



**HAL**  
open science

# Population-based estimation for PDE system - Applications in electroporation of tumor spheroids.

Annabelle Collin

► **To cite this version:**

Annabelle Collin. Population-based estimation for PDE system - Applications in electroporation of tumor spheroids.. ESAIM: Control, Optimisation and Calculus of Variations, 2024, 10.1051/cocv/2024019 . hal-04508314v2

**HAL Id: hal-04508314**

**<https://inria.hal.science/hal-04508314v2>**

Submitted on 20 Mar 2024

**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 4.0 International License

# Population-based estimation for PDE system – Applications in electroporation of tumor spheroids

Annabelle Collin<sup>1\*</sup>

<sup>1</sup> Univ. Bordeaux, CNRS, Inria, Bordeaux INP, IMB, UMR 5251, F-33400 Talence, France

\* annabelle.collin@inria.fr

## Abstract

The estimation of partial differential systems (PDE) – in particular, the identification of their parameters – is fundamental in many applications to combine modeling and available measurements. However, it is well known that parameter prior values must be chosen appropriately to balance our distrust of measurements, especially when data are sparse or corrupted by noise. A classic strategy to compensate for this weakness is to use repeated measurements collected in configurations with common priors, such as multiple subjects in a clinical trial. In the mixed-effects approach, all subjects are pooled and a global distribution of model parameters in the population is estimated. However, due to the high computational cost, this strategy is often not applicable in practice for PDE. In this paper, we propose an estimation strategy to overcome this challenge. This sophisticated method is based on two important existing methodological strategies: (1) a population-based Kalman filter and, (2) a joint state-parameter estimation. More precisely, the errors coming from the initial conditions are controlled by a Luenberger observer and the parameters are estimated using a population-based reduced-order Kalman filter restricted to the parameter space. The performance of the algorithm is evaluated using synthetic and real data for tumor spheroid electroporation.

**Keywords:** Population-based estimation ; Sequential strategies ; PDE systems

## 1 Introduction

Systems of partial differential equations (PDE) can be used to understand and even predict the evolution of biological or physical processes. A major difficulty in applying these models in practice is dealing with the many uncertain quantities – such as initial conditions or model parameters – that must be specified for the models to be truly predictive. Available observations could provide additional information that can be used to circumvent these uncertainties. Estimation strategies can be viewed as inverse methods whose goal is to recover the uncertainties – such as initial conditions and parameters – that led to particular observations. Solving this inverse problem typically leads to solving a PDE constrained optimization problem, and there are a variety of techniques in the literature based on maximum likelihood optimization strategies to answer these types of questions.

Unfortunately, regardless of the method chosen, the estimation procedure is not robust enough in many situations when measurements are too sparse or noisy without relying on strong parameter priors. This observability weakness can be compensated for by the fact that multiple independent measurements have been collected for a configuration in which the model variables describe a statistical unit of interest belonging to a larger population in which variability is constrained. This is typically the assumption in nonlinear mixed-effects models [12, 27, 13] – in which the use of available population data improves the identifiability of individual parameters through coupled maximization of the likelihood function. Many algorithms and related software have been developed using this principle. These include the First-Order Conditional Estimation (FOCE) algorithm implemented in `NONMEM` [22], the Stochastic Approximation of Expectation-Maximization (SAEM) algorithm proposed in `MONOLIX` [11], penalized maximum likelihood approaches available in `NIMROD` [24], or even full Markov chain Monte Carlo methods (MCMC) implemented in `Winbugs` [28], `jags` [7], or `Stan` [2]. Most of these estimation algorithms and software have been compared in previous work and shown to achieve similar performance results for identifiable models of intermediate size, such as ordinary differential equations, for example in pharmacokinetics [8, 23, 24, 15]. However, it remains a challenge to increase the size of the underlying dynamics

– for example, when using PDE systems, which is the goal of this paper – with acceptable complexity costs of the algorithm. To our knowledge, there is only one paper that aims to develop a strategy for using population estimation strategies for PDE systems. In [9], the PDE system is solved by interpolations of solutions precomputed on a *carefully constructed* mesh of the parameter domain.

In this paper, we propose a formulation of a population-based estimator for PDE systems to address this challenge. The errors in the initial conditions are controlled by a Luenberger observer that corrects for the state, and the parameters are estimated using a population reduced-order Kalman-based filter restricted to the parametric space. This sophisticated method for solving this difficult problem is based on two important methodological strategies: (1) the population-based Kalman filter introduced in [6], (2) the joint state-parameter estimation introduced in [19] and further developed in [18]. More precisely, the population-based Kalman filter [6] captures all population information in a unified maximum likelihood estimation procedure assuming Gaussian disturbances. However even the proposed reduced version of this population-based Kalman filter – defined by combining clusters subpopulations with common observational backgrounds – results in the famous curse of dimensionality [1] that makes it numerically intractable for partial differential equations (PDE). Assuming that the dimension of the parameters is smaller than the dimension of the system state, which is then responsible for the curse of dimensionality, one can restrict the population-based Kalman filter to parametric space and couple it with a Luenberger observer following the strategy presented in [19] and further developed in [18]. This strategy assumes that a suitable Luenberger observer has been previously designed to control the trajectory.

The paper is organized as follows. Section 2 formalizes the estimation problem and introduces the population-based estimator. Section 3 provides the complete strategy for an illustrative 1D PDE system modeling spheroid electroporation. In a first subsection, the model is presented. In a second section, a Luenberger observer is introduced and mathematically investigated. In a third section, numerical experiments are performed on synthetic data, and in a final section, the results on real data are compared with a SAEM estimation. A final conclusion is drawn in Section 4.

## 2 Population-based estimation for PDE systems

### 2.1 Problem statement

#### 2.1.1 Dynamics

We consider a population of  $N_s$  subjects. For each subject  $i \in \llbracket 1, N_s \rrbracket$ , we denote by  $z^i$  the state of the system and by  $\theta^i$  a vector concatenating all time-independent parameters of the system. Denoting by  $\Omega$  an open-bounded domain of  $\mathbb{R}^d$  ( $d = 1, 2, 3$ ),  $T$  a strictly positive constant and by  $f$  a sufficient regular operator – corresponding to the dynamics of the system – we consider the following PDE system

$$\partial_t z^i(t, x) = f(t, z^i(t, x), \theta^i), \quad \forall t \in [0, T], \quad \forall x \in \Omega, \quad (1a)$$

$$z^i(x, 0) = z_0^i(x). \quad (1b)$$

We assume that it is well-posed meaning that for a given vector  $\theta^i$  of size  $N_\theta$  and for a given sufficient regular initial solution  $z_0^i \in \mathcal{X}$ , there exists a unique solution  $z^i \in L^2(0, T; \mathcal{X})$ , where  $\mathcal{X}$  is the state space.

#### 2.1.2 Observations

In practice, we do not observe  $z^i$  directly, but we do have observations – or measurements –  $y^i \in \mathcal{Y}$ . These observations  $y^i$  are related to  $z^i$  by the measurement process. Denoting by  $h : \mathcal{X} \rightarrow \mathcal{Y}$  the observation operator, we have

$$y^i(t, \cdot) = h(z^i(t, \cdot)) + \chi^i(t, \cdot), \quad (2)$$

where  $\chi^i$  represents the error inherent to the measurement process. The errors are assumed to be centered and we denote by  $w^i$  the covariance operator of the observation errors.

**Remark 1** *In many practical situations, the data are not continuous-time variables, which means that only a time sampling of the data is available. Two strategies are conceivable: either we use the data only when it is available, or we rely on some time interpolation. For simplicity, we assume that the data have been interpolated so that we can take continuous observations into account when presenting the mathematical theory.*

### 2.1.3 Initial condition and state observer

The initial condition of each subject is decomposed into two parts

$$z_0^i(x) = \hat{z}_0^i(x) + \zeta_z^i(x),$$

where  $\hat{z}_0^i$  is the known part and  $\zeta_z^i$  the uncertainty.

Considering that the uncertainties are restricted to the initial condition, we assume that an effective state observer – corresponding to a Luenberger observer [16] – compatible with the PDE system and the observations is already available. This means that one could define a term  $g$  – called the filter – such that the solution of the observer system

$$\begin{aligned} \partial_t \hat{z}^i(t, x) &= f(t, \hat{z}^i(t, x), \theta^i) + g(\hat{z}^i(t, x), y^i(t, x)), \quad \forall t \geq 0, \forall x \in \Omega, \\ \hat{z}^i(0, x) &= \hat{z}_0^i(x), \quad \forall x \in \Omega, \end{aligned}$$

converges in time to the solution of the target system

$$\begin{aligned} \partial_t z^i(t, x) &= f(t, z^i(t, x), \theta^i), \quad \forall t \in [0, T], \quad \forall x \in \Omega, \\ z^i(0, x) &= \hat{z}_0^i(x) + \zeta_z^i(x), \quad \forall x \in \Omega, \\ y^i(t, x) &= h(z^i(t, x)), \quad \forall t \geq 0, \forall x \in \Omega, \end{aligned}$$

*i.e.*  $\|\hat{z}(t, \cdot) - z(t, \cdot)\|_{\mathcal{X}}$  decreases to 0 when  $t$  increases.

### 2.1.4 Parameters

We would like to couple this state observer with an observer able to estimate the parameters. More precisely, we will consider a population-based approach for them and by typically assuming that the  $N_s$  subjects of the population are drawn at random from a same distribution, in this case a Gaussian distribution

$$\theta^i \sim_{\text{i.i.d.}} \mathcal{N}(\theta^{\text{POP}}, P^\theta), \quad 1 \leq i \leq N_s,$$

where  $\theta^{\text{POP}}$  is the population mean and  $P^\theta$  is the associated covariance. We equivalently write

$$\theta^i = \theta^{\text{POP}} + \tilde{\theta}^i \text{ with } \tilde{\theta}^i \sim_{\text{i.i.d.}} \mathcal{N}(0, P^\theta), \quad 1 \leq i \leq N_s.$$

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

## 2.2 Discretization

Before giving the estimation algorithm, we introduce the discretization. As for the time discretization, we restrict ourselves here to a forward Euler time scheme, but it is quite easy to consider more general time schemes. Given  $N_{\Delta t} \in \mathbb{N}^*$  and  $\Delta t = \frac{T}{N_{\Delta t}}$ , we define the  $N_{\Delta t} + 1$  time points  $t_n = n \times \Delta t$  for  $n \in \llbracket 0, N_{\Delta t} \rrbracket$ . Then, as for the spatial discretization, one could indiscriminately use finite element, finite volume or finite difference methods and we define  $N_{\Delta x}$  the size of the discretized state.

The uppercase letter is used to denote the spatial discretization of the state-dependent variables. The subscript  $n$  denotes the time discretization at time  $t_n$  of the time-dependent variables.

Finally, the prefix  $\#$  implies that the variable concatenates all  $N_s$  subjects either as a vector (for functions or nonlinear operators) or as a matrix (for covariance matrices).

More precisely, we have

$$\begin{aligned} \#Z_n &= \begin{pmatrix} Z_n^1 \\ Z_n^2 \\ \vdots \\ Z_n^{N_s} \end{pmatrix}, \# \theta = \begin{pmatrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^{N_s} \end{pmatrix}, \#Y_n = \begin{pmatrix} Y_n^1 \\ Y_n^2 \\ \vdots \\ Y_n^{N_s} \end{pmatrix}, \#W_n = \begin{pmatrix} W_n^1 & & & \\ & W_n^2 & & \\ & & \ddots & \\ & & & W_n^{N_s} \end{pmatrix}, \\ \#\hat{Z}_n &= \begin{pmatrix} \hat{Z}_n^1 \\ \hat{Z}_n^2 \\ \vdots \\ \hat{Z}_n^{N_s} \end{pmatrix}, \#\hat{\theta}_n = \begin{pmatrix} \hat{\theta}_n^1 \\ \hat{\theta}_n^2 \\ \vdots \\ \hat{\theta}_n^{N_s} \end{pmatrix}, \#F(t_n, \#\hat{Z}_n, \#\hat{\theta}_n) = \begin{pmatrix} F(t_n, \hat{Z}_n^1, \hat{\theta}_n^1) \\ F(t_n, \hat{Z}_n^2, \hat{\theta}_n^2) \\ \vdots \\ F(t_n, \hat{Z}_n^{N_s}, \hat{\theta}_n^{N_s}) \end{pmatrix}, \\ \#H(\#\hat{Z}_n) &= \begin{pmatrix} H(\hat{Z}_n^1) \\ H(\hat{Z}_n^2) \\ \vdots \\ H(\hat{Z}_n^{N_s}) \end{pmatrix} \text{ and } \#G(\#\hat{Z}_n, \#Y_n) = \begin{pmatrix} G(\hat{Z}_n^1, Y_n^1) \\ G(\hat{Z}_n^2, Y_n^2) \\ \vdots \\ G(\hat{Z}_n^{N_s}, Y_n^{N_s}) \end{pmatrix}. \end{aligned}$$

### 2.3 Estimation strategy

The state  $\#Z_n$ , for all  $n \in \llbracket 0, N_{\Delta t} \rrbracket$  and the parameters  $\#\theta$  are unknown. Our goal is to propose an algorithm that specifies the time evolution of the state  $\#\hat{Z}_n$ , the parameters  $\#\hat{\theta}_n$  and the associated covariance matrix  $\#\hat{P}_n^\theta$  in such a way that we have the convergence over time of

$$\|\#\hat{Z}_n - \#Z_n\| \rightarrow 0, \quad \|\#\hat{\theta}_n - \#\theta\| \rightarrow 0 \text{ and } \|\#\hat{P}_n^\theta - \#P^\theta\| \rightarrow 0$$

starting from the priors  $\#\hat{Z}_0$ ,  $\#\hat{\theta}_0$  and  $\#\hat{P}_0^\theta$ . The first two correspond to the prior parts of the initial conditions and the parameters. Inspired by the work of [6], the initial covariance matrix  $\#\hat{P}_0^\theta$  is defined as

$$\#\hat{P}_0^\theta = \left( \frac{1}{N_s^2} \bar{\mathbf{1}}_{N_s} \bar{\mathbf{1}}_{N_s}^\top \otimes M + \left[ \mathbb{I}_{N_s} - \frac{1}{N_s} \bar{\mathbf{1}}_{N_s} \bar{\mathbf{1}}_{N_s}^\top \right] \otimes (\hat{P}_0^\theta)^{-1} \right)^{-1}.$$

This matrix couples the population members, since in fact  $\#P_0^\theta \neq \mathbb{I}_{N_s} \otimes P_0^\theta$ .

However, the population-based Kalman filter introduced in [6] is writing on the augmented defined as

$$\begin{pmatrix} \hat{Z}_n^i \\ \hat{\theta}_n^i \end{pmatrix}.$$

However, in this case the size of the full matrix covariance – that we will denote  $\#P$  – containing the state and the parameters is  $(N_s \times (N_\theta + N_{\Delta x}))^2$  which leads to high computation times which makes it inapplicable in practice for large systems such as PDE systems. That is why, we couple this first methodological strategy to a second ones presented in [19] and further developed in [18]. This strategy consists in building a joint state-parameter estimation procedure based on a simple collocated feedback strategy for state estimation, adequately extended by Kalman filtering techniques to allow the simultaneous estimation of a limited set of unknown parameters. This then assumes that an effective state observer is already available.

Indeed, assuming that the full covariance matrix  $\#P$  is of reduced rank – typically much smaller than the dimension of the space  $N_s \times (N_\theta + N_{\Delta x})$  – the main idea in the reduced-order filtering strategies [21, 20, 18] is to be able to maintain covariance matrices in their factorized form

$$\#P = \#L \#U^{-1} \#L^\top,$$

where  $\#U$  is an invertible matrix of much smaller size than  $\#P$  and represents the main uncertainties in the system. What is crucial here is to be able to perform all computations on  $\#L$  and  $\#U$  without needing to compute  $\#P$ , see [25] and references therein. This implies that we only need to define

the initial reduced covariance matrix  $\#U_0^{-1}$  and the initial extension matrix  $\#L_0$  from the initial covariance matrix  $\#P_0$ . Here, we then consider

$$\#U_0^{-1} = \#P_0^\theta \text{ and } \#L_0 \stackrel{\text{def}}{=} \#L_0^\theta = \begin{pmatrix} \mathbb{O}_{N_{\Delta x}, N_\theta} & & & & \\ \mathbb{I}_{N_\theta, N_\theta} & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \mathbb{O}_{N_{\Delta x}, N_\theta} & \\ & & & \mathbb{I}_{N_\theta, N_\theta} & \end{pmatrix} \text{ where } \mathbb{O}_{p,q} \text{ and } \mathbb{I}_{p,q}$$

are respectively the zero matrix and the identity matrix of size  $p \times q$  meaning that the size of  $\#U_0$  is  $N_\theta N_s \times N_\theta N_s$  and the size of  $\#L_0$  is  $(N_{\Delta x} + N_\theta)N_s \times N_\theta N_s$ .

In [6], the authors also introduce a reduced-order version of the population-based Kalman filter that clusters subpopulations with common observational backgrounds. Assuming that the covariance matrix  $\#P_0^\theta$  ( $= \#U_0^{-1}$ ) is of reduced rank, it is reasonable to consider another SVD approximation

$$\#P_0^\theta = \#L_0^{\theta,C} (\#U_0^{\theta,C})^{-1} (\#L_0^{\theta,C})^\top,$$

where  $\#U_0^{\theta,C}$  is an invertible matrix of size  $N_\theta N_C \times N_\theta N_C$  with  $N_C$  the number of clusters. The definition of adequate matrices  $\#L_0^{\theta,C}$  and  $\#U_0^{\theta,C}$  are given in [6]. The algorithm given below combines both strategies *i.e.* we consider:

$$\#U_0 = (\#P_0^\theta)^{-1} \text{ and } \#L_0 = \#L_0^\theta, \text{ when no clustering is applied,} \quad (3a)$$

$$\#U_0 = \#U_0^{\theta,C} \text{ and } \#L_0 = \#L_0^\theta \#L_0^{\theta,C}, \text{ when clustering is applied.} \quad (3b)$$

Concerning the Kalman-based filter, 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 [25, 14]. 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  $N_\sigma$  sample points – the so-called sigma points [10]. Different choices of sigma-points can be used in practice from canonical sigma-points ( $N_\sigma = 2N_s \times N_\theta$ ) which are aligned with the canonical base to star sigma-points ( $N_\sigma = 2N_s \times N_\theta + 1$ ) in which the origin is added to the canonical points or simplex sigma-points which are located on a regular polyhedron and represents the smallest number of necessary sigma-points ( $N_\sigma = N_s \times N_\theta + 1$ ), see [10, 18] for more details. Here in the numerical part, we will use simplex sigma-points. To do so, we 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  $(\mathbf{e}^{(j)})_{1 \leq j \leq N_\sigma} \in \mathbb{R}^{N_\theta \times N_s}$  with the following empirical mean and covariances

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

where  $\delta$  is a scaling factor. As we consider a reduced version restricted to the parameters and even sometimes restricted to few clusters, we use the Reduced-order Unscented Kalman Filter (RoUKF) introduced in [18]. The full algorithm is given in Algorithm 1.

One can notice that the time evolution of the parameters is conditioned only by the Kalman filter (4c), which means

$$\#\hat{\theta}_{n+1}^{(j)+-} = \#\hat{\theta}_{n+1}^{(j)-} = \#\hat{\theta}_n^{(j)-}, \forall j \in \llbracket 1, N_\sigma \rrbracket,$$

while the state evolves first with the model dynamics (prediction (4a)), then with the state observer (state correction (4b)), and finally by the Kalman filter (parameter correction (4c)). This last step is fundamental to integrate the sensitivity of the parameters to the state and is the key point of the Reduced-order Unscented Kalman Filter (RoUKF) introduced in [18].

**Remark 2** Possible extensions of Algorithm 1 exist as:

- One could consider a transformation of the state and parameter variables allowing to add constraints (positive or bounded values for example). It is not difficult to extend the UKF filter to manage variables that follow a Gaussian distribution up to a certain transformation, see Remark 4 of [6] for more details.

---

**Algorithm 1** Luenberger observer coupled with a RoUKF
 

---

**Initialization**

Priors on initial conditions:  $\# \hat{Z}_0^+ = \# \hat{Z}_0$

Priors on parameters:  $\# \hat{\theta}_0^+ = \# \hat{\theta}_0$

Priors on the covariance matrix of the parameters:  $\# U_0$

Initial extension matrix:  $\# L_0$

Error covariances:  $(\# W_n)_{n \in [0, N_{\Delta t}]}$

Sigma points:  $\delta, \alpha_j, \mathbf{e}^{(j)}, \forall j \in [1, N_\sigma]$ ,

**Loop**  $\forall n \in [0, N - 1]$

[1] *Sampling*

$$\begin{pmatrix} \# \hat{Z}_n^{(j)+} \\ \# \hat{\theta}_n^{(j)+} \end{pmatrix} = \begin{pmatrix} \# \hat{Z}_n^+ \\ \# \hat{\theta}_n^+ \end{pmatrix} + \sqrt{\delta} \# L_n \sqrt{(\# U_n^+)^{-1}} \mathbf{e}^{(j)}, \forall j \in [1, N_\sigma]$$

[2] *Prediction (Forward)*

$$\begin{pmatrix} \# \hat{Z}_{n+1}^{(j)-} \\ \# \hat{\theta}_{n+1}^{(j)-} \end{pmatrix} = \begin{pmatrix} \# \hat{Z}_n^{(j)+} \\ \# \hat{\theta}_n^{(j)+} \end{pmatrix} + \Delta t \begin{pmatrix} \# F(t_n, \# \hat{Z}_n^{(j)+}, \# \hat{\theta}_n^{(j)+}) \\ 0 \end{pmatrix}, \forall j \in [1, N_\sigma] \quad (4a)$$

[3] *Correction (Analyze)*

State Correction

$$\begin{pmatrix} \# \hat{Z}_{n+1}^{(j)+-} \\ \# \hat{\theta}_{n+1}^{(j)+-} \end{pmatrix} = \begin{pmatrix} \# \hat{Z}_{n+1}^{(j)-} \\ \# \hat{\theta}_{n+1}^{(j)-} \end{pmatrix} + \Delta t \begin{pmatrix} \# G(\# \hat{Z}_{n+1}^{(j)-}, \# Y_n) \\ 0 \end{pmatrix}, \forall j \in [1, N_\sigma] \quad (4b)$$

Parameter Correction

$$\# L_{n+1} = \text{Cov}_{\alpha, N_\sigma}(\mathbf{e}^{(\cdot)}, \# \hat{Z}_{n+1}^{(\cdot)+-})$$

$$\Lambda_{n+1} = \text{Cov}_{\alpha, N_\sigma}(\mathbf{e}^{(\cdot)}, (\# H(\# \hat{Z}_{n+1}^{(\cdot)+-}) - \# Y_n))$$

$$\# U_{n+1} = \text{Cov}_{\alpha, N_\sigma}(\mathbf{e}^{(\cdot)}, \mathbf{e}^{(\cdot)}) + \Lambda_{n+1}^T (\# W_n)^{-1} \Lambda_{n+1}$$

$$K_{n+1} = \# L_{n+1} (\# U_{n+1})^{-1} \Lambda_{n+1} (\# W_n)^{-1}$$

$$\begin{pmatrix} \# \hat{Z}_{n+1}^+ \\ \# \hat{\theta}_{n+1}^+ \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{\alpha, N_\sigma}(\# \hat{Z}_{n+1}^{(\cdot)+-}) \\ \mathbb{E}_{\alpha, N_\sigma}(\# \hat{\theta}_{n+1}^{(\cdot)+-}) \end{pmatrix} + K_{n+1} (\# Y_n - \mathbb{E}_{\alpha, N_\sigma}(\# \hat{Z}_{n+1}^{(\cdot)+-})) \quad (4c)$$

- 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 [25, Section 7.4]. In this respect, we could introduce a forgetting factor, see [6] for more details.
- In the spirit of [14, Chapter 2], one could introduce some stochasticity into the discrete-time system by introducing an additive modeling error contribution, see [6] for more details. This has been applied in an ODE context in [5].
- If necessary, one strategy for estimating the initial condition is to parameterize it with few parameters and then add them to the parameter vector, see [4] for an illustration.

For the sake of the readability, we do not integrate here.

### 3 Application in 3D spheroids electroporation

#### 3.1 Problem statement

##### 3.1.1 Modeling

We will consider 3 compartments of cells: the proliferative ones whose the density denoted by  $p$ , the quiescent ones whose the density denoted by  $q$  and the cells with a functioning altered by the impact of the electrical shock whose the density denoted by  $f$ . By denoting the tumor spheroid domain at time  $t$  by  $\Omega(t)$ , the evolution of the three densities are supposed to satisfy the following equations

$$\begin{cases} \partial_t p + \nabla \cdot (\vec{v}p) &= \tau_G(t)(p + q) - \tau_O(t, x)p, & \Omega(t), \\ \partial_t q + \nabla \cdot (\vec{v}q) &= \tau_O(t, x)p, & \Omega(t), \\ \partial_t f + \nabla \cdot (\vec{v}f) &= 0, & \Omega(t), \end{cases} \quad (5)$$

where the function  $\tau_G$  defined by  $\tau_G(t) = ae^{-bt}$  is the growth rate which is decreasing over time ( $a$  and  $b$  are two strictly positive constants) and  $\vec{v}$  denotes the velocity field that describes the motion of tumor cells. Using the saturation hypothesis  $p + q + f = 1$  in  $\Omega(t)$ , we have  $\nabla \cdot v = \tau_G(t)(p + q) = ae^{-bt}(p + q)$ . The function  $\tau_O$  depending on  $(t, x)$  has to be chosen to describe the appearance of a high quiescent proportions of cell in the center of the spheroid. Denoting by  $t_{as}$  the time just after the electrical shock, the three following phenomenas will be mathematically modeled:

- (1) the death of proliferative and quiescent cells leads to  $P(t_{as}, x) = 0$ , for  $x \in \omega_P \subset \Omega(t < t_{as})$  and  $Q(t_{as}, x) = 0$ , for  $x \in \omega_Q \subset \Omega(t < t_{as})$  then  $\Omega(t_{as})$  decreases by a percentage noted  $p_c$ ,
- (2) the functioning modification of proliferative and quiescent cells leads to  $F(t_{as}, x) = \lambda(P(t_{as}, x) + Q(t_{as}, x))$ , for  $x \in \Omega(t_{as})$ ,
- (3) the value of the parameter  $a$  is decomposed into two parts:

$$a = \begin{cases} a_{bs}, \forall t < t_{as}, \\ ma_{bs}, \forall t \geq t_{as}, \end{cases} \quad (6)$$

where  $m$  is a positive constant which can be interpreted as a boost of the growth when the value is superior to 1 after the electrical shock.

Inspiring from the work of [17], the spherical symmetry of spheroid can be used to rewrite the model in radial and relative coordinates. Interestingly, this implies that the field is entirely determined by its divergence, therefore no assumption on the rheology of the tumor spheroid is needed.

For all  $x \in \Omega(t)$ , we denote the normalized radial coordinate ( $r = 0$  at the center of the spheroid,  $r = 1$  at the surface)

$$r(t, x) = \frac{\|x\|}{R(t)},$$

where  $R$  is the spheroid radius. Under the hypothesis of invariance by rotation, we can then define, for all  $(t, x)$ :

$$U(t, r(t, x)) = u(t, x), \text{ for unknowns } u = \{p, q, f, \tau_O, \vec{v}\} \text{ and } U = \{P, Q, F, \mathcal{T}_O, \vec{V}\}.$$

We denote by  $V$  the radial component of the velocity  $\vec{V}$  and the invariance by rotation implies that  $\vec{V} = V\vec{e}_r$ .

In [3], the authors show that System (5) becomes

$$\begin{cases} R' &= R\tau_G I(t, 1), \\ \partial_t P &= -\tau_G (r^{-2}I(t, r) - rI(t, 1))\partial_r P + \tau_G (1 - F)(1 - P) - \mathcal{T}_O P, \\ \partial_t F &= -\tau_G (r^{-2}I(t, r) - rI(t, 1))\partial_r F - \tau_G (1 - F)F, \end{cases} \quad (7)$$

where  $I$  is the following function of  $t$  and  $r$

$$I(t, r) = \int_0^r (P(t, \underline{r}) + Q(t, \underline{r}))\underline{r}^2 d\underline{r} = \int_0^r (1 - F(t, \underline{r}))\underline{r}^2 d\underline{r},$$

and then  $Q$  can be determined using the fact that  $Q = 1 - (P + F)$ .

To close the system, the evolution of the rate  $\mathcal{T}_O$  has to be described. Seen as representing the lack of oxygen, one can consider the following logistic function

$$\mathcal{T}_O(t, r) = \mathcal{T}_O^{begin} - \frac{\mathcal{T}_O^{begin} - \mathcal{T}_O^{end}}{1 + \exp\left(\frac{R(t)(1 - r) - \mathcal{T}_O^d}{\mathcal{T}_O^s}\right)}.$$

The parameter  $\mathcal{T}_O^d > 0$  corresponds to the distance from the tumor front for which the oxygen is easily accessible. The parameter  $\mathcal{T}_O^s > 0$  corresponds to the slope of the function. The parameters  $\mathcal{T}_O^{begin} > 0$  (resp.  $\mathcal{T}_O^{end} > 0$ ) corresponds to the value of  $\mathcal{T}_O$  far from (resp. close to) the tumor boundary.

Initial conditions are given by:

$$P(0, r) = \hat{P}_0 + \xi_P, \quad Q(0, r) = 1 - P(0, r), \quad F(0, r) = 0, \quad \forall r,$$

and  $R(0) = \hat{R}_0 + \xi_R$  where  $\hat{R}_0$  and  $\hat{P}_0$  (resp.  $\xi_P$  and  $\xi_R$ ) are constants corresponding to the prior (resp. uncertainty) parts. We choose them as  $\hat{R}_0 + \xi_R \geq 0$ ,  $\hat{R}_0 \geq 0$ ,  $\hat{P}_0 + \xi_P \in [0, 1]$  and  $\hat{P}_0 \in [0, 1]$ .

The parameters that we want to estimate are  $a_{bs}$ ,  $b$ ,  $\lambda$ ,  $p_c$  and  $m$ .

**Remark 3** *The parameters  $\mathcal{T}_O^{begin}$ ,  $\mathcal{T}_O^{end}$ ,  $\mathcal{T}_O^s$  and  $\mathcal{T}_O^d$  are not identifiable with the kind of observations considered in this article. To reproduce heterogeneity similar to the data available in [17] (Figure 6, CAPAN-2 control), we fix the values  $\mathcal{T}_O^d = 0.3$  mm,  $\mathcal{T}_O^{begin} = 0.04$  s<sup>-1</sup>,  $\mathcal{T}_O^{end} = 0.005$  s<sup>-1</sup> and  $\mathcal{T}_O^s = 0.05$  mm.*

### 3.1.2 Mathematical analysis

We admit here that there exists a unique solution to the dynamical system (7) and this solution is smooth (this can be proved by a Banach contraction mapping theorem).

**Proposition 1** *We have*

- (1)  $\forall t \geq 0, \forall r \in [0, 1], U(t, r) \in [0, 1]$ , for  $U = \{P, Q, F\}$  and  $P(t, r) + Q(t, r) \in [0, 1]$ ,
- (2)  $\forall t \geq 0, R(t) \in \left[\min(R(0), R(t_{as})), R(0)e^{\frac{m a_{bs}}{3b}}\right]$ .

The proof of Proposition 1 is given in Appendix A.1.

### 3.1.3 Observations

We observe the radius at different times for a population of spheroids submitted to the same high intensity pulsed field.

## 3.2 Analysis of the state observer

### 3.2.1 State observer

We design the following Luenberger-type observer

$$\begin{cases} \hat{R}' &= \hat{R}\tau_G\hat{I}(t,1) - \gamma_{\text{obs}}(t)(\hat{R} - R), \\ \partial_t\hat{P} &= -\tau_G(r^{-2}\hat{I}(t,r) - r\hat{I}(t,1))\partial_r\hat{P} + \tau_G(1 - \hat{F})(1 - \hat{P}) - \hat{\mathcal{T}}_O\hat{P}, \\ \partial_t\hat{F} &= -\tau_G(r^{-2}\hat{I}(t,r) - r\hat{I}(t,1))\partial_r\hat{F} - \tau_G(1 - \hat{F})\hat{F}, \end{cases} \quad (8)$$

with

$$\hat{I}(t,r) = \int_0^r (1 - \hat{F})\underline{r}^2 d\underline{r}, \quad \text{and} \quad \hat{\mathcal{T}}_O(t,r) = \mathcal{T}_O^{\text{begin}} - \frac{\mathcal{T}_O^{\text{begin}} - \mathcal{T}_O^{\text{end}}}{1 + \exp\left(\frac{\hat{R}(t)(1-r) - \mathcal{T}_O^d}{\mathcal{T}_O^s}\right)}.$$

The function  $\gamma_{\text{obs}}$  is the gain function. We assume that the function goes from  $[0, T]$  to  $]0, \infty[$  and is bounded *i.e.* there exists  $\gamma_{\text{obs}}^{\min}, \gamma_{\text{obs}}^{\max} \in ]0, \infty[$  such that  $\gamma_{\text{obs}}^{\min} \leq \gamma_{\text{obs}}(t) \leq \gamma_{\text{obs}}^{\max}$ , for all  $t \in [0, T]$ . Initial conditions are given by

$$\hat{R}(0) = \hat{R}_0, \quad \hat{P}(0, r) = \hat{P}_0 \in [0, 1], \quad \hat{Q}(0, r) = 1 - \hat{P}_0, \quad \hat{F}(0, r) = 0, \quad \forall r.$$

In this part, we only consider uncertainties reduced to the initial condition. The values of the parameters are fixed to the true values. As previously, we admit that there exists a unique solution to the dynamical system (8) and this solution is smooth. One can easily show that Proposition 1-(1) is still valid for  $\hat{P}, \hat{Q}$  and  $\hat{F}$ .

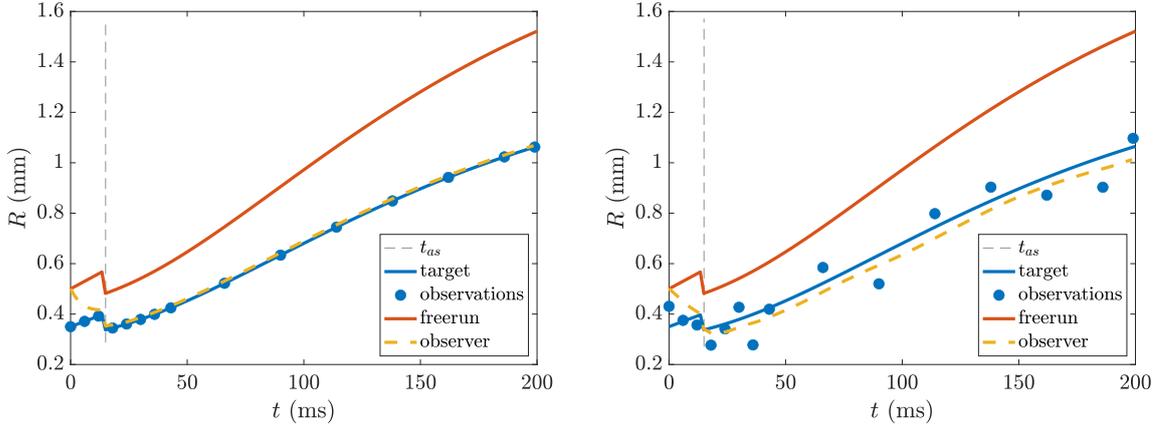


Figure 1: Radius evolution. Blue curve: target solution ( $R$ ) ; blue circles: observations ; orange curve: freerun solution ( $\hat{R}$  with  $\gamma_{\text{obs}} = 0$ ) and yellow dashed curve: observer solution ( $\hat{R}$  with  $\gamma_{\text{obs}} = 0.01 + 0.2(t < 20)$ ). Left: no noise. Right:  $\sigma_{\text{err}} = 0.08$ .

### 3.2.2 Analysis

In order to analyze the convergence of the observer we will study the estimation error defined by  $\tilde{R} = R - \hat{R}$  and  $\tilde{P} = P - \hat{P}$ . One can prove the two following propositions (see the proofs in Appendix A.2).

**Proposition 2** *If  $\gamma_{\text{obs}}^{\min} > \frac{ma_{\text{bs}}}{3}$ , the radius  $t \mapsto \tilde{R}(t)$  converges exponentially to 0 when  $t$  goes to  $+\infty$ .*

**Proposition 3** *If  $\gamma_{\text{obs}}^{\min} > \frac{ma_{\text{bs}}}{3} + \mathcal{T}_O^{\text{end}}$  and  $\tilde{P}(0, \cdot) \in H^s(]0, r[)$ , the norm  $t \mapsto \|\tilde{P}(t, \cdot)\|_{L^2(]0, 1])}^2$  converges exponentially to 0 when  $t$  goes to  $+\infty$ .*

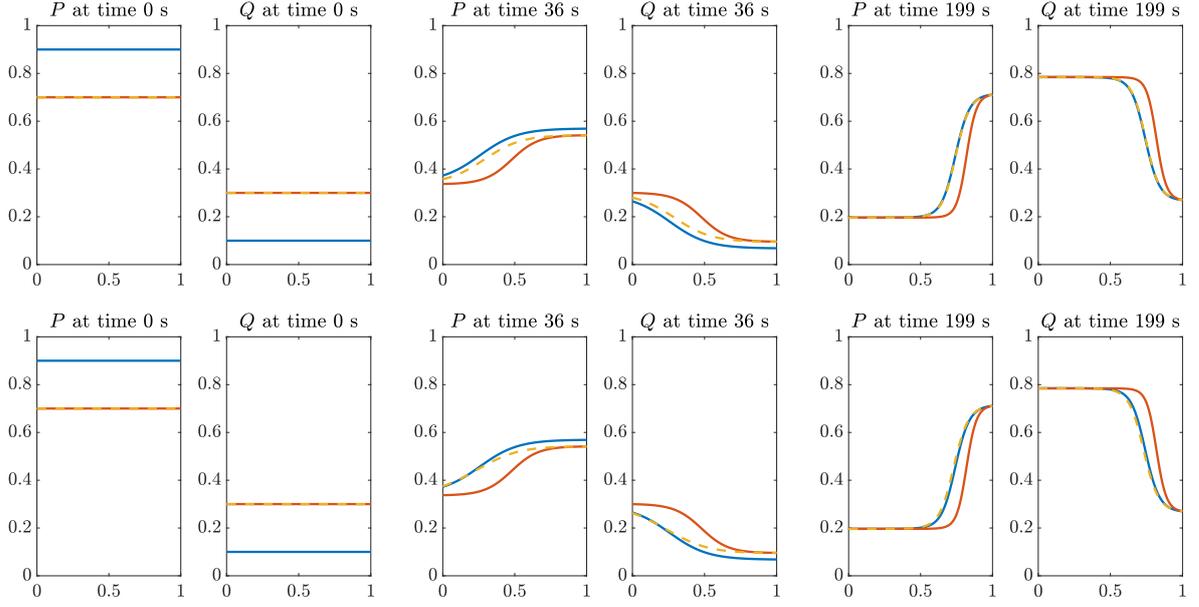


Figure 2: Time evolutions of proliferative and quiescent cells. Blue curve: target solution  $(P, Q)$ , orange curve: freerun solution  $(\tilde{P}, \tilde{Q}$  with  $\gamma_{\text{obs}} = 0 \text{ s}^{-1}$ ) and yellow dashed curve: observer solution  $(\tilde{P}, \tilde{Q}$  with  $\gamma_{\text{obs}} = 0.01 + 0.2(t < 20)$ ). Top: no noise. Bottom:  $\sigma_{\text{err}} = 0.08$ .

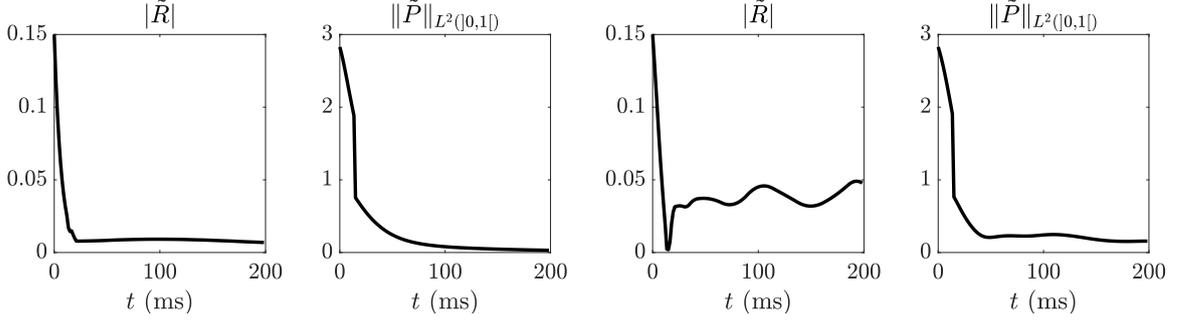


Figure 3: Time evolutions of the errors:  $|\tilde{R}| := |R - \hat{R}|$  and  $\|\tilde{P}\|_{L^2([0,1])} := \|P - \hat{P}\|_{L^2([0,1])}$ . Left: no noise. Right:  $\sigma_{\text{err}} = 0.08$ .

### 3.2.3 Numerical illustrations with synthetic data

We illustrate the observer model by considering for initial conditions:

$$R(0) = 0.35 \text{ mm}, \hat{R}_0 = 0.5 \text{ mm}, P(0) = 0.9 \text{ and } \hat{P}_0 = 0.7.$$

The numerical approximation is obtained using the explicit Euler scheme ( $t_{\text{max}} = 200 \text{ s}$ ,  $\Delta t = 0.2 \text{ s}$ ) for the time discretization. Concerning the spatial discretization, we will consider for  $P$  and  $F$  a finite difference of order 1 with a spatial step of  $\Delta_x = 0.05$  meaning that the size of the size discretized state equals to  $N_{\Delta_x} = 2 \times \frac{1}{\Delta_x} + 1$ . The values of the parameters are fixed at:  $a = 0.03 \text{ s}^{-1}$ ,  $b = 0.01 \text{ s}^{-1}$ ,  $\lambda = 0.6$ ,  $p_c = 0.85$ ,  $m = 2$  and  $t_{as} = 15 \text{ s}$ . Concerning the observations, we consider the following  $N^{\text{obs}}$  time points for the observations:

$$t^{\text{obs}} = [0, 6, 12, 18, 24, 30, 36, 43, 66, 90, 114, 138, 162, 186, 199] \text{ s},$$

and we denote by  $Y_j^R = R(t^0(j))$  for all  $j \in [1, N^{\text{obs}}]$  the corresponding observations. 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. Here, we choose to interpolate.

Scenarios		<b>b</b>	<b>k</b>	<b>R<sub>0</sub></b>	<b>P<sub>0</sub></b>
<b>Target</b>		$m_b = 0.01$ $\sigma_b = 0.5$	$m_k = 3$ $\sigma_k = 0.5$	$m_{R_0} = 0.35$ $\sigma_{R_0} = 0.3$	$m_{P_0} = 0.9$ $\sigma_{P_0} = 0.3$
<b>str. pr.</b>	$\gamma_{\text{obs}} = 0.0$	$m_{\hat{b}} = 0.01$ $\sigma_{\hat{b}} = 0.5$	$m_{\hat{k}} = 3$ $\sigma_{\hat{k}} = 0.5$	$\hat{R}_0 = R_0$	$\hat{P}_0 = P_0$
<b>str. pr. - so</b>	$\gamma_{\text{obs}} = 0.01 + 0.2(t < 20)$	$m_{\hat{b}} = 0.01$ $\sigma_{\hat{b}} = 0.5$	$m_{\hat{k}} = 3$ $\sigma_{\hat{k}} = 0.5$	$\hat{R}_0 = R_0$	$\hat{P}_0 = P_0$
<b>w. pr.</b>	$\gamma_{\text{obs}} = 0.0$	$m_{\hat{b}} = 0.05$ $\sigma_{\hat{b}} = 1$	$m_{\hat{k}} = 1$ $\sigma_{\hat{k}} = 1$	$\hat{R}_0 = R_0$	$\hat{P}_0 = P_0$
<b>w. pr. - so</b>	$\gamma_{\text{obs}} = 0.01 + 0.2(t < 20)$	$m_{\hat{b}} = 0.05$ $\sigma_{\hat{b}} = 1$	$m_{\hat{k}} = 1$ $\sigma_{\hat{k}} = 1$	$\hat{R}_0 = R_0$	$\hat{P}_0 = P_0$
<b>w. pr., f. IC</b>	$\gamma_{\text{obs}} = 0.0$	$m_{\hat{b}} = 0.05$ $\sigma_{\hat{b}} = 1$	$m_{\hat{k}} = 1$ $\sigma_{\hat{k}} = 1$	$\hat{R}_0 = 0.5$	$\hat{P}_0 = 0.7$
<b>w. pr., f. IC - so</b>	$\gamma_{\text{obs}} = 0.01 + 0.2(t < 20)$	$m_{\hat{b}} = 0.05$ $\sigma_{\hat{b}} = 1$	$m_{\hat{k}} = 1$ $\sigma_{\hat{k}} = 1$	$\hat{R}_0 = 0.5$	$\hat{P}_0 = 0.7$

Table 1: Values of priors for the 6 scenarios: strong priors (str. pr.), strong priors with state observer (str. pr. - so), weak priors (w. pr.), weak priors with state observer (w. pr. - so), weak priors and false initial conditions (w. pr., f. IC), weak priors and false initial conditions with state observer (w. pr., f. IC - so)

The following gain function is considered:  $\gamma_{\text{obs}} = 0.01 + 0.2(t < 20) \text{ s}^{-1}$ . It is classical to consider a highest gain at the beginning to correct error of initial condition without integrating the noise too much.

Figure 1-Left shows the time evolution of the radius for the target, the freerun and the observer models. One can see that the observer solution (yellow dashed curve) converges in time to the target solution (blue curve). Figure 2-Top shows the time evolution of proliferative and quiescent cells for the target, the freerun and the observer models. One can see that the observer solution (yellow dashed curve) converges in time to the target solution (blue curve). Figure 3-Left shows the evolution of the following errors:  $|\tilde{R}| := |R - \hat{R}|$  and  $\|\tilde{P}\|_{L^2([0,1])} := \|P - \hat{P}\|_{L^2([0,1])}$ .

We then consider additive noise:  $Y_j^R = R(t^0(j)) + \chi_j^R$ , where  $\chi_j^R \sim \mathcal{N}(0, \sigma_{\text{err}}^2)$  for all  $j \in [1, N^{\text{obs}}]$ , with  $\sigma_{\text{err}}$  the standard deviation of the error. We consider  $\sigma_{\text{err}} = 0.08$  to see the ability of our observer to deal with measurement errors. Results are given in Figures 1-Right, 2-Bottom and 3-Right. One can see that our observer is sensitive to the noise for the estimation of the radius but the results are still reasonable. The gain function  $\gamma_{\text{obs}}$  has to be chosen very carefully to avoid the overfitting.

### 3.3 State and parameter estimation

#### 3.3.1 Synthetic data benchmark

In this section, we would like to test the ability of our algorithm to estimate parameters. First, a synthetic data benchmark is considered to validate our implementation. To do so, we run different simulation series with different numbers of spheroids and with different measurement errors – *i.e.* with different values of the standard deviation of the additive Gaussian noise always denoted by  $\sigma_{\text{err}}$  – for 6 scenarios: strong priors with  $\gamma_{\text{obs}} = 0$  (**str. pr.**), strong priors with  $\gamma_{\text{obs}} > 0$  (**str. pr. - so**), weak priors with  $\gamma_{\text{obs}} = 0$  (**w. pr.**), weak priors with  $\gamma_{\text{obs}} > 0$  (**w. pr. - so**), weak priors and false initial conditions with  $\gamma_{\text{obs}} = 0$  (**w. pr., f. IC**), weak priors and false initial conditions with  $\gamma_{\text{obs}} > 0$  (**w. pr., f. IC - so**). To limit the energy consumption of this validation, we will focus on estimating only two parameters:  $b$  and  $a_{bs}$  considering the model in the free growth case, *i.e.* without electroporation. More precisely, instead of estimating  $a_{bs}$ , we will estimate  $k = \frac{a_{bs}}{b}$  which is assumed to be fixed in the population when using the reduced Gompertz model introduced in [26].

For the  $N_R$  replicates of all the cases, the data have been generated with the following Gaussian

laws

$$\begin{pmatrix} \log(R_0) \\ \text{logit}(P_0) \\ \log(b) \\ \log(k) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \log(m_{R_0}) \\ \text{logit}(m_{P_0}) \\ \log(m_b) \\ \log(m_k) \end{pmatrix}, \begin{pmatrix} \sigma_{R_0}^2 & 0 & 0 & 0 \\ 0 & \sigma_{P_0}^2 & 0 & 0 \\ 0 & 0 & \sigma_b^2 & 0 \\ 0 & 0 & 0 & \sigma_k^2 \end{pmatrix} \right),$$

whose values are given in Table 1. For all cohorts, the value of  $k$  in the cohort is fixed (but varies from cohort to cohort). The prior values of the initial conditions are

$$\begin{pmatrix} \log(\hat{R}_0) \\ \text{logit}(\hat{P}_0) \end{pmatrix}$$

and the prior values and standard deviations of the parameters used for the algorithm correspond to the following Gaussian laws

$$\begin{pmatrix} \log(\hat{b}) \\ \log(\hat{k}) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \log(m_{\hat{b}}) \\ \log(m_{\hat{k}}) \end{pmatrix}, \begin{pmatrix} \sigma_{\hat{b}}^2 & 0 \\ 0 & \sigma_{\hat{k}}^2 \end{pmatrix} \right),$$

whose values are given in Table 1. For the algorithm, the prior mixed covariance matrix

$$\begin{pmatrix} (\sigma_b^f)^2 & 0 \\ 0 & (\sigma_k^f)^2 \end{pmatrix}$$

is also needed. When there is no information about the distribution of the parameters, one can consider that the standard deviation is divided by 2:  $\sigma_b^f = 0.5 \sigma_b$ . Concerning the parameter  $k$  which is assumed to be fixed in the population, one can consider  $\sigma_k^f = \sigma_k$ .

**Remark 4** *One might note that in this particular case, the prior mixed covariance matrix is of size 2 by 2 because the population Kalman filter is reduced to the parameters as the state of the system is managed by the Luenberger state observer.*

To validate our algorithm, we consider several statistical validation criteria whose the definitions are given in Appendix B: the relative bias (rBIAS), the relative mean squared error (rMSE), the mean estimated standard deviation (STD), the empirical standard deviation of estimates, defined as the standard deviation of estimates at convergence (ESTD) and the coverage (COV).

### 3.3.2 Results using Algorithm 1 with Eq. (3a)

We first study the results obtained with Algorithm 1 when no clustering is considered meaning that the initial reduced covariance matrix  ${}^\#U_0$  and the initial extension matrix  ${}^\#L_0$  are given by Eq. (3a). We generate  $N_R = 100$  simulation replicates for  $N_s = 1, 5, 10, 20$  and  $\sigma_{err} = 0, 0.04, 0.08$  for the 6 scenarios. We also generate  $N_R = 100$  simulation replicates for  $N_s = 40$  but only for the scenario **w. pr., f. IC - so**.

To illustrate how the algorithm works, Figure 4 shows the time evolution of the states and parameters of two spheroids belonging to a cohort of 10 spheroids and  $\sigma_{err} = 0.08$ . The generated data and the two selected spheroids are indicated (top, left). The second line corresponds to Spheroid 2 and the third line to Spheroid 6. In both cases, the first column corresponds to radius  $R$ , the second to proliferative and quiescent densities  $P, Q$ , and the third to parameter  $b$ . The parameter  $k$  – estimated fixed in the population – is given in the upper right corner. The target states and target values of the parameters are shown in blue. The blue points correspond to the noisy observed data. The states and parameter values of the observer are shown in yellow. The dashed lines correspond to the 95% confidence interval when available (*i.e.* for the parameters only). One can see that the yellow lines converge to blue lines over time, showing the ability of our algorithm to estimate well.

Figure 5 (resp. Figure 6) shows the rMSE, the rBIAS, and the COV concatenated for both parameters  $b$  and  $k$  and the STD, the ESTD, and the BMIXED for parameter  $b$  against the number of spheroids considered  $N_s$  for a standard deviation error  $\sigma_{err}$  of 0 (resp. 0.08) for the

Algo	$\Delta_x$	$N_s = 1$	$N_s = 5$	$N_s = 10$	$N_s = 20$	$N_s = 40$
Algorithm 1	0.05	0.03	0.3	1	4	12
Algorithm 1	0.01	x	x	3.5	x	x
Algorithm 1	0.001	x	x	60	x	x
Function <i>nlmefitsa</i>	0.05	x	80	x	x	x

Table 2: Computation times given in **minutes** when estimating 2 parameters ( $b$  and  $k$ ) in Scenario **w. pr., f. IC - so**. Comparison with the **Matlab** function *nlmefitsa* (average of computation times for 10 replicates).

6 scenarios. First, one can see that, as expected, the results are better when strong priors and true initial conditions are considered. Second, using the state observer in conjunction with the Kalman observer for the parameters leads to better results when weak priors are considered and especially when the initial conditions are false. Since Scenario **w. pr., f. IC - so** is the most interesting, we will focus on it. Figure 7 shows the rMSE, the rBIAS, and the COV concatenated for both parameters  $b$  and  $k$  and the STD, the ESTD and the BMIXED for parameter  $b$  against the considered errors. As expected, increasing the number of spheroids gives better results and increasing the errors worsens them, although it can be seen that they are still acceptable.

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**, see Table 2. This gives an order of magnitude numerical complexity of the algorithm, albeit with a not fully optimized implementation. In particular, we give the computation times when no parallelization of the code is performed, although it is very easy to parallelize the loop on sigma points. Our strategy must solve  $N_\sigma$  ( $= N_s \times N_\theta + 1$  using simplex sigma-points) times a time step of the PDE system at each time step. Therefore – even if the estimation part was reduced to the parameters with the Kalman filter, which means, that the dimension of the covariance matrix goes to  $N_s \times (N_\theta + N_{\Delta x}) = N_s \times (2 + 2 \times \frac{1}{\Delta_x} + 1) = 43N_s$  to  $N_s N_\theta = 2N_s$  – the computation times are always strongly related to the size of  $N_{\Delta x}$ . To illustrate this, we also shrink the spatial step  $\Delta_x$  for  $N_s = 10$ . We can see that the computation times for a non-optimized algorithm are still reasonable and this shows the possibility of applying it to 2D or 3D PDE systems.

Finally, for comparison, we use the **Matlab** function *nlmefitsa*, which is an implementation of the nonlinear mixed effects model (NLME) with the stochastic EM algorithm, but only for  $N_s = 5$  because of the high computational cost. We report the mean of the computation times for 10 replicates. The rMSE – equal to 0.0067 (mean for only 10 replicates) – is very close to the rMSE – equal to 0.0074 (mean for 100 replicates) – obtained with our algorithm in exactly the same context.

### 3.3.3 Results using Algorithm 1 with Eq. (3b)

Second, we study the results obtained with Algorithm 1 when clustering is considered, meaning that the initial reduced covariance matrix  ${}^\sharp U_0$  and the initial extension matrix  ${}^\sharp L_0$  are given by Eq. (3b). We consider  $N_R = 100$  replicates of  $N_s = 20$  and  $N_s = 40$  spheroids for Scenario **w. pr., f. IC - so** and for  $\sigma_{err} = 0.08$ . We compare the strategy without using clustering and the strategy with using clustering with  $N_C = 5$  for  $N_s = 20$  and  $N_C = 10$  for  $N_s = 40$ . Table 3 shows the results. It can be seen that this strategy allows to reduce the computation times a bit, but it comes at a price. The rMSE, the rBIAS and the BMIXED, the ESTD are still comparable and very reasonable, but the fact that it couples more the subjects together could be seen on the STD – which is too small compared to the target value of 0.5 – and the COV also – which is too small compared to the target value of 95%. The larger the number of spheroids, the higher the price to be paid. The fact that the reduction in computational time is small in this context is not surprising. Eq. (3a) reduces the dimension of the covariance matrix from  $N_s \times (N_\theta + N_{\Delta x}) = N_s \times (2 + 2 \times \frac{1}{\Delta_x} + 1) = 43N_s$  to  $N_s \times N_\theta = 2N_s$  and Eq. (3b) on  $N_C \times N_\theta = 2N_C$ , which is not a big difference. This strategy is only interesting when  $N_s$  and  $N_\theta$  are very large and can also be used, for example, as a first step to improve priors.

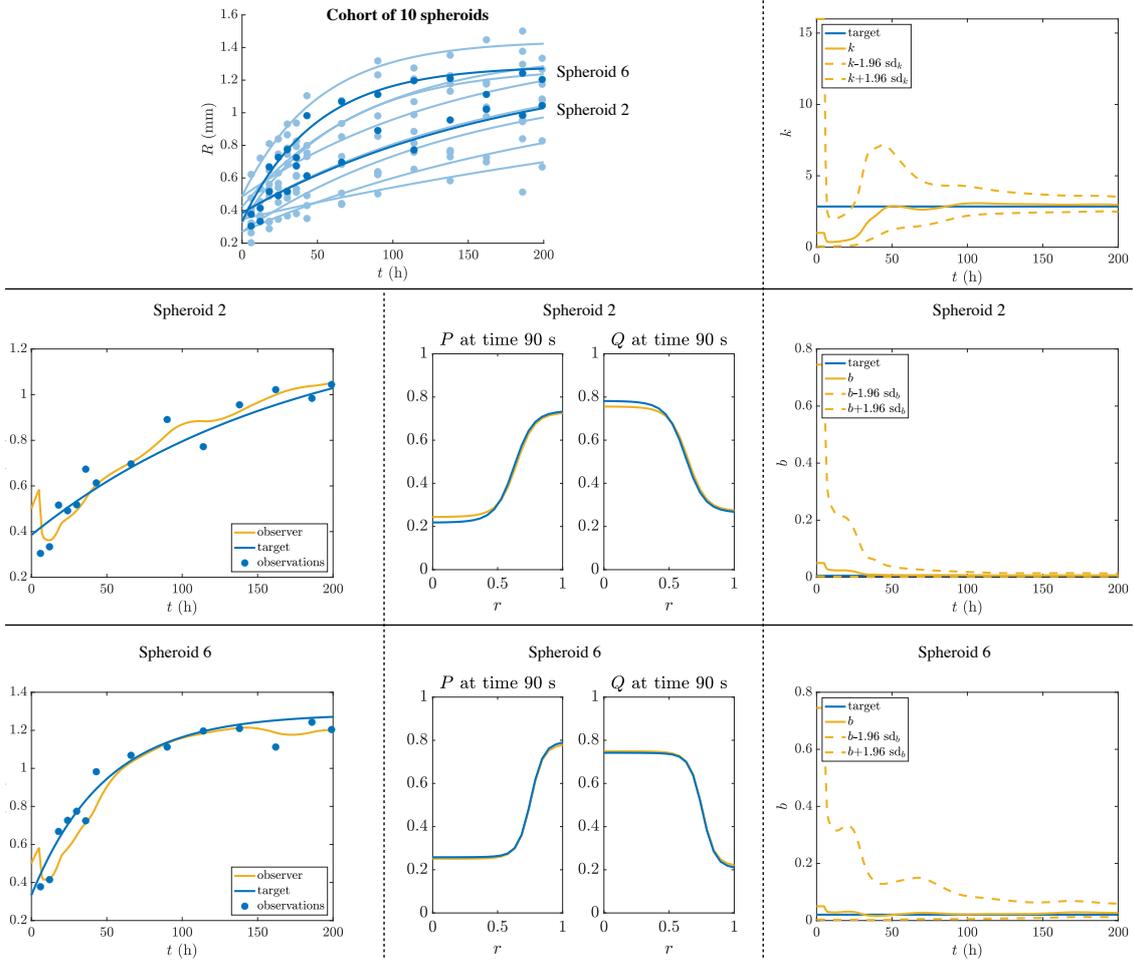


Figure 4: Time evolution of states and parameters for two spheroids using Algorithm 1 with  $N_S = 10$  and  $\sigma_{err} = 0.08$ . The target states and target values of the parameters are shown in blue. The blue points in the graphs of the radius correspond to the noisy observed data. The states and parameter values with population estimation are shown in yellow. The dashed lines correspond to the 95% confidence interval. First line, left: generated data of a cohort of the 10 spheroids. The two selected spheroids are in dark blue. First line, right: evolution of parameter  $k$  (assumed fixed in the cohort). Second (resp. third) line: Spheroid 2 (resp. 6) with time evolution state of variables  $R$  on the left, variables  $P$  and  $Q$  at a time step in the middle and parameter  $b$  on the right.

Algo	$N_S$	$N_C$	rMSE	rBIAS	COV	STD ( $b$ )	ESTD ( $b$ )	BMIXED ( $b$ )	CT (min)
Eq. (3a)	20	x	0.004	-0.0032	96%	0.4	0.57	0.066	4
Eq. (3b)	20	5	0.006	0.022	82%	0.18	0.4	0.077	3.2
Eq. (3a)	40	x	0.0027	-0.0019	95%	0.38	0.55	0.046	12
Eq. (3b)	40	10	0.0061	0.023	63%	0.13	0.38	0.052	10

Table 3: Performances of Algorithm 1 with and without clustering evaluated over 100 replicates in Scenario **w. pr., f. IC - so** when  $\sigma_{err} = 0.08$ . Results are aggregated across all parameters of the model for rMSE, rBIAS and COV. The computation times (CT) are also given.

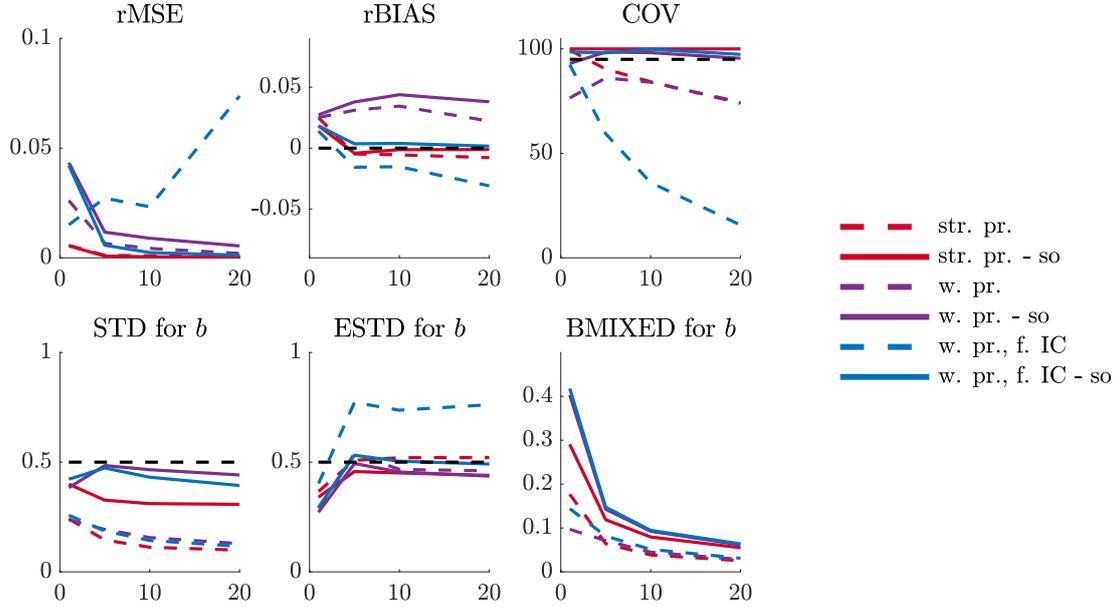


Figure 5: Performances of Algorithm 1 with Eq. (3a) evaluated over 100 replicates for the 6 scenarios when  $\sigma_{err} = 0$ . First line: rMSE, rBIAS and COV concatenated for both parameters  $b$  and  $k$  against the number of spheroids  $N_S$  considered. Second line: STD, ESTD and BMIXED for parameter  $b$  ( $k$  is fixed in the population) as a function of the number of considered spheroids  $N_S$ .

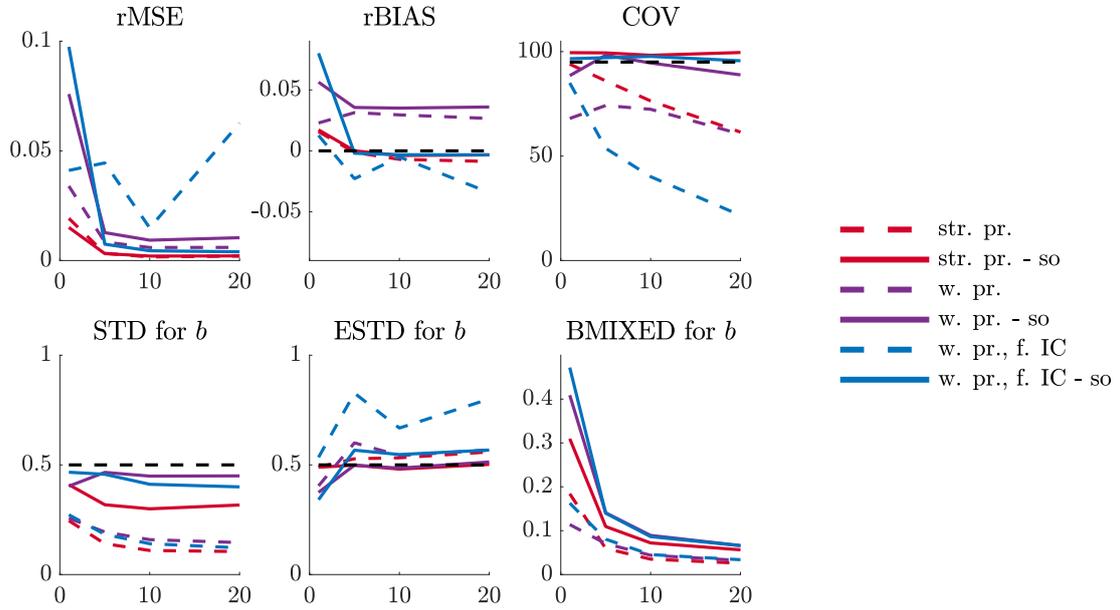


Figure 6: Performances of Algorithm 1 with Eq. (3a) evaluated over 100 replicates for the 6 scenarios when  $\sigma_{err} = 0.08$ . First line: rMSE, rBIAS and COV concatenated for both parameters  $b$  and  $k$  against the number of spheroids  $N_S$  considered. Second line: STD, ESTD and BMIXED for parameter  $b$  ( $k$  is fixed in the population) as a function of the number of considered spheroids  $N_S$ .

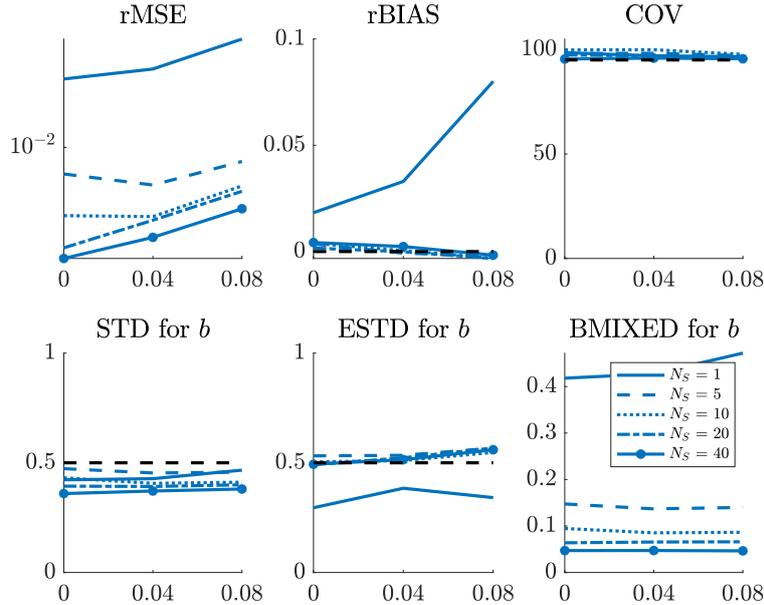


Figure 7: Performances of Algorithm 1 with Eq. (3a) evaluated over 100 replicates for Scenario **w. pr., f. IC - so**. First line: rMSE (log-scale), rBIAS and COV concatenated for both parameters  $b$  and  $k$  against the errors considered. Second line: STD, ESTD and BMIXED for parameter  $b$  ( $k$  is fixed in the population) as a function of the considered errors.

### 3.3.4 Anisotropic case

Before turning to the application to real data, we will perform final experiments on synthetic data to test the ability of our estimation strategy to deal with anisotropic disturbances of the parameters. We will consider Scenario **w. pr., f. IC - so** for 4 cases:  $\sigma_{err} = 0$  and  $0.08$  and  $N_S = 10$  and  $20$ . Only a few changes were made to the benchmark specified in Subsection 3.3.1. First, the two parameters are generated in a coupled manner meaning that

$$\begin{pmatrix} \log(R_0) \\ \text{logit}(P_0) \\ \log(b) \\ \log(k) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \log(m_{R_0}) \\ \text{logit}(m_{P_0}) \\ \log(m_b) \\ \log(m_k) \end{pmatrix}, \begin{pmatrix} \sigma_{R_0}^2 & 0 & 0 & 0 \\ 0 & \sigma_{P_0}^2 & 0 & 0 \\ 0 & 0 & \sigma_b^2 & \sigma_{b,k} \\ 0 & 0 & \sigma_{b,k} & \sigma_k^2 \end{pmatrix} \right),$$

where  $\sigma_{b,k}$  is the covariance of  $b$  and  $k$ . We consider the values given in Table 1 (target line) and  $\sigma_{b,k}$  is fixed at  $0.4^2$  ( $k$  is not fixed in the population in this context). Regarding the prior values for the estimation, only one value was changed:  $\sigma_{\hat{k}}^f = 0.5\sigma_{\hat{k}}$  as this parameter is no longer constant in the population. Table 4 summarizes the results, which are very good in terms of rMSE and rBIAS. Moreover, the results are very acceptable for the mean estimated standard deviations (STD) and very good for the standard deviation of estimates at convergence (ESTD) confirming the ability of our estimator to deal with anisotropic disturbances.

## 3.4 Application to real data

To conclude this study, we propose a validation using a real data set. Multicellular HCT-116-GFP spheroids were exposed to different electric field intensities – more specifically, 80 unipolar pulses of  $100 \mu\text{s}$  and a frequency of  $1 \text{ Hz}$  at  $0 \text{ V.cm}^{-1}$  (referred to as the control group) or  $500 \text{ V.cm}^{-1}$  (referred to as EF500 group) or  $1000 \text{ V.cm}^{-1}$  (divided into 2 groups due to their very different behaviour, referred to as EF1000-A and EF1000-B) or  $2000 \text{ V.cm}^{-1}$  (referred to as EF2000 group) – and their volumes were monitored by fluorescence and bright field microscopy. The experiments were designed and performed by Jelena Kolosnjaj, Muriel Golzio, and Marie-Pierre Rols of the IPBS Institute at the University of Toulouse and are described in detail in [3].

Criteria	True Value	$N_S = 10$	$N_S = 10$	$N_S = 20$	$N_S = 20$
		$\sigma_{err} = 0$	$\sigma_{err} = 0.08$	$\sigma_{err} = 0$	$\sigma_{err} = 0.08$
rMSE	×	0.0041	0.015	0.0040	0.016
rBIAS	×	0.0012	0.012	0.0010	0.0065
STD <sub><i>b</i></sub>	0.5	0.55	0.56	0.54	0.55
STD <sub><i>k</i></sub>	0.5	0.35	0.36	0.36	0.37
$(\text{Cov}_{b,k})^{\frac{1}{2}}$	0.4	0.34	0.36	0.37	0.37
ESTD <sub><i>b</i></sub>	0.5	0.47	0.51	0.49	0.54
ESTD <sub><i>k</i></sub>	0.5	0.53	0.56	0.51	0.55
$(\text{ECov}_{b,k})^{\frac{1}{2}}$	0.4	0.43	0.41	0.44	0.43

Table 4: Performances of Algorithm 1 with Eq. (3a) evaluated over 100 replicates for Scenario **w. pr., f. IC - so** for 4 cases:  $\sigma_{err} = 0$  and 0.08 and  $N_S = 10$  and 20. Results are aggregated across the two parameters of the model for rMSE and rBIAS. The six last lines correspond to the estimation of the variances and covariance of both parameters  $b$  and  $k$ .

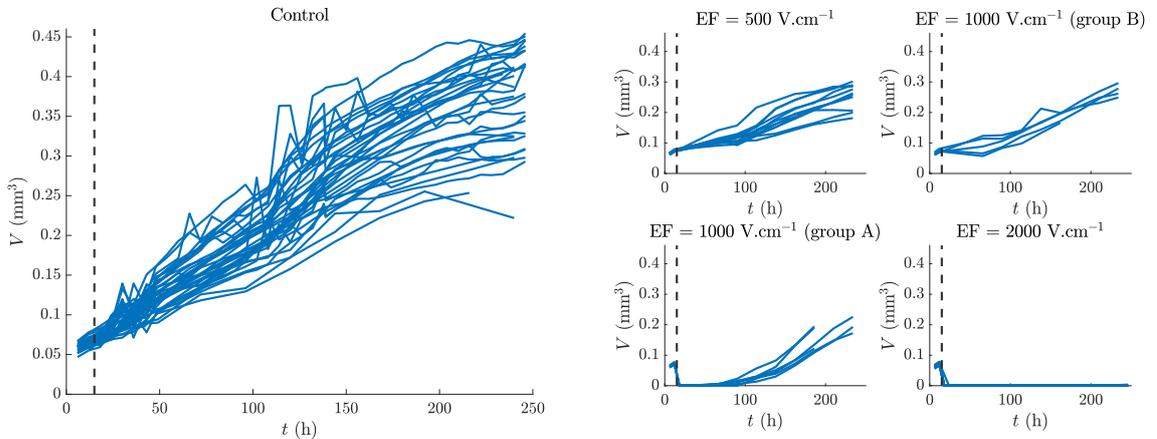


Figure 8: Volume evolutions of multicellular spheroids. Left: Control experiment (35 cases). Right: Electroporation experiment (37 cases divided into 4 categories: 12 for EF500, 6 for EF1000-A and 5 for EF1000-B and 14 for EF2000).

Very large volume values were measured because of the osmotic shock in the cells forming the spheroids after the electric shock. Because our model does not account for this phenomenon, these values are not used to estimate the parameters. Figure 8 shows the data used in this article. To avoid identifiability issues,  $b$  and  $k$  are estimated with the control cohort, and their mean values are used to estimate the other parameters in the 4 cases: EF500, EF1000A, EF1000B, and EF2000. Table 5 gives the estimated values. We compare the values obtained with the `Matlab` function `nlmefitsa` in exactly the same context. We can see that the results are very comparable. This last step with a real data set completes the validation of our algorithm on this example.

## 4 Conclusion

In this paper, we present a strategy to combine the principle of data assimilation through joint state and parameter estimation with the population configuration classically considered in the formulation of nonlinear mixed-effects models [12, 27, 13]. Due to the high computational cost, the classical strategies for solving nonlinear mixed-effects models are not applicable in practice for very large systems such as PDE systems. Our alternative estimation strategy to address this challenge is based on two important existing methodological strategies: (1) a population-based Kalman filter [6] and (2) a joint state-parameter estimation [19, 18]. We illustrated and evaluated

Case		Control ( $N_s = 35$ )	
$b$	our	0.0064 (mean) - 0.0017 (std)	
	Function <i>nlmefitsa</i>	0.010 (mean) - 0.0018 (std)	
$k$	our	2.28 (mean) - 0 (std)	
	Function <i>nlmefitsa</i>	2.19 (mean) - 0 (std)	

Case		EF500 ( $N_s = 12$ )	EF1000-B ( $N_s = 5$ )
$\lambda$	our	0.37 (mean) - 0.035 (std)	0.36 (mean) - 0.045 (std)
	Function <i>nlmefitsa</i>	0.22 (mean) - 0.092 (std)	0.37 (mean) - 0.012 (std)
$p_c$	our	0.90 (mean) - 0.0028 (std)	0.86 (mean) - 0.012 (std)
	Function <i>nlmefitsa</i>	0.89 (mean) - 0.014 (std)	0.82 (mean) - 0.086 (std)
$m$	our	1.14 (mean) - 0.18 (std)	1.21 (mean) - 0.16 (std)
	Function <i>nlmefitsa</i>	1.11 (mean) - (std) 0.024	1.40 (mean) - 0.16 (std)

Case		EF1000-A ( $N_s = 6$ )	EF-2000 ( $N_s = 14$ )
$\lambda$	our	0.69 (mean) - 0.04 (std)	0.92 (mean) - 0.0015 (std)
	Function <i>nlmefitsa</i>	0.77 (mean) - 0.10 (std)	0.97 (mean) - 0.065 (std)
$p_c$	our	0.11 (mean) - 0.015 (std)	0.00097 (mean) - $2.1 \times 10^{-6}$ (std)
	Function <i>nlmefitsa</i>	0.17 (mean) - 0.010 (std)	0.0016 (mean) - $9 \times 10^{-5}$ (std)
$m$	our	5.3 (mean) - 0.77 (std)	x
	Function <i>nlmefitsa</i>	5.19 (mean) - 1.02 (std)	x

Table 5: Mean (mean) and standard deviation (std) of estimated parameters using real data. The values were compared with the results obtained with the NLME algorithm using the `Matlab` function *nlmefitsa*. The post-pulse data for EF2000 do not really allow us to estimate the  $m$  parameter, since the data for all spheroids are close to 0. Therefore, this parameter is not estimated.

our proposed strategy on a 1D PDE model for tumor spheroid electroporation. Our approach was shown to be successful on synthetic data even with anisotropic disturbances. Moreover, on real data, our results compare very well with the estimation procedures used in the literature as standard for estimating parameters of nonlinear mixed-effects models.

The choice of the 1D PDE model stems from the fact that the computational cost of solving the PDE system is still reasonable and allows to perform many replicates – 7500 simulations in total<sup>1</sup> – to validate the ability to estimate the parameters well. Even though our strategy is very efficient in terms of computation time, we need to solve the PDE system  $N_\sigma$  (equivalent to  $N_s \times N_\theta + 1$  when simplex sigma points are considered) at each time step to apply it. This implies that we recommend the use of an efficient solver for 2D or 3D PDE systems.

The strategy of double reduction of the Kalman filter (reduction to parameter space and reduction by clustering) seems interesting only when  $N_s$  and  $N_\theta$  are very large. Since it seems to worsen the results, we advise to use it cautiously. We emphasize that it can be used, for example, as a first step to improve the prior values.

It is clear that using the state observer in conjunction with the Kalman observer for the parameters leads to better results when weak priors are considered and especially when the initial conditions are false. This implies that we recommend the use of a state observer especially when the confidence in the initial conditions is low. We are aware that building a state observer to handle errors in the initial conditions can be a very difficult step and is the main drawback of our strategy.

Having demonstrated the concept of our strategy and these performances in terms of accuracy, manageability and computation times, the main perspective of this work is to develop a code that can be easily used by the community.

<sup>1</sup>(6 scenarios  $\times$  4 numbers of subjects ( $N_s = \{1, 5, 10, 20\}$ ) + 1 scenario  $\times$  1 number of subject ( $N_s = 40$ ))  $\times$  3 considered errors ( $\sigma_{err} = \{0, 0.04, 0.08\}$ )  $\times$   $N_R = 100$  replicates

## 5 Acknowledgement

The author thanks Philippe Moireau for scientific discussions and Jelena Kolosnjaj, Muriel Golzio, and Marie-Pierre Rols for permission to use the data originally presented in [3].

## References

- [1] R. Bellman. Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America*, 42(10):767, 1956.
- [2] 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.
- [3] A. Collin, H. Bruhier, J. Kolosnjaj, M. Golzio, M.-P. Rols, and C. Poinard. Spatial mechanistic modeling for prediction of 3D multicellular spheroids behavior upon exposure to high intensity pulsed electric fields. *AIMS Bioengineering*, 9(2):102–122, 2022.
- [4] A. Collin, D. Chapelle, and P. Moireau. A Luenberger observer for reaction–diffusion models with front position data. *Journal of Computational Physics*, 300:288–307, 2015.
- [5] A. Collin, B. P. Hejblum, C. Vignals, L. Lehot, R. Thiébaud, P. Moireau, and M. Prague. Using population based Kalman estimator to model COVID-19 epidemic in france: estimating the effects of non-pharmaceutical interventions on the dynamics of epidemic. *International Journal of Biostatistics*, 2022.
- [6] A. Collin, M. Prague, and P. Moireau. Estimation for dynamical systems using a population-based Kalman filter–applications in computational biology. *MathematicS In Action*, 2022.
- [7] M. J. Denwood. runjags: A R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(9):1–25, 2016.
- [8] S. B. Duffull, C. M. J. Kirkpatrick, B. Green, and N. H. G. Holford. Analysis of population pharmacokinetic data using NONMEM and WinBUGS. *Journal of Biopharmaceutical Statistics*, 15(1):53–73, 2004.
- [9] E. Grenier, V. Louvet, and P. Vigneaux. Parameter estimation in non-linear mixed effects models with SAEM algorithm: extension from ODE to PDE. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(5):1303–1329, 2014.
- [10] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.
- [11] E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038, 2005.
- [12] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.
- [13] M. Lavielle. *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014.
- [14] K. Law, A. Stuart, and K. Zygalakis. *Data assimilation: A mathematical introduction*, volume 62 of *Texts in Applied Mathematics*. Springer, Cham, 2015.
- [15] 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.

- [16] D. Luenberger. *Determining the State of a Linear with Observers of Low Dynamic Order*. PhD thesis, Stanford University, 1963.
- [17] T. Michel, J. Fehrenbach, V. Lobjois, J. Laurent, A. Gomes, T. Colin, and C. Poignard. Mathematical modeling of the proliferation gradient in multicellular tumor spheroids. *Journal of theoretical biology*, 458:133–147, 2018.
- [18] 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.
- [19] P. Moireau, D. Chapelle, and P. Le Tallec. Joint state and parameter estimation for distributed mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 1987(6–8):659–677, 2008.
- [20] D. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207, 2001.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] M. Prague, D. Commenges, J. Guedj, J. Drylewicz, and R. Thiébaud. NIMROD : A program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations. *Computer methods and programs in biomedicine*, 111(2):447–458, 2013.
- [25] D. Simon. *Optimal State Estimation: Kalman,  $H^\infty$ , and Nonlinear Approaches*. Wiley-Interscience, 2006.
- [26] C. Vaghi, A. Rodallec, R. Fanciullino, J. Ciccolini, J. P. Mochel, M. Mastri, C. Poignard, J. M. Ebos, and S. Benzekry. Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLoS computational biology*, 16(2):e1007178, 2020.
- [27] G. Verbeke. Linear mixed models for longitudinal data. In *Linear mixed models in practice*, pages 63–153. Springer, 1997.
- [28] H. Wu. Statistical methods for HIV dynamic studies in aids clinical trials. *Statistical methods in medical research*, 14(2):171–192, 2005.

## A Mathematical proofs

### A.1 Proof of Section 3.1

The proof of Proposition 1 is given in what follows.

**Proof First part** - Using the characteristic curves defined by the velocity field  $V$  and the assumption on the initial condition, it is easy to prove that  $P + Q$  belongs to  $[0, 1]$  as the right-hand side of

$$\partial_t(P + Q) + V\partial_r(P + Q) = \tau_G(P + Q)(1 - (P + Q))$$

vanishes for  $P + Q = 0$  and  $P + Q = 1$ . With this fact, it is then easy to prove that  $P \in [0, 1]$  as the right-hand side of the equation verified by  $P$

$$\partial_t P + V\partial_r P = \tau_G(P + Q)(1 - P) - \tau_O P,$$

is positive for  $P = 0$  and non-positive for  $P = 1$ . This directly implies that  $Q \in [0, 1]$  (as  $P \in [0, 1]$  and  $P + Q \in [0, 1]$ ) and  $F \in [0, 1]$  (as  $F = 1 - (P + Q)$ ).

**Second part -** As  $(P + Q)(t, r) \in [0, 1], \forall t \geq 0, \forall r \in [0, 1]$ , one can show that

$$0 \leq I(t, 1) = \int_0^1 (P(t, r) + Q(t, r))r^2 dr \leq \frac{1}{3}. \quad (9)$$

It then implies that  $R'(t), R(t) \geq 0, \forall t$  and more precisely that

$$0 \leq R' \leq \frac{ma_{bs}e^{-bt}}{3}R.$$

As the solution of  $\underline{R}'(t) = \frac{ma_{bs}e^{-bt}}{3}\underline{R}(t)$  is the solution of the Gompertz model given by

$$\underline{R}(t) = \underline{R}(0)e^{\frac{ma_{bs}}{3b}(1-e^{-bt})},$$

it concludes the proof.  $\square$

## A.2 Proofs of Section 3.2

Before to start the proofs of Propositions 2 and 3, we need the following preliminary lemma.

**Lemma 1** *We have  $\hat{F} = F$ . This directly implies that  $\hat{I} = I$ .*

**Proof** The last equation of System (7) and the last equation of System (8) are the same as the velocity only depends on the density  $F$  or  $\hat{F}$ . The initial conditions  $F(0, r) = 0$  and  $\hat{F}(0, r) = 0$  imply that for all  $t < t_{as}$ ,  $F(t, r) = \hat{F}(t, r) = 0$ . The key point of the proof lies in the fact at  $t = t_{as}$ , as  $F$  (or  $\hat{F}$ ) equals to 0,  $P + Q = 1$  (or  $\hat{P} + \hat{Q} = 1$ ) then  $F(t_{as}, r) = \hat{F}(t_{as}, r) = \lambda, \forall r \in [0, 1]$ . By unicity, we then have  $F(t, r) = \hat{F}(t, r), \forall t \geq 0, \forall r \in [0, 1]$ .  $\square$

**Remark 5** *All the following proofs are done in the context of  $t_{as} = 0$  just for the sake of simplicity.*

Using Lemma 1 and the fact that  $\tau_O P - \hat{\tau}_O \hat{P} = ((\tau_O - \hat{\tau}_O)P) + \hat{\tau}_O \tilde{P}$ , our objective is to study the solution of the following system of equations

$$\begin{cases} \tilde{R}' &= \tilde{R}\tau_G I(t, 1) - \gamma_{\text{obs}}(t)\tilde{R}, \\ \partial_t \tilde{P} &= -\tau_G(r^{-2}I(t, r) - rI(t, 1))\partial_r \tilde{P} - \tau_G(1 - F)\tilde{P} - ((\tau_O - \hat{\tau}_O)P) + \hat{\tau}_O \tilde{P}. \end{cases}$$

with

$$I(t, r) = \int_0^r (1 - F)r^2 dr.$$

This system can be rewritten as

$$\begin{cases} \tilde{R}' &= \tilde{R}\tau_G I(t, 1) - \gamma_{\text{obs}}(t)\tilde{R}, \\ \partial_t \tilde{P} + V\partial_r \tilde{P} &= -\tau_G(1 - F)\tilde{P} - ((\tau_O - \hat{\tau}_O)P) + \hat{\tau}_O \tilde{P}, \\ V &= \tau_G(r^{-2}I(t, r) - rI(t, 1)). \end{cases} \quad (10)$$

Now, one can prove Proposition 2.

**Proof** The solution of the equation on  $\tilde{R}$  is

$$\tilde{R}(t) = \tilde{R}(0) \exp\left(\int_0^t (\tau_G(\underline{t})I(\underline{t}, 1) - \gamma_{\text{obs}}(\underline{t}))d\underline{t}\right), \quad \forall t \geq 0. \quad (11)$$

Using Eq. (9) and the formula of  $\tau_G(\underline{t}) = ma_{bs}e^{-b\underline{t}}$ , one can show that

$$0 \leq \int_0^t (\tau_G(\underline{t})I(\underline{t}, 1) - \gamma_{\text{obs}}(\underline{t}))d\underline{t} \leq t\left(\frac{ma_{bs}}{3} - \gamma_{\text{obs}}^{\min}\right)$$

leading to

$$0 \leq |\tilde{R}(t)| \leq |\tilde{R}(0)|e^{(\frac{ma_{bs}}{3} - \gamma_{\text{obs}}^{\min})t}. \quad (12)$$

$\square$

Now, one can prove Proposition 3.

**Proof** We have

$$\begin{aligned} (\tau_O - \hat{\tau}_O) &= (\mathcal{T}_O^{begin} - \mathcal{T}_O^{end}) \left( \frac{1}{1 + e^{\frac{(\hat{R}(1-r)-d)}{s}}} - \frac{1}{1 + e^{\frac{(R(1-r)-d)}{s}}} \right), \\ &= (\mathcal{T}_O^{begin} - \mathcal{T}_O^{end}) \frac{e^{\frac{(1-r)(R+\hat{R})-2d}{2s}}}{(1 + e^{\frac{(R(1-r)-d)}{s}})(1 + e^{\frac{(\hat{R}(1-r)-d)}{s}})} \sinh \left( \frac{(1-r)\hat{R}}{2s} \right). \end{aligned}$$

The term  $(\mathcal{T}_O^{begin} - \mathcal{T}_O^{end}) \frac{e^{\frac{(1-r)(R+\hat{R})-2d}{2s}}}{(1 + e^{\frac{(R(1-r)-d)}{s}})(1 + e^{\frac{(\hat{R}(1-r)-d)}{s}})}$  is bounded since  $R$  and  $\hat{R}$  are bounded as  $R$  is bounded and  $\hat{R}$  converge to  $R$ . Using the mean value theorem and Eq. (12), one can show that

$$|\tau_O - \hat{\tau}_O| \leq C|\tilde{R}| \leq Ce^{(\frac{mabs}{3} - \gamma_{obs}^{\min})t}. \quad (13)$$

Multiplying by  $\tilde{P}$  and integrating on  $r \in ]0, 1[$  the second equation of System (10), we obtain

$$\frac{1}{2} \frac{d}{dt} \|\tilde{P}\|_{L^2(]0,1])}^2 = - \int_0^1 \tau_G(1-F)\tilde{P}^2 dr - \int_0^1 (\tau_O - \hat{\tau}_O)P\tilde{P} dr - \int_0^1 \hat{\tau}_O\tilde{P}^2.$$

Using the fact that  $0 \leq F(t, r) \leq F_0, \forall t, \forall r \in ]0, 1[$ , we have

$$\frac{1}{2} \frac{d}{dt} \|\tilde{P}\|_{L^2(]0,1])}^2 = - \int_0^1 (ae^{-bt}(1-F_0) + \mathcal{T}_O^{end})\tilde{P}^2 dr - \int_0^1 (\tau_O - \hat{\tau}_O)P\tilde{P} dr.$$

The last term is bounded using Eq. (13)

$$\begin{aligned} - \int_0^1 (\tau_O - \hat{\tau}_O)P\tilde{P} dr &\leq \int_0^1 |\tau_O - \hat{\tau}_O|P|\tilde{P}| dr \\ &\leq \int_0^1 |\tau_O - \hat{\tau}_O| \\ &\leq Ce^{(\frac{mabs}{3} - \gamma_{obs}^{\min})t}, \end{aligned}$$

leading to

$$\frac{1}{2} \frac{d}{dt} \|\tilde{P}\|_{L^2(]0,1])}^2 \leq -\mathcal{T}_O^{end} \|\tilde{P}\|_{L^2(]0,1])}^2 + Ce^{-(\gamma_{obs}^{\min} - \frac{mabs}{3})t}.$$

The solution of the ODE  $y' = C_1y + C_2e^{C_3t}$  is given by  $y = \underline{C}e^{C_1t} + \frac{C_2}{C_3 - C_1}e^{C_3t}$  if  $C_3 \neq C_1$ . With  $C_1 = -\mathcal{T}_O^{end}$ ,  $C_2 = C$  and  $C_3 = \frac{mabs}{3} - \gamma_{obs}^{\min}$ , we have  $C_1 > C_3$ .

$$\|\tilde{P}\|_{L^2(]0,1])}^2 \leq C(e^{-\mathcal{T}_O^{end}t} + e^{-(\gamma_{obs}^{\min} - \frac{mabs}{3})t}) \leq Ce^{-\mathcal{T}_O^{end}t}.$$

□

## B Statistical validation criteria

We quickly define the statistical validation criteria used to validate the ability of our estimator to estimate parameters. For a given parameter  $\theta$ , we define by  $\theta_t$  (respectively,  $\theta_e$ ) the matrix of size  $N_R \times N_S$  that contains in each row  $r$  the target (respectively, estimated) values of  $\theta$  for all subjects of replicate  $r$ . We calculate the relative bias (rBIAS) and the relative mean squared error (rMSE)

$$\text{rBIAS} = \frac{1}{N_R N_S} \sum_{r=1}^{N_R} \sum_{i=1}^{N_S} \frac{\theta_t(r, i) - \theta_e(r, i)}{|\theta_t(r, i)|} \quad \text{and} \quad \text{rMSE} = \frac{1}{N_R N_S} \sum_{r=1}^{N_R} \sum_{i=1}^{N_S} \left( \frac{\theta_t(r, i) - \theta_e(r, i)}{\theta_t(r, i)} \right)^2.$$

Using the estimated covariance matrix, we compute  $\sigma_e^\theta$  the matrix of size  $N_R \times N_S$  that contains in each row  $r$  the estimated standard deviation of  $\theta$  for all subjects of replicate  $r$ . We also compute the mean estimated standard deviation (STD) and the empirical standard deviation of estimates, defined as the standard deviation of estimates at convergence (ESTD)

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

$$\text{and ESTD} = \sqrt{\frac{1}{N_R N_S} \sum_{r=1}^{N_R} \sum_{i=1}^{N_S} \left( \theta_e(r, i) - \frac{1}{N_R N_S} \sum_{n=1}^{N_R} \sum_{i=1}^{N_S} \theta_e(r, i) \right)^2}.$$

Following the same idea, we also define for two parameters  $\theta^1$  and  $\theta^2$ , the mean estimated covariance (Cov) and the empirical covariance of estimates (ECov) by

$$\text{Cov} = \frac{1}{N_R N_S} \sum_{r=1}^{N_R} \sum_{i=1}^{N_S} \sigma_e^{\theta^1, \theta^2}(r, i)$$

$$\text{and ECov} = \sqrt{\frac{1}{N_R N_S} \sum_{r=1}^{N_R} \sum_{i=1}^{N_S} \left( \theta_e^1(r, i) - \frac{1}{N_R N_S} \sum_{n=1}^{N_R} \sum_{i=1}^{N_S} \theta_e^1(r, i) \right) \left( \theta_e^2(r, i) - \frac{1}{N_R N_S} \sum_{n=1}^{N_R} \sum_{i=1}^{N_S} \theta_e^2(r, i) \right)},$$

where  $\sigma_e^{\theta^1, \theta^2}$  the matrix of size  $N_R \times N_S$  that contains in each row  $r$  the estimated covariance between  $\theta^1$  and  $\theta^2$  for all subjects of replicate  $r$ . Finally, we can calculate the relative bias of the standard deviation for random effects, called BMIXED, defined by

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

where  $\sigma_{e,m}^\theta$  is the matrix of size  $N_R \times N_S$  containing in each row  $r$  the estimated mixed standard deviation of  $\theta$  for all subjects of replicate  $r$  and therefore formed using the estimated covariance matrix. 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.