.i.d}} \mathcal{N}(0, Q_n^i)$ such that $x^i = (x_n^i)_{n \in \mathbb{N}}$ is now a Markov chain defined by the random map

$$x_{n+1}^i = \phi_{n+1|n}(x_n^i, \theta^i) + G_n \nu_n^i, \quad 0 \leq n < N_T - 1, \quad 1 \leq i \leq N_P. \tag{5}$$

Note that we typically have the following scaling $|||G_n||| \propto \delta T_n$ due to discretization and $\mathrm{Tr}(W_k^i) \propto \frac{1}{\delta T_n}$ [17] to match a potential continuous-time limit. Consequently, upon completion of (4), there exists a joint operator B_n such that for the joint state-parameter sequence random map we have

$$\begin{cases} z_{n+1}^i = \psi_{n+1|n}(z_n^i) + B_n \nu_n^i, \quad 0 \leq n < N_T - 1, \quad 1 \leq i \leq N_P. \\ z_0^i = z_0 + \xi^i \end{cases} \tag{6}$$

In addition to the stochasticity introduced in the dynamics, we assume that the initial state and the parameters of interest are subject to some stochasticity in the population under consideration. 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. A population-based approach typically assumes that the $N_P$members of the population – called subjects in this paper – are drawn at random from a same distribution, in this case 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 and P_0 is the associated covariance. We equivalently write

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

In mixed effects models, ξ^{pop} is called the population intercept, while $\tilde{\xi}^i$ is the random effect. Normally, ξ^{pop} is not known, nor is P_0. However, we can assume that ξ^{pop} is bounded as a deterministic quantity with respect to a norm defined by a given positive matrix M, which is assumed to be known. Moreover, we have at our disposal some a priori $\hat{P}_0$ for P_0.

2.3 Observations

Our goal is to estimate the unknown parts of the initial conditions in our population using the available data. In practice, we do not observe x^i directly, but we do have 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 by the measurement process. This process is modeled by introducing an observation operator h_k that depends on time $t_k \in [0, T]$ such 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 covariance matrix of the observation errors. Note that to define a discrete-time white noise that could be consistent with a potential continuous-time limit, we have the following scaling $\text{Tr}(W_k^i) \propto \frac{1}{\Delta t_k^i}$ with $\Delta t_k^i = t_{k+1}^i - t_k^i$.

For simplicity, in 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 account for subject-dependent sampling.

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

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

We note that for simplicity of exposition, the model discretization and the observation discretization are consistent, namely $\max_i(\delta t_i)$ is small enough, such that $N_T \geq N_{\text{obs}}$ and there exists a sequence of integers $(n_k)_{0 \leq k \leq N_{\text{obs}}}$ such that for all observation times t_k, $1 \leq k \leq N_{\text{obs}}$, there exists $T_{n_k} = t_k$.

Note that when we combine the augmented dynamics and the observations, we find that there is a nonlinear mapping Γ such that

$$y_k^i = \Gamma(\xi^{\text{pop}}, \xi^i, (\nu_n^i)_{0 \leq n < N_T - 1},\, t_k) + \chi_k^i, \quad 1 \leq i \leq N_P, \quad 0 \leq k \leq N_{T_{\text{obs}}} \tag{7}$$

where χ_k^i is Gaussian noise, ξ^{pop} is 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 [28]. And even if the model dynamics are linear and the observation operator is linear, the dependence of the dynamics on the parameters ultimately leads to Γ being a nonlinear operator with respect to z_0^{pop} and ξ^i.

Thus, to summarize the problem, our goal is 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 (7), benefiting - from a data assimilation point of view - from the dynamic 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 $N_z \times N_P$

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

Similarly, we denote $(\boldsymbol{\nu}_n)_{0\leq n\leq N_T-1} = ((\nu_n^1 \cdots \nu_n^{N_P})^\intercal)_{0\leq n\leq N_T-1}$. Throughout 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

$$\min_{\substack{\xi^{\text{pop}}\in\mathcal{X},\boldsymbol{\xi}\in(\mathcal{X}\times\mathcal{P})^{N_P},\\ (\boldsymbol{\nu}_n)_{0\leq n\leq N_T-1}\in\mathcal{Q}^{N_P\times N_T}}} \Bigg\{ \mathscr{L}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\langle \xi^{\text{pop}}, M\xi^{\text{pop}}\rangle + \sum_{i=1}^{N_P}\Bigg[\frac{1}{2}\big\langle (\xi^i-\xi^{\text{pop}}), \hat{P}_0^{-1}\,(\xi^i-\xi^{\text{pop}})\big\rangle + \frac{1}{2}\sum_{n=0}^{N_T-1}\big\langle \nu_n^i, (Q_n^i)^{-1}\nu_n^i\big\rangle + \frac{1}{2}\sum_{k=0}^{N_{T_{\text{obs}}}}\big\langle (y_k^i - c_k^i(z_{n_k}^i)), (W_k^i)^{-1}(y_k^i-c_k^i(z_{n_k}^i)\big\rangle\Bigg]\Bigg\}, \quad (8)$$

where all $(z_n^i)_{0\leq n\leq N_T}$, $1\leq i\leq N_P$, are forced to follow the dynamics (6). Here, this minimization includes the population intercept ξ^{pop}. To simplify the minimization problem, we note that the population intercept represents the expected mean of all the subject uncertainties. We therefore propose to model ξ^{pop} as an empirical mean over all subjects namely

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

Therefore, the minimization problem becomes

$$\min_{\substack{\boldsymbol{\xi}\in(\mathcal{X}\times\mathcal{P})^{N_P},\\ (\boldsymbol{\nu}_n)_{0\leq n\leq N_T-1}\in\mathcal{Q}^{N_P\times N_T}}} \Bigg\{ \mathscr{L}_T(\boldsymbol{\xi},\boldsymbol{\nu}) = \frac{1}{2}\big\langle \mathrm{E}_{N_P}(\boldsymbol{\xi}), M\mathrm{E}_{N_P}(\boldsymbol{\xi})\big\rangle + \sum_{i=1}^{N_P}\Bigg[\frac{1}{2}\big\langle (\xi^i-\mathrm{E}_{N_P}(\boldsymbol{\xi})), \hat{P}_0^{-1}\,(\xi^i-\mathrm{E}_{N_P}(\boldsymbol{\xi}))\big\rangle + \frac{1}{2}\sum_{n=0}^{N_T-1}\big\langle \nu_n^i, (Q_n^i)^{-1}\nu_n^i\big\rangle + \frac{1}{2}\sum_{k=0}^{N_{T_{\text{obs}}}}\big\langle (y_k^i - c_k^i(z_{n_k}^i)), (W_k^i)^{-1}(y_k^i-c_k^i(z_{n_k}^i)\big\rangle\Bigg]\Bigg\}, \quad (9)$$

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 (9) depends on fewer unknowns than in (8), which means that our coupling reduces the minimization space, which is already quite 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 to

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

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 indices in $[(i-1)\,N_z+1, i\,N_z]\times\Big[(j-1)\,N_z+1, j\,N_z\Big]$ is given by

$$(\boldsymbol{M}_0)_{i,j} = \frac{1}{{N_P}^2}M + \Big(\delta_{ij}-\frac{1}{N_P}\Big)\hat{P}_0^{-1}, \quad (10)$$

where δ_{ij} is the Kronecker delta equal to 1 when $i=j$ and 0 otherwise. 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

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

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, $(\mathbb{I}_{N_P} - \frac{1}{N_P}\vec{1}_{N_P}\vec{1}_{N_P}^\intercal)\otimes\hat{P}_0^{-1}$, is definite on $\mathrm{Ker}(\vec{1}_{N_P}\vec{1}_{N_P}^\intercal)^\perp\otimes\mathbb{R}^{N_x}$, while 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 $\mathrm{Tr}(M)$ – and therefore can be interpreted as a regularization term with respect to the second matrix. Indeed, if $\mathrm{Tr}(M)$ is small with respect to $Tr(\hat{P}_0^{-1})$, our criterion means that the estimation is strongly supported with a shared population variable ξ^{pop} and small variations $\tilde{\xi}^i$ around it. Conversely, $\mathrm{Tr}(M)$that is 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, the introduction of

$$\hat{\boldsymbol{P}}_0 = \boldsymbol{M}_0^{-1} = \left(\frac{1}{{N_P}^2}\vec{1}_{N_P}\vec{1}_{N_P}^{\intercal} \otimes M + \left[\mathbb{I}_{N_P} - \frac{1}{N_P}\vec{1}_{N_P}\vec{1}_{N_P}^{\intercal}\right] \otimes \hat{P}_0^{-1}\right)^{-1}, \tag{11}$$

our estimator is defined as the minimizer of the least-squares objective function (9), which is rewritten into the general form

$$\mathscr{L}_T(\boldsymbol{\xi}, \boldsymbol{\nu}) = \frac{1}{2}\langle \boldsymbol{\xi}, \hat{\boldsymbol{P}}_0^{-1}\,\boldsymbol{\xi}\rangle + \frac{1}{2}\sum_{n=0}^{N_T-1}\left\langle \boldsymbol{\nu}_n, \boldsymbol{Q}_n^{-1}\boldsymbol{\nu}_n\right\rangle + \frac{1}{2}\sum_{k=0}^{N_{T_{\mathrm{obs}}}}\left\langle \boldsymbol{y}_k - \boldsymbol{c}_k(\boldsymbol{z}_k), \boldsymbol{W}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{c}_k(\boldsymbol{z}_{n_k}))\right\rangle. \tag{12}$$

We have gathered all the individual states in $\boldsymbol{z} = (z^1 \cdots z^{N_P})^{\intercal}$, which must follow the discrete-time "populations" dynamic

$$\begin{cases} \boldsymbol{z}_{n+1} = \boldsymbol{\psi}_{n+1|n}(\boldsymbol{z}_n) + \boldsymbol{B}_n\boldsymbol{\nu}_n, \quad \forall t \in [0,T] \\ \boldsymbol{z}_0 = \vec{1}_{N_P} \otimes z_0 + \boldsymbol{\xi}, \end{cases} \tag{13}$$

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 in: $\boldsymbol{y} = (y^1 \cdots y^{N_P})^{\intercal}$ so that

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

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 the covariance

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

Minimizing (12) under the constraint imposed by the dynamics (13) becomes a classical data assimilation problem [2, 30]. The key to our uncertainty modeling is that $\hat{\boldsymbol{P}}_0$ – constructed from (10) – couples the population members, since in fact

$$\hat{\boldsymbol{P}}_0 \neq \hat{\boldsymbol{P}}_0^{\mathrm{un}} \stackrel{\mathrm{def}}{=} \mathbb{I}_{N_P} \otimes \hat{P}_0. \tag{14}$$

In the upcoming numerical section, the effects 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

Minimization of the functional (12) under the constraint of dynamics of the form (13) can be naturally performed using variational data assimilation methods [5, 9, 2]. We denote by $\bar{\boldsymbol{z}}_{n|N_T}$, $0 \le n \le N_T$, 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 for the case where all operators are "linear"– namely for all n, where $\boldsymbol{\psi}_{n+1|n}$, $\boldsymbol{c}_n$ and $\boldsymbol{B}_n$ are linear [23, 49, 4, 30]. Given a nonlinear model operator

or a nonlinear observation operator, it is classical to rely on an approximate optimal sequential estimator based on the generalization of the Kalman filter to nonlinear operators [49, 30]. Here, we have chosen to use the Unscented Kalman Filter (UKF), where the mean and covariance operators are computed from the empirical mean and empirical covariance based on N_σ sample points – the so-called sigma points [22]. We define for a given set of positive coefficients $\boldsymbol{\alpha} = (\alpha_j)_{1\leq j\leq N_\sigma}$ and a given set of particles $(v^{(j)})_{1\leq j\leq N_\sigma}$, the empirical mean and covariance by

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

Remark 1 *Different choices of sigma-points can be used in practice from canonical sigma-points ($N_\sigma = 2N_P \times N_z$) which are aligned with the canonical base to star sigma-points ($N_\sigma = 2N_P \times N_z + 1$) in which the origin is added to the canonical points or simplex sigma-points, see [22, 34] for more details. Here in the numerical part, we will use canonical sigma-points.*

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

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

where δ is a scaling factor. The UKF algorithm reads

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

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 [19, 49, 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

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

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

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

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 ${N_P}^{-1}\sum_{i=1}^{N_P}(\hat{\boldsymbol{P}}^{+}_{N_T})^{\theta}_{i,i}$ 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 $({N_P}^{-2}\sum_{i,j=1}^{N_P}(\hat{\boldsymbol{P}}^{+}_{N_T})^{\theta}_{i,j})^{-1}$ 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 2 (Initial condition estimation through smoothing) *One classical drawback of Kalman estimation is that a quantity of the form* $\mathrm{E}(\boldsymbol{z}_0|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n)$ *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* $\mathrm{E}(x_0^i|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n)$ *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 [49]. 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* $1\le n\le N_T$, $z_n^i=(x_n^i\ \theta^i\ x_0^i)^{\intercal}$*. 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* $\mathrm{E}(x_0^i|\boldsymbol{y}_0,\cdots,\boldsymbol{y}_n)$.

Remark 3 (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 [49, Section 7.4]. In this respect, we introduce a forgetting factor* $0\le\rho\le 1$*, and modify the covariance prediction into*

$$\hat{\boldsymbol{P}}_{n+1}^{-}=\frac{1}{\rho}\mathrm{Cov}_{\boldsymbol{\alpha},N_\sigma}\left(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-}\right)+\boldsymbol{B}_n\boldsymbol{Q}_n\boldsymbol{B}_n^{\intercal},\ 0\le n\le N_T. \tag{16}$$

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 4 (Non-linear transformation of state variables) *In many cases, the state variable* x_n^i *modeled for each subject i in iteration n does not follow a Gaussian distribution. However, they can be modeled as a nonlinear transformation of a Gaussian variable, as is often proposed in nonlinear mixed effects models [35, 28]. This is the case when you consider a variable that takes non-negative values and can be modeled as a Gaussian variable after being transformed by a log transformation. Similarly, if you consider a variable that takes bounded values and can be modeled as a Gaussian variable after transformation by a logit transformation. Then it is not difficult to extend the UKF filter to manage variables that follow a Gaussian distribution up to a certain transformation. By introducing the general mapping* $\boldsymbol{\varphi}$ *that transforms the augmented state into Gaussian variables, we modify the classical UKF*

algorithm in

$$
\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\big(\boldsymbol{\varphi}^{-1}(\boldsymbol{\zeta}_n^{(j)-})\big) - \boldsymbol{y}_n, & 0 \leq n \leq N_T, 1 \leq j \leq N_\sigma, \\
\boldsymbol{K}_n = \mathrm{Cov}_{\boldsymbol{\alpha},N_\sigma}\big(\boldsymbol{\zeta}_n^{(\cdot)-}, \tilde{\boldsymbol{y}}_n^{(\cdot)}\big)[\mathrm{Cov}_{\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^+ = \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, \\
\hat{\boldsymbol{P}}_n^+ = \hat{\boldsymbol{P}}_n^- - \boldsymbol{K}_n[\mathrm{Cov}_{\boldsymbol{\alpha},N_\sigma}\big(\tilde{\boldsymbol{y}}_n^{(\cdot)}\big) + \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}\big(\boldsymbol{\varphi}^{-1}(\boldsymbol{\zeta}_n^{(j)+})\big), & 0 \leq n \leq N_T - 1, 1 \leq j \leq N_\sigma, \\
\hat{\boldsymbol{z}}_{n+1}^- = \boldsymbol{\varphi}^{-1}\Big(\mathrm{E}_{\boldsymbol{\alpha},N_\sigma}\big(\boldsymbol{\varphi}(\hat{\boldsymbol{z}}_{n+1}^{(\cdot)-})\big)\Big), & 0 \leq n \leq N_T - 1, \\
\hat{\boldsymbol{P}}_{n+1}^- = \frac{1}{\rho}\mathrm{Cov}_{\boldsymbol{\alpha},N_\sigma}\big(\boldsymbol{\varphi}(\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{17}
$$

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

Remark 5 (Iterative Kalman procedure) *In the presence of strong nonlinearity, we know that the Kalman filter is only an approximation to the MLE, which is often overcome in the literature by requiring a small number of iterations of the Kalman filter [19, Section 6.1]. In a case where we estimate the parameters and initial condition by a smoothing procedure, an iterative procedure can be easily set up by considering multiple loops where the new* a-priori *comes from the final* a-posteriori *of the previous estimation. Such a procedure is very similar to the Bayesian approach described in [45].*

Remark 6 (Alternative filtering strategies) *We rely here on the Unscented Kalman Filter, but we could also have proposed an equivalent strategy using an Extended Kalman Filter (EKF). The advantage of the UKF in this work was the ease of integration and implementation of the nonlinear transformation of variables, since UKF can be seen as a derivative-free alternative to EKF [2]. In addition, we could have used an Ensemble Kalman Filter [18], which would have provided greater sampling latitude without going up to the complexity of a fully justified particle filter. Our choice was in fact guided by the convenience of having at our disposal a reduced-order version of UKF that will further limit the computational complexity of our strategy as described in the next section.*

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 [42, 40, 34] is to be able to maintain covariance matrices in their factorized form

$$
\hat{\boldsymbol{P}}_n^+ = \boldsymbol{L}_n(\boldsymbol{U}_n^+)^{-1}\boldsymbol{L}_n^\intercal \text{ follow by } \hat{\boldsymbol{P}}_{n+1}^- = \boldsymbol{L}_{n+1}(\boldsymbol{U}_{n+1}^-)^{-1}\boldsymbol{L}_{n+1}^\intercal \tag{18}
$$

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 [34]. 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 7 – 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

$$(\boldsymbol{U}_0^{-1})_{r,s} = \frac{1}{N_{c_r} N_{c_s}} \sum_{i \in c_r} \sum_{j \in c_s} (\hat{\boldsymbol{P}}_0)_{i,j}. \tag{19}$$

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

$$(\boldsymbol{L}_0)_{i,r} = \begin{cases} \mathbb{I}_{N_z}, & \text{if } i \in c_r, \\ \not\vdash_{N_z}, & \text{otherwise.} \end{cases} \tag{20}$$

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 (20) by

$$(\boldsymbol{L}_0)_{i,r} = \beta_{i,r} \mathbb{I}_{N_z}. \tag{21}$$

Remark 7 (Other reduction techniques) *We propose here a practical definition of $\boldsymbol{L}_0$, $\boldsymbol{U}_0$ based on the* k-means*, but alternative strategies could have been followed, for example hierarchical clustering or principal component analysis of the $\mathbb{Y}\mathbb{Y}^\intercal$ as performed in the Proper-Orthogonal-Decomposition approach presented in [10].*

We conclude this section by presenting our reduced estimator algorithm based on the reduced-order Unscented-Kalman Filter (RoUKF) [34]. We emphasize that in this algorithm the number of sigma-points $N_\sigma^r \ll N_\sigma$ used in the full UKF approach (15). Therefore, the associated unitary sigma-points $\boldsymbol{e}_r^{(j)}$ generate a subspace of small dimension N_σ^r. Moreover, here we here complement [34] by several features. First, we integrate the nonlinear transformation introduced in Remark 4. Second, we add the impact of the model error operator, which must be projected into the reduced subspace generated by the column of $\boldsymbol{L}_n$ to preserve the decomposition (18) over time iteration. In this respect, we follow [41, 40] 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 introduced in the UKF algorithm is preserved in the RoUKF algorithm by analogy

with [41, 40]. We finally obtain

$$
\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 = \mathrm{Cov}_{\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(\mathrm{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} = \mathrm{Cov}_{\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 \mathrm{Cov}_{\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{22}
$$

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 8 (Fisher information matrix) *An additional advantage of the reduced-order formulation is that it is based on a square-root formulation of Kalman filtering, which allows easy computation of the 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

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

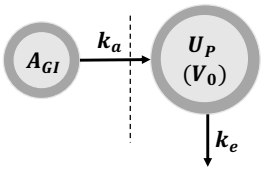
which is an estimation of the Fisher information matrix as computed in other Unscented Kalman Filter works [6].

4 Numerical results

In this section, we validate and illustrate our population-based Kalman filter with examples of different nature. In Section 4.1, we consider a toy problem corresponding to a one-compartment pharmacokinetic model with first-order absorption and elimination. First, we validate our strategy on synthetic data by evaluating the performance of Algorithm 15 using the coupled covariance matrix $\hat{\boldsymbol{P}}_0$. We then consider the performance of Algorithm (22) 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 real data estimation.

In Section 4.2, we consider a more complex configuration from a real estimation problem of the COVID-19 epidemic early evolution in 2 geographic resolutions – in regions ($N_P = 12$) and in departments ($N_P = 94$) – of the non-island French metropolitan area. The objective of the estimation is to reconstruct the impact of lockdown health policies on the transmission rate in the first months of virus spread using an extended model of type SIR [37] (from March 1$^{\text{st}}$ and May 11$^{\text{th}}$, 2020). We find this structure emblematic, as it requires the use of all the components proposed in our methodology.

(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)

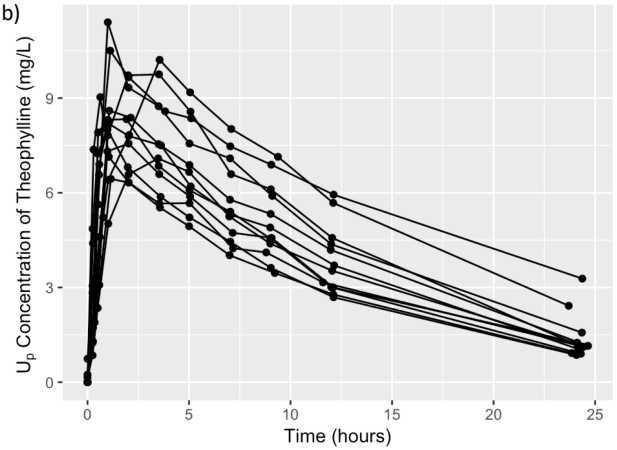


Figure 1: A one-compartment model PK with oral absorption: (a) illustration of the mathematical model, (b) a benchmark data set of the anti-asthmatic drug theophylline for 12 patients

.

4.1 A Pharmacokinetics toy problem

4.1.1 Model and observations

We consider a one-compartment pharmacokinetic 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 plasma (in mg). The interactions between A_{GI} and A_P are described by the following system

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}A_{GI} = -k_a A_{GI}, \\ \dfrac{\mathrm{d}}{\mathrm{d}t}A_P \ = k_a A_{GI} - k_e A_P, \end{cases} \tag{23}$$

where k_a corresponds to the absorption rate (in h^{-1}) and k_e corresponds to the elimination rate (in h^{-1}). In clinical practice, it is assumed that only the plasma concentration $U_P = \frac{A_P}{V_0}$ (in $\mathrm{mg.L}^{-1}$) is observed at the $N_{T_{\mathrm{obs}}}$ observation times, where V_0 is the volume of the plasma compartment (in L). System (23) can be rewritten as follows

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}A_{GI} = -k_a A_{GI}, \\ \dfrac{\mathrm{d}}{\mathrm{d}t}U_P \ = \dfrac{k_a}{V_0}A_{GI} - k_e U_P, \end{cases} \tag{24}$$

see Figure 1(a) for an illustration.

Remark 9 *Obviously, the analytical solution of System* (24) *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, since 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).

As for the initial values, $A_{GI}(0)$ corresponds to a known initial dose 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 from available observations. These parameters are all positive and are constrained by a log transformation as the estimation proceeds. Following the notations in 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 thus based on a fully deterministic model, as is often the case in data assimilation [30, Section 2.1]. As for the measurements, the observation operator is given by

$$c_k^i(z^i(t_k^i)) = U_P^i(t_k)^{0.25}, \quad 1 \le i \le N_P,\ 0 \le k \le N_{T_{\text{obs}}}. \tag{25}$$

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

Another important aspect is the ability to handle the time sampling of the data, which is imposed in practice and can be very different from the time step of the model. In fact, the time sampling of the data may be different for each patient. To deal with the time sampling of the data, we can think of two strategies: We can either use the data only when they are available, or we can rely on time interpolation. To simplify our implementation, we chose to interpolate the observations. More specifically, using the time discretization notations of the system introduced in Section 3, we first reconstruct the measurements for $1 \le i \le N_P, 0 \le n < N_T$

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

These measurements are consistent with an observation operator c_k^n of the form (25), with additional rescaling of $W_n^i \propto \frac{t_{k+1}-t_k}{t_{n+1}-t_n} w$. Finally, we set $\rho = 1$, $\delta = 1$ in our algorithms.

4.1.2 Validation using synthetic data

First, we will evaluate the performance of Algorithm 15 using the coupled covariance matrix $\hat{\boldsymbol{P}}_0$ (see Eq. (11)) (population estimation) compared to using the diagonal covariance matrix $\hat{\boldsymbol{P}}_0^{\text{un}}$, see Eq. (14) (individual estimation). We then evaluate the performance of the reduced approach (22) 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

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

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

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

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

To evaluate the performance of the estimation, several statistical validation criteria can be considered. For a parameter θ, we define by θ_t (respectively, θ_e) the matrix of size $N_R \times N_P$ that contains in each row r the target (respectively, estimated) values of θ for all patients of replicate r. We also calculate 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 square error MSE

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

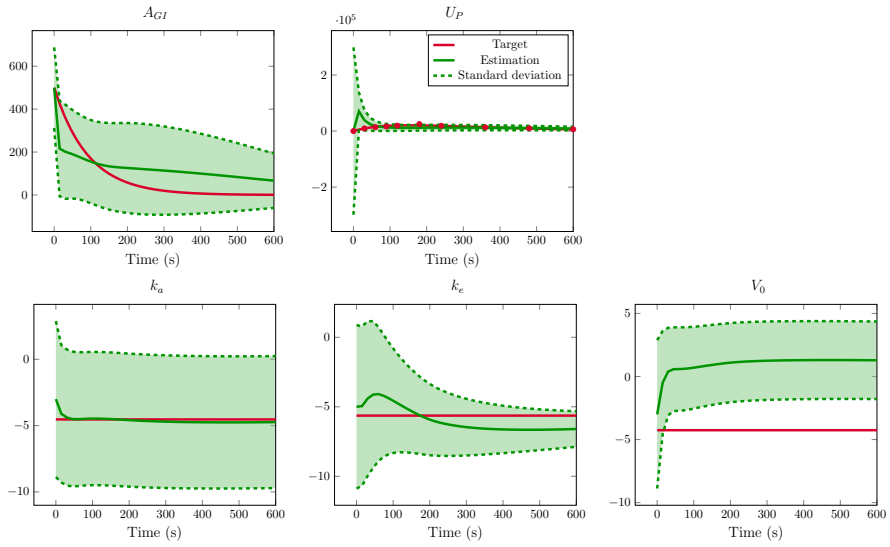


Figure 2: Time evolution of states A_{GI} and U_P and parameters k_a, k_e and V_0 for a patient using an individual estimation with N_P=20 and $w_t = 0.3$. The target states and target values are shown in red. The red points in the graph of U_P correspond to the noisy observed data. The states and parameter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

The RBIAS describes the deviation of the mean across the estimated parameters from their true values. The MSE summarizes both the bias and the variability of the estimates. Using $\hat{P}_n^i$, we compute σ_e^θ the matrix of size $N_R \times N_P$ that contains in each row r the estimated standard deviation of θ for all patients of replicate r. We also supplement the RBIAS and MSE with the mean estimated standard deviation (STD)

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

and the empirical standard deviation of estimates, defined as the standard deviation of 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 calculate the relative bias of the standard deviation for random effects, called BMIXED, defined by

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

where $\sigma_{e,m}^\theta$ is the matrix of size $N_R \times N_P$ containing in each row r the estimated mixed standard deviation of θ for all patients of replicate r and therefore formed using $\tilde{\hat{P}}_n^i$. This indicator is completed by coverage (COV), which is defined as the percentage of time that the true value is contained within the 95% confidence interval of the estimated 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 readability, the statistical criteria are aggregated over the 3 parameters. As expected when using an individual approach, the estimates are similar for all sample sizes in terms of precision and accuracy. We note that the individual Kalman filter does not estimate the parameters correctly, as underscored in Figure 2, in which the time evolution of the state and parameters is plotted for a patient. Indeed, the poor estimation performance is mainly due to the poor parameters *a-priori*, which are not compensated by the individual approach.

In contrast, as shown in Table 1, the population approach really strengthens the estimation because the RBIAS and MSE decrease with sample size and are dramatically smaller than for the individual approach (for $N = 100$, the RBIAIS goes from 0.374 for the individual approach to -0.007 for the

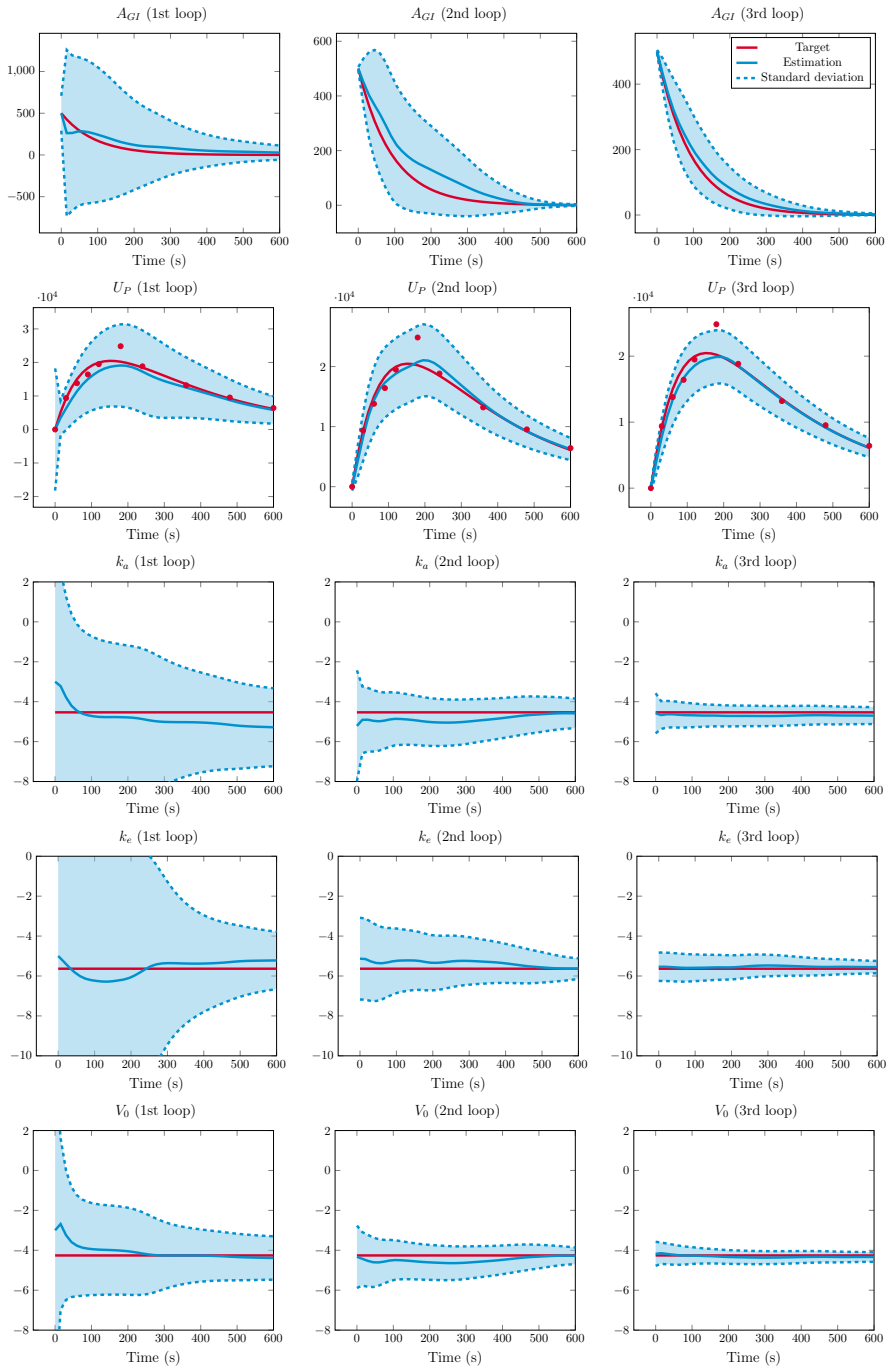


Figure 3: Time evolution of states A_{GI} and U_P and parameters k_a, k_e and V_0 for a patient using a population estimation with N_P=20 and $w_t = 0.3$. Each column corresponds to one loop of the algorithm. The target states and target values of the parameters are shown in red. The red dots in the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

Meas. Error	Sample Size	Individual RBIAS	Individual MSE	Individual STD	Individual ESTD	Population RBIAS	Population MSE	Population STD	Population 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 were evaluated over 100 replicates in 6 different scenarios. Summary statistics are relative bias (RBIAS), mean squared error (MSE), mean estimated standard deviation (STD), and empirical standard deviation of estimates (ESTD). Results are aggregated across all parameters of the model.

		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.

population approach). This is also confirmed by the resulting individual trajectories as shown in Figure 3 (first column) showing the time evolution of a patient's state and parameters. We note that the STD is significantly larger than the ESTD. It indicates an overestimation of the standard deviation of the estimates due to the fact that the initial covariances of the estimates were chosen very roughly (3^2 instead of $0.2^2, 0.25^2$, and 0.1^2 for k_a, k_e, and V_0, respectively). Then, to obtain a more accurate covariance for the final estimate, we iterate the estimation procedure according to remark 5. Each run is referred to as a loop. The population procedure is evaluated for $w_t = 0.3$ and N_P=20. Table 2 summarizes the results for each parameter of the model. As expected, the RBIAS and MSE decrease when multiple loops are used. The second (respectively third) column of Figure 3 shows the time evolution of the states and parameters for a patient for the second (respectively third) loop. In addition, STD get closer to the ESTD and COV approaches its nominal value of 95%. Indeed, we naturally recommend performing multiple estimation loops when starting from poor covariance *a-priori*. The estimation of the standard error of the random effect is also less biased with 3 loops, as BMIXED shows, but still important. Perhaps this can be explained by the fact that we do not minimize the full functional (8), which includes the population intercept, but the functional (9), which assumes that the population intercept can be considered an empirical mean over all patients. It would be a perspective for this work to compare our results with 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. This is why the reduced-order version of the population Kalman filter that we will present here is so interesting. 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 plotting the observed variable $U_P^{0.25}$ for a replicate. Table 3 presents the results given for each parameter and also aggregated across the 3 parameters. These estimates are supplemented by computation times obtained using a 2.3 GHz quad-core Intel Core i7 computer and an implementation in `MATLAB-2020-b`. This gives an order of magnitude numerical complexity of the algorithm, but with a not fully optimized implementation. Focusing on the MSE, we obtain equivalent results for the full covariance approach and the reduced covariance one. As for the RBIAS, the results obtained with the reduced version remain acceptable. As expected, both standard deviation criteria (STD and ESTD)

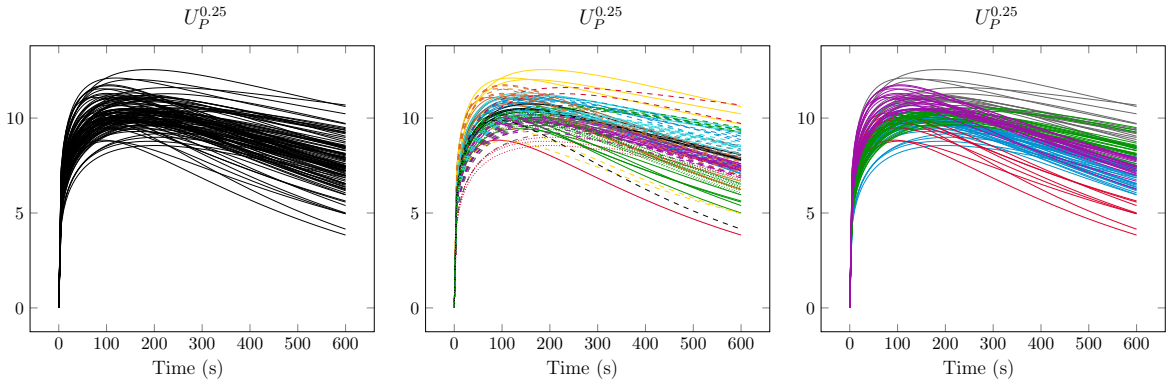


Figure 4: Illustration of the clustering algorithm (*k-means* algorithm) for the observed variable $U_P^{0.25}$ for a synthetic data set of 100 patients. Left: complete data set. Middle: results of the *k-means* algorithm with 20 clusters. Right: results of the *k-means* algorithm with 5 clusters. Each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

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

Table 3: Performances of the population estimation algorithms (full and reduced versions) evaluated over 100 replicates for N_P=100 and $\sigma_m = 0.3$.

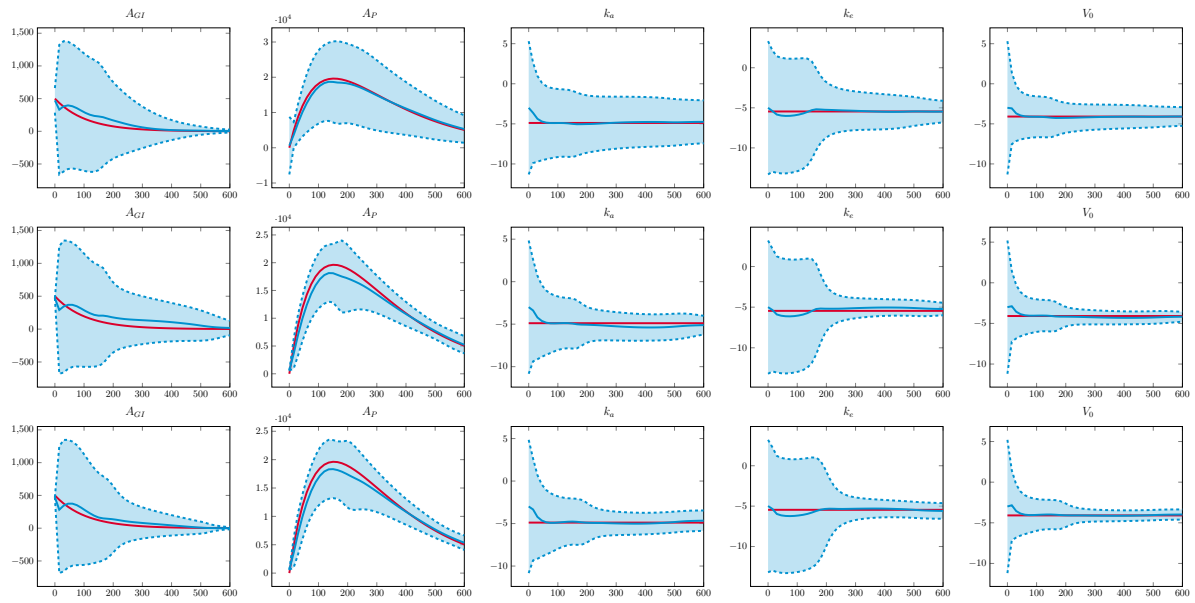


Figure 5: Time evolution of states A_{GI} and U_P and parameters k_a, k_e, and V_0 for a patient using the full version of the algorithm (top) and reduced-order versions (middle: $N_C = 20$, bottom: $N_C = 5$) for N_P=100 and $\sigma_m = 0.3$. The target states and parameters are shown in red. The estimated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

decrease with the reduced-order filter, but their values are also acceptable. As for the ESTD, it is even better. By using the reduced-order filter, the COV approaches its nominal value of 95% and finally the BMIXED also improves. These 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 must be chosen carefully so as not to constrain the system too much. Figure 5 shows the time evolutions of the states and the parameters of a 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 show similar behavior in the four cases. This means that the system is not too constrained even if the COV for V_0 is low (in the acceptable range) - which is known to be the most difficult parameter of the system to estimate. Figure 5 shows that increasing the coupling by reducing the number of clusters reduces the standard deviations, eliminating the need for multiple loops. Finally, in terms of computation times, these can be drastically reduced with the reduced-order version of the population Kalman filter: for example, by a factor of 7 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 a one-compartment pharmacokinetic model with first-order absorption and elimination using the theophylline data set: Individual nonlinear estimation and individual UKF, and population nonlinear estimation and population UKF. Estimates and estimated standard deviations are reported.

To conclude this study on a toy problem, we propose validation against a real data set. Theophylline is a methylxanthine drug used in the treatment of respiratory diseases such as chronic obstructive pulmonary disease and asthma under various brand names. Data were collected by Upton et al. on 12 subjects who received a single oral dose of theophylline and then gave 11 blood samples over a 25-hour period [51]. The 12 subjects received $A_{GI}(0) = 320$mg of theophylline. A spaghetti plot showing the plasma concentration of the drug with time is available in Figure 1-(b). Considering the small number of patients, only the filter for the whole population is tested. For illustration, it is possible to compare our results with a simple nonlinear mixed effects (NLME) approach (based on the analytic solution in Remark 9).

The individual nonlinear estimation is implemented with the `R`-package *nls* and the population NLME model is implemented with the `R`-package *saemix*. [11]. Our population-based Kalman filter yields very similar results to the population-based NLME, see Table 4. It is also interesting to note that the individual Kalman filter gives better results than the individual nonlinear strategy, especially in estimating V_0, the most difficult parameter to estimate, since it determines the slope of U_p and is very sensitive to observational 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 [37] inferred with hospitalization data. The objective of the estimation is to derive the impact of the lockdown on the transmission rate between March 2, 2020 and May 11, 2020.

4.2.1 SEIRAH Model

More specifically, we use a SEIRAH model adapted from [37] in which 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 (ie, both recovered and deceased). The dynamics of such a model is

given by

$$\begin{cases} \dfrac{\mathrm{d}S}{\mathrm{d}t} = -\dfrac{b\,S(I+\alpha A)}{N} \\ \dfrac{\mathrm{d}E}{\mathrm{d}t} = \dfrac{b\,S(I+\alpha A)}{N} - \dfrac{E}{D_E} \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \dfrac{r_E}{D_E}E - \dfrac{1-r_I}{D_q}I - \dfrac{r_I}{D_I}I, \\ \dfrac{\mathrm{d}R}{\mathrm{d}t} = \dfrac{r_I I + A}{D_I} + \dfrac{H}{D_H} \\ \dfrac{\mathrm{d}A}{\mathrm{d}t} = \dfrac{1-r_E}{D_E}E - \dfrac{A}{D_I} \\ \dfrac{\mathrm{d}H}{\mathrm{d}t} = \dfrac{1-r_I}{D_q}I - \dfrac{H}{D_H}. \end{cases} \tag{26}$$

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 that models the disease transmission over time resulting from the public health policies followed in France during the first wave of the epidemic of COVID-19.

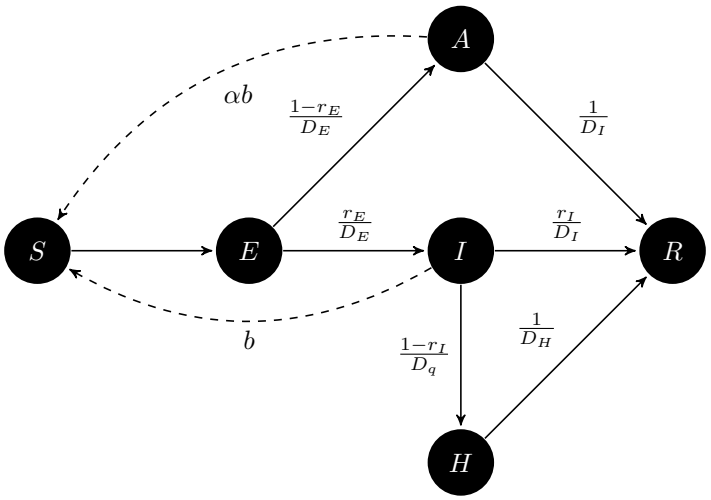


Figure 6: SEIRAH model representation – adapted from [37]

Parameter	Interpretation	Value
b	Transmission rate of ascertained cases	Region/Department Specific - **Estimated**
r_E	Ascertainment rate	0.844 [20]
r_I	Non hospitalized rate	0.966 [1]
α	Ratio of transmission between A and I	0.55 [31]
D_E	Latent (incubation) period (days)	5.1 [27]
D_I	Infectious period (days)	5 [8]
D_q	Duration from I onset to H (days)	11-D_E [13]
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 correlation between data when considering region data, see Appendix A.

In this respect, System (26) should be complemented by a dynamic for b in each region or department. We consider a dynamics of the form

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

where g is a function describing the health policy effect on the transmission rate b. For example, a transmission rate decreases from b_{init} to b_{end} due to a lockdown policy centered around the time t_ℓ with

an efficiency time-rate $\frac{1}{\tau}$ can typically be modeled with a logistic function denoted by ℓ

$$\ell(t) = b_{\text{init}} - \frac{b_{\text{init}} - b_{\text{end}}}{1 + e^{-\frac{t-t_\ell}{\tau}}}$$

leading to the assumption that $g(t) = \ell'(t)$. Moreover, the dynamics of multiple lockdown can be modeled by a combination of such g-functions during intervals, while unlockdown can also be represented by such g-functions with reversed sign.

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

$$\begin{pmatrix} x^i_{n+1} \\ b^i_{n+1} \\ \theta^i_{n+1} \end{pmatrix} = \begin{pmatrix} x^i_n + \delta T_n\, f(x^i_n, b^i_n, \theta^i_n) \\ b^i_n + \delta T_n\, g^i(t_n, \theta^i_n) \\ \theta^i_n \end{pmatrix} + \begin{pmatrix} \not{\vdash}_5 \\ \delta T_n \\ \not{\vdash}_{N_p} \end{pmatrix} \nu^i_n, \qquad \text{for each region/department } i$$

where f accounts for the dynamics in (26) and S is subsequently reconstructed by using $S = N - (E + I + R + A + H)$ in each region/department. Again, the discretization is completed by introducing model uncertainties, which for all $1 \leq i \leq N_P$ are given by the sequence $(\nu^i_n)_{1\leq n\leq N_T} \sim_{\text{i.i.d.}} \mathcal{N}(0, Q^i_n)$. Finally, recall that $\theta \in \mathbb{R}^{N_p}$ gathers all unknown time-independent parameters that need to be estimated, namely here b_{end} and b_{init} characterizing the g^i functions.

Since all state variables E, I, R, A, H are positive and bounded by N, the total population size of the department, following our Remark 4, they are going to be mapped with a scaling and *logit* transform

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

As for the variable b, we all use a weighted *logit* function: $\max_b \text{logit}(b)$, where $\max_b$ was set to 1.5 after preliminary investigation. This allows b to be estimated positive and smaller than $\max_b$.

Neglecting deaths, the proportion of infected individuals in the population in each region / department at a given time is given by

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

This quantity, which can also be calculated using other methods, is used to validate our strategy.

4.2.2 Available Data

To proceed with the estimation of these uncertain quantities, we have at our disposal the hospitalization data, prevalence and daily number incidence, denoted H_{data} and $H_{in,data}$, respectively. They are collected in the database SI-VIC, which records COVID-19 hospitalizations of patients in French hospitals. These two quantities are represented with the solutions of System (26) 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
$100\Big(1 - \frac{b_{\text{end}}}{b_{\text{init}}}\Big)\%$	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 parameters b_{init} and b_{end} at the regional level. The third line gives the corresponding decrease rate: $100\Big(1 - \frac{b_{\text{end}}}{b_{\text{init}}}\Big)\%$.

We develop the following strategy, which allows us to avoid any *a priori* about the shape of b:

- Step 1 – Estimate the parameter b_{init} (resp. b_{end}) using data obtained before lockdown (resp. 10 days after lockdown). The values are given in Table 6. It can be seen that the values are quite similar in all regions. In this first step, the transmission rate b is assumed to be 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.* assuming $g(t) = 0$, $\forall t$ – but setting b_{init} to the value estimated in Step 1. The results are shown in Figure 7a for all regions. It shows the form of a logistic function, confirming our modeling choice. Since there is no information about b, the model error is very important and leads to overfitting of the data. For many regions, oscillations with a mean of 0 and a period of up to 7 days are visible.
- Step 3 – To assess whether these weekly plateaus are just compensation for 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 shown in Figure 7b. In Figure 7c, we calculate the difference between the *a priori* logistic function b and its estimate. We can see 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 can be read in Appendix B, the effective reproductive ratio can be written as a function of the model parameters as follows

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

Figure 7d compiles the estimated effective reproductive ratios. Before the lockdown the values of the effective reproductive ratios are around 3.5, a value comparable to those found in the literature for the basic reproductive ratio, see [47, 15]. The effective reproductive ratios at the end of the lockdown are less than 0.7, showing the efficiency of the lockdown.

The proportion of infected individuals in the population in each region at the end of the first lockdown (May 11, 2020) is shown in Figure 8a. These results are very similar to those in [47] that allow us to validate our strategy.

4.2.4 Estimation in the 94 departments

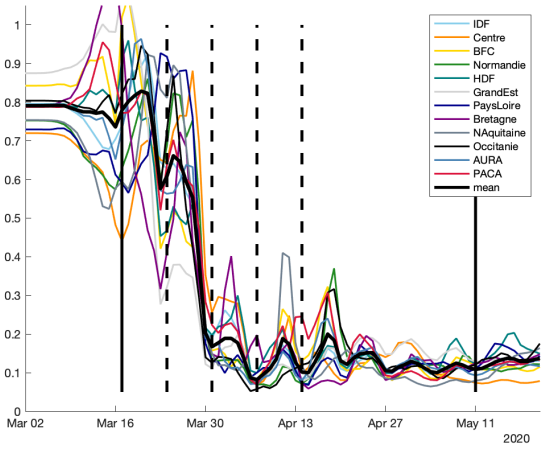
Let us now proceed to a more refined population and consider departments. Here we apply only the last step 3, using the estimated mean values of b_{init} and b_{end} obtained in the estimation for the 12 regions. Thus, we obtain the proportion of infected individuals in the population at the department level, as shown in Figure 8b. These results obtained with the Kalman filter for the entire population are then compared with the results obtained with a reduced-order version with $N_C = 24$. Figure 8d shows that the estimated infected individuals are very close to the values obtained with the full population Kalman filter, while the computation time is divided by 5: 40 minutes for the full filter versus 8 minutes 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 note that the cluster – see Figure 8c – does not fully match the partition of the region, as the epidemic in France follows a north-east to south-west gradient. Therefore, our reduced-order strategy allows better results at the departmental scale than estimation at the region scale, but with comparable computational complexity.

All these results will be discussed in more detail in a forthcoming paper devoted to the epidemic. However, we believe that they already show how versatile our approach is in terms of model complexity with appealing estimation performance and controlled computational complexity.

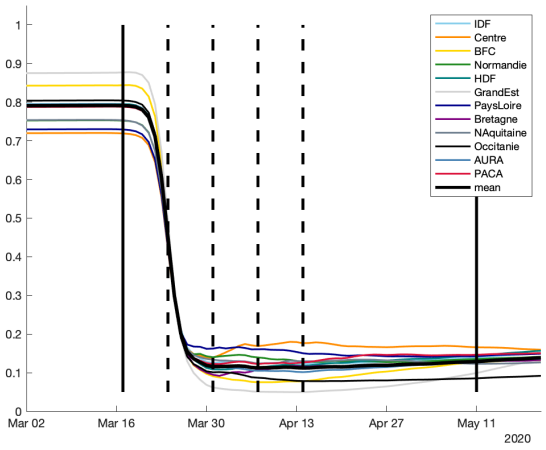
5 Conclusion

In this work, we have proposed a new strategy that allows combining the principle of data assimilation with the population configuration classically considered in the formulation of mixed effects models. In this regard, assuming Gaussian disturbances up to some variable transformation, we defined a log-likelihood functional that couples subjects in an elegant way. We approximate the resulting minimization with a dynamic programming approach, here the Unscented Kalman Filter [22]. To reduce computation time, we augment our approach with a reduced-order version of our population filter after clustering the measurements. Note that our choice of UKF could be replaced by an Extended Kalman Filter [49] or an Ensemble Kalman Filter [18] assuming that reduced-order versions are available [42, 40].

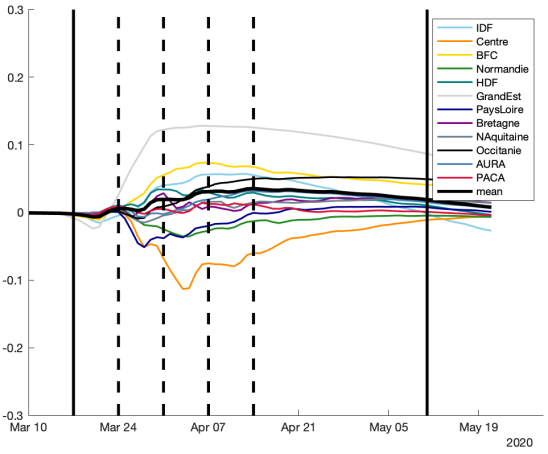
We illustrated and evaluated our proposed strategy on a simple pharmacokinetics model. Our approach has been shown to be successful on synthetic data. Moreover, on real data, our results compare very well with estimation procedure used as standard in the literature for estimation of parameters of nonlinear mixed effects models. We then moved to a more complex configuration, with stronger nonlinearities and modeling errors.



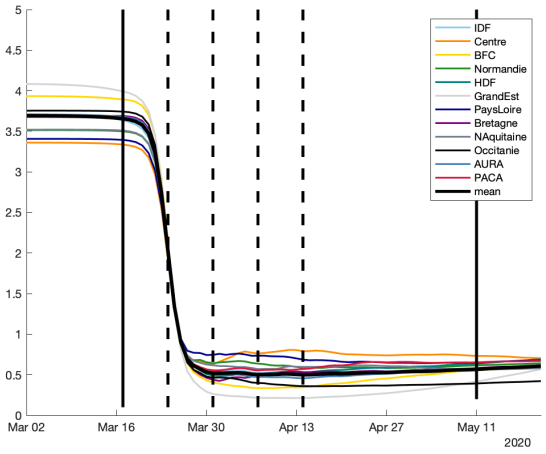
(a) Time evolution of b in the 12 French regions with $g = 0$ (no *apriori*)



(b) Time evolution of b in the 12 French regions when g is a logistic function and $\nu \neq 0$.

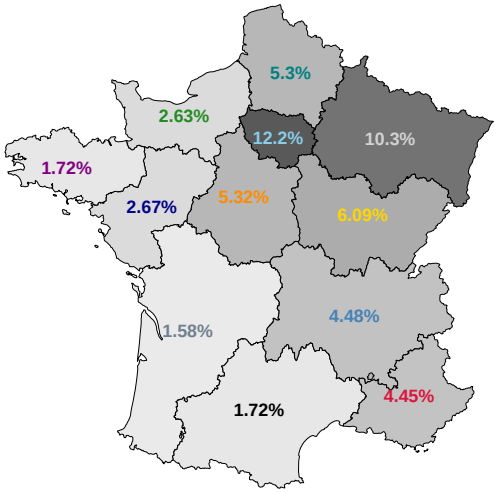


(c) Difference between g and b when g corresponds to a logistic function.

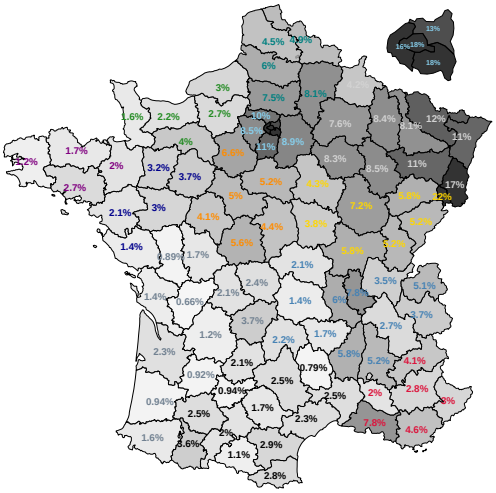


(d) Time evolution of R_{eff} in the 12 French regions when g is a logistic function and $\nu \neq 0$.

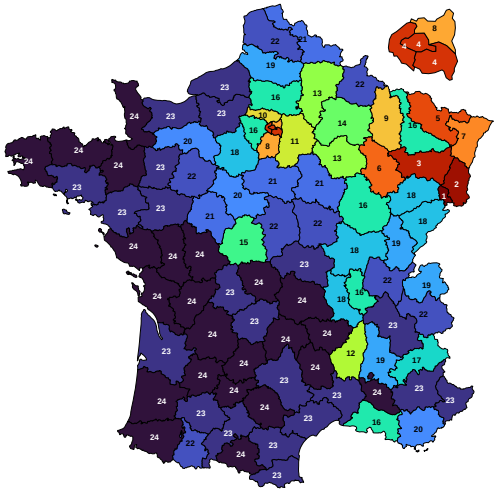
Figure 7: Illustration of the estimation strategy. The first (resp. last) vertical black line corresponds to the first (resp. last) day of lockdown. The other dashed lines are spaced 7 days apart.



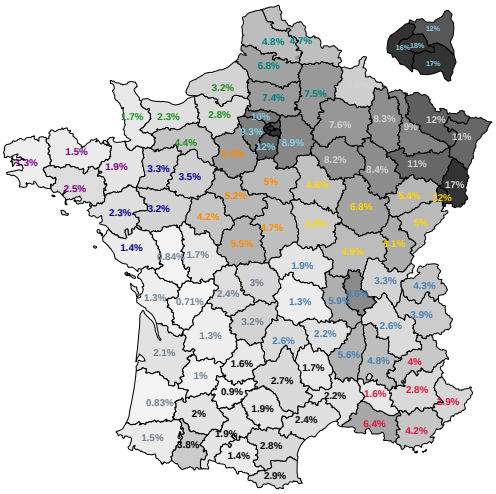
(a) 12 regions



(b) 94 departments



(c) Clustering illustration



(d) 94 departments - reduced-order version

Figure 8: Model estimate for the proportion of vaccinated individuals in the population (deaths not included) on May 11, 2020.

As explained in the introduction, the goal of this work was to propose a strategy that remains accurate, manageable, and within reasonable runtime as the dimension of the ODE system increases. The results obtained are very encouraging and the computational time achieved paves the way for the use of our strategy in large dimensional ODE models or possibly partial differential systems (PDE).

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

$$\frac{\mathrm{d}}{\mathrm{d}t}H_{\text{data}} = -\frac{H_{\text{data}}}{D_H} + H_{\text{in,data}}. \tag{27}$$

The correlation between H_{data} and $H_{\text{in,data}}$ at the regional level allows us to calculate D_H in each region using a mean square estimate. The values obtained are given in Table 7. We set D_H to the mean value.

Department-level hospitalization data cannot be trusted. This is because each region consists of several departments, and the sum of hospitalizations in all departments of a region is not consistent with hospitalizations in the region (which is the case for the data on daily case rates). In addition, when we calculate D_H at the department level, we get unrealistic values with a minimum value at 10 days and a maximum value at 30.5 days. This is partly due to poor reporting in patient transfers between hospitals from different departments. For this reason, we use only daily case rate data when considering departments and set D_H to the mean value estimated from regional-level data (see 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

B Computation of the effective reproductive ratio R_{eff}

To calculate the reproductive ratio R_{eff} of our SEIRAH model (26), we apply the *Next Generation Matrix* approach – see for example [21]. The principle is to focus 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 [38] 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.M. Navon. Data assimilation for geophysical fluids. In R. Temam and J. Tribbia, editors, *Handbook of Numerical Analysis: Computational Methods for the Atmosphere and the Oceans*. Elsevier, 2008.

[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, August 2017.

[7] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[8] 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.

[9] 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.

[10] 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.

[11] 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*, 80:1–41, 2017.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] L. C. Evans. *An introduction to stochastic differential equations*, volume 82. American Mathematical Soc., 2012.

[18] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2nd edition, 2009.

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

[20] 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.

[21] 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.

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

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

[24] 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.

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

[26] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[27] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin 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.

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

[29] 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.

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

[31] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.

[32] 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.

[33] 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.

[34] 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.

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

[36] 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.

[37] A. Pan, L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, et al. Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china. *Journal of American Medecine Association*, 323(19):1915–1923, 2020.

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

[39] 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.

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

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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. *International Journal for Numerical Methods in Biomedical Engineering*, 14(1):39–57, December 2021.

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

[50] 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.

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

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

[53] 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.

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