Ensemble Kalman Filter and Extended Kalman Filter for State-Parameter Dual Estimation in Mixed Effects Models defined by a Stochastic Differential Equation

José Soto[1,2] and Saba Infante[1,3]

[1] Escuela de Ciencias Matemáticas y Computacionales, Universidad Yachay Tech, Ecuador.
[2] Universidad de los Andes, Venezuela.
[3] Departamento de Matemáticas, FACyT, Universidad de Carabobo, Venezuela.

Abstract. The biological processes that occur in the real world have complex dynamics. Mathematical models that try to describe these phenomena have nonlinear structures, observations are made at discrete time points and include measurement errors, and are difficult to estimate. In particular, when modeling dynamics of repeated measurements on individuals or objects, they are analyzed by mixed-effects diffusion models. The standard estimation methods in these cases are: maximum likelihood, EM, SAEM, Newton Rapson, among others. In this paper we propose a specific inference methodology for models. Apply extended Kalman Filter (EKF) and the ensemble Kalman filter (EnKF) to the estimation of both states and parameters of nonlinear state-space models. To illustrate the methodology, the states and parameters of an Ornstein-Uhlembeck (O-U) mixed-effects model were estimated, obtaining precise estimates with small standard deviations. To measure the estimation quality of the algorithms was used as a measure of goodness-of-fit known as the square root of the mean quadratic error, obtaining very small errors.

Keywords: mixed effects models, stochastic differential equation, dual extended kalman filter, dual ensemble kalman filter.

1 Introduction

Many biological process experiments have repeated responses that are measured over time over an individual or groups of individuals, or objects, and that are obtained through a design of experiments randomized block design, least square designs, nested designs, repeated measurement designs, longitudinal data design previously established; these observations arise in many fields of research: clinical trials, agricultural growth studies, tumor growths, biological process curves, financial series, among other applications. The classical statistical models used to analyze this type of information are mixed models; which include terms of random effects, components of variance and associated covariance. A generalization of these models are the models mixtures defined by a stochastic differential equation (SDE), which can be seen as an extension of a mixed Markov chain with finite state space, for space state and continuous time. These models cover a wider class of problems, especially with irregular observation times and at different time scales. Despite their great interest, these models are really difficult to study and handle, and statistical inference becomes a challenging subject. Recently in the statistical literature methodologies have been developed based on mixed-effect diffusion models where the diffusion process satisfies the SDE type Itô; that is, there is a class of continuous time stochastic processes that models the SDE solution as a first-order Markov process. This methodology has been applied in: system dynamics [1]; experiments in the pharmaceutical industry ([6],[26] and [14] among others); in neuroscience ([12]); in prognosis of solar and wind energy ([19]); growing cracks ([15]). Other applications include analysis of transmitted diseases such as such as epidemics, financial series, population growth dynamics ([18]), and intracellular processes ([29]). In particular; [10] carried out an analysis of a continuous process of growth curves using mixed models defined in terms of an SDE, they proposed Bayesian techniques to make the inferences and estimate the parameters; In addition, they took into consideration cases when the SDE have an explicit solution using a Gibbs algorithm; and when the conditional distribution of the diffusion process has no explicit form,they use an Euler-Maruyama scheme to make the approximations. [5] combine a SAEM algorithm with the extended Kalman filter (EKF) to estimate the parameters of a mixed-effects model; additionally, they provided tools for the recovery of the trajectories of the underlying diffusion of the model; The methodology is illustrated by simulations and using data from the pharmaceutical industry.[4] modeled panel data using a mixed-effects model with heterogeneity in individual random effects. They compared two approaches using mixtures of Gaussian and Dirichlet distributions; The methodology, it was illustrated using synthetic data and real data to obtain the a posterior estimates of the parameters. [29] propose a scheme Monte Carlo Markov chains (MCMC) in the scenario of mixed effects diffusion models.The proposal includes an algorithm that allows efficient sampling of marginal distributions, which may have linear nonlinear dynamics between observation times, they illustrate the proposal using synthetic data from an SDE model of orange tree growth and using data real results obtained from pest control in cotton plants under different treatment regimes. [27] proposed a mixed effects model of an SDE to model the growth dynamics of repeated measurements of tumor volumes in mice, they considered two sections divided into a representation through fractions of dead tumors and those that survive a treatment; they estimated the model parameters using an exact Bayesian methodology and an approximate method. Recently, [24] studied a general mixed effects model, where the variability of the random effects of the individuals or experimental units were incorporated through an SDE, in that work was implemented a Monte Carlo Markov chains algorithm to estimate the parameters. A diagnos-

tic analysis was performed on the estimated parameters to detect if the model is adequate and show its convergence.
The rest of the article follows: in Session 2 the problem is formulated, in Session 3 the methodology is established, in Session 4 the results are shown and in Session 5 a discussion and conclusions are established.

2 Problem formulation

Let $y_{ij}, i = 1, \ldots, I, j = 0, \ldots, J_i$ where y_{ij} be the $j-$th observation of the unit i at time t_{ij}, where $t_{i0} \leq t_{i0} \leq \cdots \leq t_{iJ_i} \leq T$. If we assume that y_{ij} is the noisy observation of a latent time-dependent process x_{ij}, then

$$y_{ij} = x(t_{ij}, \phi_i) + \epsilon_{ii} \quad , \quad \epsilon_{ij} \sim N(0, \sigma^2) \tag{1}$$

where $x(t_{ij}, \phi_i)$ is the realization of a diffusion process defined as the solution of a SDE describing the observed dynamic process

$$dx(t, \phi) = a(x, t, \phi)dt + b(x, t, \phi, \gamma)dB(t)$$

with $\{B_t\}$ is a collection of independent Brownian movements, $a(x, t, \phi)$ and $b(x, t, \phi, \gamma)$ are the functions of drift and volatility. The mixed effect SDE model, in its compact form is summarized

$$\begin{aligned} y_{ij} &= x(t_{ij}, \phi_i) + \epsilon_{ij}, \qquad\qquad \phi_i \sim p(\phi|\beta) \quad\; \epsilon_{ij} \sim N(0, \sigma^2) \\ dx(t, \phi_i) &= a(x, t, \phi_i)dt + b(x, t, \phi_i, \gamma)dB(t) \\ x(t_0, \phi_i) &= x_0(\phi_i) \qquad \text{for } 1 \leq i \leq I \quad , 0 \leq j \leq J_i \end{aligned} \tag{2}$$

where $\phi_i \in \mathbb{R}^d$ is the individual parameter for subject i, randomly distributed with the density $p(\phi|\beta)$, depending on the parameter $\beta \in \mathbb{R}^p$, $\epsilon_{ij} \in \mathbb{R}$ represents the measurement error, with a measurement noise variance σ^2, The regression term $x(t_{ij}, \phi_i)$ for subject i is a realization of the diffusion process $x : \mathbb{R} \to \mathbb{R}$ defined by equation (2), $x_0(\phi_i)$ is a initial condition, $B(t)$ is a collection of independent Brownian motions, and are independent of ϵ_{ij}, $a\,(x, \theta) : \mathbb{R}^d \times \mathbb{R}^p \longrightarrow \mathbb{R}^d$ is a drift function, $b\,(x, \theta) : \mathbb{R}^d \times \mathbb{R}^p \longrightarrow M_{d\times d}(\mathbb{R})$ is a volatility function, where $M_{d\times d}(\mathbb{R})$ is the set of $d \times d$ matrices with real elements, are known functions depending on an unknown parameter $\theta \in \mathbb{R}^d$. The parameters of interest are $\theta = (\beta, \gamma, \sigma)$ and the states solutions $x(t, \phi_i)$. Statistical inference makes sense only if the existence and uniqueness of a solution of the SDE (2) for all $x(t_0)$, ϕ and $b(.)$ is ensured. Sufficient conditions of existence and uniqueness are the following globally Lipschitz, linear growth and boundedness conditions (Donnet and Samson (2008)). To obtain an approximation of the transition density, the Euler-Maruyama approximation can be used,([7]). Let $t_0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots < \tau_N = t_J$ denotes a discretization in the interval $[t_0, t_J]$. Suppose that for all $j = 0, \ldots, J$, exist an integer n_j what verifies $t_j = t_{n_j}$ with $n_0 = 0$. Let $(\Delta t_n)_{1\leq n\leq N}$ a sequence of steps where $\Delta t_n = \tau_n - \tau_{n-1}$. Let $\Delta = \max_{1\leq n\leq N} \Delta t_n$ the maximum step; then the diffusion process that approximates the solution x denoted by $\tilde{x}$ of the SDE that is obtained by the numerical method Euler-Maruyama is given by

$$\tilde{x}_{i,n} = \tilde{x}_{i,n-1} + \Delta t_n a(\tilde{x}_{i,n-1}, \tau_{n-1}, \phi_i) + b(\tilde{x}_{i,n-1}, \tau_{n-1}, \phi_i, \gamma)\sqrt{\Delta t_n}\epsilon_n$$

where $\epsilon_n \sim N(0,1)$ and $\tilde{x}_{i,n}$ denotes the realization of a stochastic process in time τ_n for parameters ϕ_i; thus $(\tilde{x}_{i,n_0}, \tilde{x}_{i,n_1}, \ldots, \tilde{x}_{i,n_J})$ is an approximation of the original process $(x(t_0,\phi_i)(x(t_1,\phi_i),\ldots,(x(t_J,\phi_i))$. The statistical model is now

$$\begin{aligned}
y_{ij} &= \tilde{x}_{i,n_j} + \epsilon_{ij} \qquad \phi_i \sim p(\phi|\beta) \qquad \epsilon_{ij} \sim N(0,\sigma^2) \\
\tilde{x}_{i,n} &= \tilde{x}_{i,n-1} + \Delta t_n a(\tilde{x}_{i,n-1}, \tau_{n-1}, \phi_i) \\
&\quad + b(\tilde{x}_{i,n-1}, \tau_{n-1}, \phi_i, \gamma)\sqrt{\Delta t_n}\epsilon_n \qquad \epsilon_n \sim N(0,1) \\
x(t_0,\phi_i) &= x_0(\phi_i) \quad \text{for } 1 \le i \le I, \quad 0 \le j \le J, 1 \le n \le N
\end{aligned} \tag{3}$$

The main contribution of this research consists of implementing two algorithms: the extended Kalman filter (EKF) and the Ensamble Kalman filter (EnKF) for dual estimation of states and parameters in a mixed effects model defined by an SDE. To illustrate the methodology, an SDE that arises from the field of finance is considered. The Ornstein Uhlenbeck (O-U) equation in N dimensions posits that the N dimensional state vector X satisfies the SDE

$$dx_t = (B - Ax_t)dt + CdW_t$$

where B is an $N \times 1$ vector, A is a nonsingular $N \times N$ matrix and C is an $N \times M$ matrix while dW is the differential of the M dimensional Wiener process W_t. Classical Calculus suggests that Itô's lemma should be applied to

$$\omega(t) = \exp(At)(x_t - A^{-1}B)$$

The result is

$$\begin{aligned}
d\omega(t) &= A\exp(At)\left(x_t - A^{-1}B\right)dt + \exp(At)dx_t \\
&= A\exp(At)\left(x_t - A^{-1}B\right)dt + \exp(At)\left((B - Ax_t)dt + CdB_t\right) \\
&= C\exp(At)dB_t
\end{aligned}$$

Integration over $[0,t]$ now gives $\omega(t) = \omega(0) + \int_0^t C\exp(At)dB_t$; substituting $\omega(t)$ and $\omega(0)$, is obtained $\exp(At)(x_t - A^{-1}B) = (x_0 - A^{-1}B) + \int_0^t C\exp(At)dB_t$. So the SDE solution is $x_t = (I - \exp(-At))\,A^{-1}B + x_0\exp(-At) + \int_0^t C\exp\left(-A(t-s)\right)dB_s$ The solution is a multivariate Gaussian process with mean and covariance given by $\mu_x = (I - \exp(-At))\,A^{-1}B + x_0\exp(-At)$, and

$$\begin{aligned}
\Sigma_x &= \mathbb{E}\left\{\left(\int_0^t C\exp\left(-A(t-s)\right)dB_s\right)^T\left(\int_0^t C\exp\left(-A(t-s)\right)dB_s\right)\right\} \\
&= \int_0^t\int_0^t \mathbb{E}\left\{\left(C\exp\left(-A(t-s)\right)dB_s\right)\left(C\exp\left(-A(t-s)\right)dB_s\right)^T\right\} \\
&= \int_0^t \exp\left(-Au\right)CQC^T\exp\left(-Au\right)du
\end{aligned}$$

3 Methodology for estimating states and parameters

The methods of estimating mixed-effects models using an SDE require evaluating the likelihood function of data, which usually can not be known in closed form, because it is necessary to explicitly determine the transition density. Following the notation given in [27], let $y_i = \{y_{ij}\}_{j=1,\ldots,n_i}$ denote the observations for subject i and by $x_i = \{x_{ij}\}_{j=1,\ldots,n_i}$ the corresponding values of the latent process. Let $y_{1:t} = (y_1, \ldots, y_t)$ denote the full dataset containing measurements for all subjects in the considered group. Standard methods for frequentist as well as Bayesian estimation of the model parameters, require the evaluation of the likelihood function $p(y_{1:t}|\theta) = \prod_{k=1}^{t} p(y_i|\theta)$. The Markov structure implies the following $p(y_i|\theta) = pos$

$$pos = \int\int \prod_{j=0}^{n_i} p(y_{ij}|x_{ij}, \phi_i, \theta)p(x_{ij}|x_{i,j-1}, \phi_i, \theta)p(x_{i,0}|\phi_i, \theta)dx_{ij}p(\phi_i|\theta)d\phi_i \quad (4)$$

Note that the term $p(x_{i,0}|\phi_i, \theta)$ vanish in either of the cases where $x_{i,0}$ is included among the random effects or is assumed to be a known constant $x_{i,0} = x_0$. In general the likelihood function (4) is not analytically tractable by conventional methods. Thus, inference for those models rely on either more specific assumptions such as the latent processes being Gaussian, or the use of Bayesian methods. Some methods have been used in the literature are: EM algorithm (SAEM) ([5],[27]), particle MCMC algorithm ([9]), and MCMC algorithms ([10],[29] and [24], among others). In this work we propose two alternative algorithms to carry out the estimation.

3.1 Extended Kalman filter (EKF) for dual estimations

When inferring on combined mixed models using a SDE, it is necessary to work with nonlinear equations expressed in the form space-state, the estimation of states and parameters in nonlinear dynamic systems remains a task not yet resolved. The methods of dual estimation to estimate states solutions and parameters was initially proposed in [13]; later [25] proposed a method of estimation of states and parameters in biochemical network models, which are essential for the study of dynamics of biological systems. In this paper we propose to implement a computational algorithm that will allow us to simultaneously estimate states solutions and parameters in mixed models defined by an SDE. To define the algorithm, consider a dynamic state-parameters system defined by

$$\mathbf{x}_n = \mathcal{M}_{n-1}(x_{n-1}, \theta) + \mathbf{u}_{n-1}, \quad \mathbf{u}_{n-1} \sim N(0, Q_n) \quad (5)$$

$$\mathbf{y}_n = \mathcal{H}_n(x_{n-1}) + \mathbf{v}_n, \quad \mathbf{v}_n \sim N(0, R_n) \quad (6)$$

where $\mathbf{x}_n \in \mathbb{R}^{N_x}$, $\mathbf{y}_n \in \mathbb{R}^{N_y}$ denote the system states and the observation at time t_n of dimensions N_x and N_y respectively, $\theta \in \mathbb{R}^{N_\theta}$ is the parameter vector of dimension N_θ, $\mathcal{M}_n$ is a nonlinear operator integrating the system state from time t_n to t_{n+1}, $\mathcal{H}_n$ is observational operator at time t_n, $\mathbf{u}_n$ and $\mathbf{v}_n$ are errors of observation and state respectively. One of methods of estimation is to use EKF (see, [25]). Let $\mathbf{z} = \begin{pmatrix}\mathbf{x}^T & \theta^T\end{pmatrix}^T$, then the augmented state equations are given by

$$z_{n+1} = \begin{pmatrix} x_{n+1} \\ \theta_{n+1} \end{pmatrix} = \begin{pmatrix} \mathcal{M}_{n-1}(x_n, \theta_n) \\ \theta_n \end{pmatrix} + \begin{pmatrix} \mathbf{u}_n \\ \mathbf{v}_n \end{pmatrix} = \mathcal{M}_{n-1}(z_n) + \zeta_n$$

where $\mathbf{u}_n$ and $\mathbf{v}_n$ be Gaussian with zero means and covariances Q_n and R_n, respectively. After extending the state variables with the parameter vector, the observation equation becomes $y_n = \mathcal{H}_n(x_n, \theta_n) + v_n$. Following the notation given in [25], and the basic idea behind the Kalman filter is that it operates by propagating the mean and covariance of the state over time, as follows (see [23]). Define

$$\hat{z}_{n|n} = \mathbb{E}\{z_n|y_{1:n}\}; \qquad \hat{z}_{n+1|n} = \mathbb{E}\{z_{n+1}|y_{1:n}\}$$
$$\hat{\mathcal{P}}_{n|n} = \mathbb{E}\{\left(z_n - \hat{z}_{n|n}\right)\left(z_n - \hat{z}_{n|n}\right)^T|y_{1:n}\};$$

and

$$\hat{\mathcal{P}}_{n+1|n} = \mathbb{E}\{\left(z_{n+1} - \hat{z}_{n+1|n}\right)\left(z_{n+1} - \hat{z}_{n+1|n}\right)^T|y_{1:n}\}$$

where $\hat{z}_{n|n} = \begin{pmatrix} \hat{x}_{n|n} \\ \hat{\theta}_{n|n} \end{pmatrix}$. Writing the model space state in compact form

$$z_n = \mathcal{M}(z_n) + \zeta_n \tag{7}$$
$$y_n = \mathcal{H}(z_n) + v_n \tag{8}$$

To estimate the vector of states and parameters, the Jacobian matrices are calculated as follows

$$\mathcal{J}_n = \begin{pmatrix} \frac{\partial\mathcal{M}}{\partial x^T} & \frac{\partial\mathcal{M}}{\partial\theta^T} \\ 0 & I \end{pmatrix}_{\hat{z}_{n|n},\hat{\theta}_{n|n}} \quad \text{and} \quad \mathcal{H}_{n+1} = \begin{pmatrix} \frac{\partial\mathcal{H}}{\partial x^T} & \frac{\partial\mathcal{H}}{\partial\theta^T} \end{pmatrix}_{\hat{z}_{n+1|n},\hat{\theta}_{n+1|n}}$$

Then the EKF algorithm for dual estimation consists of two steps Prediction and filtering:

– **Forecast Step:** Given an estimated state $\hat{z}_{n|n}$ at one time $t = n$, and the observations $y_{1:n}$, the system moves to a new state in time $t = n+1$

$$\begin{aligned}\hat{z}_{n+1|n} &= \mathbb{E}\left\{z_{n+1}|y_{1:n}\right\} = \mathbb{E}\left\{\mathcal{M}(z_n) + \zeta_n|y_{1:n}\right\} \\ &\approx \mathbb{E}\left\{\mathcal{M}(\hat{z}_{n|n}) + \tfrac{\partial\mathcal{M}}{\partial z^T}\left(z_n - \hat{z}_{n|n}\right)|y_{1:n}\right\} \\ &= \mathcal{M}(\hat{z}_{n|n})\end{aligned}$$

The variance and covariance matrix of the prediction error can be calculated as follows

$$\begin{aligned}\hat{\mathcal{P}}_{n+1|n} &= \mathbb{E}\left\{\left(z_{n+1} - \hat{z}_{n+1|n}\right)\left(z_{n+1} - \hat{z}_{n+1|n}\right)^T|y_{1:n}\right\} \\ &= \mathbb{E}\left\{\left(\mathcal{M}(z_n) + \zeta_n - \hat{z}_{n+1|n}\right)\left(\mathcal{M}(z_n) + \zeta_n - \hat{z}_{n+1|n}\right)^T|y_{1:n}\right\} \\ &= \mathbb{E}\left\{\left[\tfrac{\partial\mathcal{M}}{\partial z^T}(z_n - \hat{z}_{n|n}) + \zeta_n\right]\left[\tfrac{\partial\mathcal{M}}{\partial z^T}(z_n - \hat{z}_{n|n}) + \zeta_n\right]^T|y_{1:n}\right\} \\ &= \mathcal{J}_n\mathcal{P}_{n|n}\mathcal{J}_n^T + \psi_n\end{aligned}$$

where

$$\mathcal{P}_{n|n} = \mathbb{E}\left\{\left(z_n - \hat{z}_{n|n}\right)\left(z_n - \hat{z}_{n|n}\right)^T\right\} \quad \text{and} \quad \psi_n = \begin{pmatrix} Q_n & 0 \\ 0 & \Phi_n \end{pmatrix}$$

– **Update Step:** In the filtration cycle, an observation is used y_{n+1} in time $n+1$ to estimate the state update. The measurement prediction error is defined as

$$\epsilon_n = y_{n+1} - \mathcal{H}_n(\hat{z}_{n+1|n})$$

The estimator of the state in time $n+1$ is

$$\begin{aligned}\hat{z}_{n+1|n} &= \mathbb{E}\left\{z_{n+1}|y_{1:n+1}\right\} = \mathbb{E}\left\{z_{n+1}|y_{1:n}, \epsilon_n\right\} \\ &= \mathbb{E}\left\{z_{n+1}|y_{1:n}\right\} + Cov\left(z_{n+1}, \epsilon_n\right)\left[Var\left(E_n\right)\right]^{-1}\epsilon_n \\ &= \hat{z}_{n+1|n} + \mathcal{K}_{n+1}\left(y_{n+1} - \mathcal{H}(\hat{z}_{n|n})\right)\end{aligned}$$

Where $\mathcal{K}_{n+1}$ is the Kalman gain matrix defined as

$$\mathcal{K}_{n+1} = \mathcal{P}_{n+1|n}\mathcal{H}_{n+1}^T\left(\mathcal{H}_{n+1}\mathcal{P}_{n+1|n}\mathcal{H}_{n+1}^T + R_{n+1}\right)^{-1}$$

The updated estimate covariance matrix is given by

$$\begin{aligned}\mathcal{P}_{n+1|n+1} &= \mathbb{E}\left\{\left(z_{n+1} - \hat{z}_{n+1|n}\right)\left(z_{n+1} - \hat{z}_{n+1|n}\right)^T |y_{1:n+1}\right\} \\ &= \left(I - \mathcal{K}_{n+1}\mathcal{H}_{n+1}\right)\mathcal{P}_{n+1|n}\end{aligned}$$

3.2 Ensemble Kalman Filter (EnKF) for dual estimations

The ensemble Kalman filter (EnKF) can be viewed as an approximate version of the Kalman filter, in which the state distribution is represented by a sample from the distribution. This sample is then propagated forward through time and updated when new data become available. The ensemble representation is a form of dimension reduction, which leads to computational feasibility even for very high-dimensional systems.
The EnFK was developed by [2],[3], [28], with applications in [17], [21], and [22]. among others. The algorithm represents a generalization of the Kalman filter for nonlinear models with many parameters, the procedure consists in generating a set of trajectories using sequential Monte Carlo techniques, from which the expectation and covariance are updated as time advance and serve to approximate the a posterior distribution of dynamic system states and parameters.
The EnFK is expressed in the form space state considering the augmented state $z_n = \left(x_n^T, \theta^T\right)^T$, then the model defined by the equations (5) and (6) is transformed in

$$\mathbf{z}_n = \tilde{\mathcal{M}}_{n-1}(z_{n-1}) + \tilde{\zeta}_n \tag{9}$$

$$\mathbf{y}_n = \tilde{\mathcal{H}}_n(z_n) + \tilde{\mathbf{v}}_n \tag{10}$$

where

$$\tilde{\mathcal{M}}_{n-1}(z_{n-1}) = \begin{pmatrix} \mathcal{M}_{n-1}(z_{n-1}) \\ \theta \end{pmatrix},$$

$\tilde{\mathbf{u}}_n = (\mathbf{u}_{n-1}^T, \mathbf{0})^T$, $\tilde{\mathcal{H}}_n(z_{n-1}) = \left(\mathcal{H}_{n-1}^T(z_n), \mathbf{0}\right)^T$, $\tilde{\mathbf{v}}_n = (\mathbf{v}_n^T, \mathbf{0})^T$, with $\mathbf{0}$ a matrix of zeros with an appropriate dimension.
The state vector can be written in the augmented form, which includes the state vector $x_{0:n} = (x_0, \ldots, x_n)$ and the parameter vector $\theta = (\theta_1 \ldots, \theta_p)$, this

$$\mathbf{z}_{n-1} = \begin{pmatrix} x_{n-1}^a \\ \theta_{n-1} \end{pmatrix}$$

Let $\bar{x}_n^f$ the mean and P_n^f the covariance matrix of the forecast of the filtered posterior distribution $p(z_{n-1}|y_{1:n-1})$. Samples are generated from the filtered posterior distribution $p(z_n|y_{1:n})$ and update the mean $\bar{x}_n^a$ and covariance P_n^a.
To achieve this goal it is assumed that independent samples are generated from a filtered density $p(z_{n-1}|y_{1:n-1})$. The EnFK uses the first-order Markov property in the system given in (9) and (10), and generates samples to approximate the density $p(z_n|y_{1:n})$, sampling of filtered density $p(z_n|x_{n-1}^a, \theta_{n-1})$. The procedure is as follows: suppose that the posterior density $p(z_{n-1}|y_{1:n-1})$ is known in time $t-1$, the prior density in time t is estimated by $p(z_n|y_{1:n-1})$, when a new observation y_n is obtained from the system, the subsequent density $p(z_n|y_{1:n})$ is updated recursively using the Bayes theorem. The algorithm proceeds as follows

Forecast step:

$$x_{n,i}^f = \tilde{\mathcal{M}}_{n-1}(x_{n-1,i}^a, \theta_{n-1}^i) + \eta_{n-1} \quad , \quad \eta_{n-1} \sim N(0, Q_{n-1}) \tag{11}$$

where $x_{n,i}^f$ is a vector of predicted values of order $k \times 1$ of the states of the $i-$th member of the ensemble for $i = 1, \ldots, n$. From the state vector ensembles, the error covariance matrix $k \times k$ can be calculated using

$$C = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_{n,i}^f - \bar{x}_n^f\right) \left(x_{n,i}^f - \bar{x}_n^f\right)^T$$

where $\bar{x}_n^f$ denotes a vector of order $k \times 1$ of averages of the states in time $t = n$.
In order to estimate the online parameters it is considered that $\Theta_{1:n} = (\theta_1, \ldots, \theta_n)$ varies over time, that is

$$\theta_n = \theta_{n-1} + \nu_n \quad ; \quad \nu_n \sim N(0, \sigma_\nu^2)$$

To approximate $p(\theta_n^f|y_{1:n})$, sample $\theta_n^{(j)}$ and weights $\omega_n^{(j)}$ using the algorithm of Liu and West (2003), a generalization of the auxiliary particulate filter of Pitt and Shephard (1999) to carry out the learning parameters, the method consists of the following

$$p(\theta_n^f|y_{1:n}) = \sum_{j=1}^{n} \omega_t^{(j)} N(m_t^{(j)}, h^2 V_n) \quad , \quad \sum_{j=1}^{n} \omega_t^{(j)} = 1$$

where $m_n^{(j)} = \alpha\theta_n^{(j)} + (1-\alpha)\bar{\theta}_n$, $\bar{\theta}_n = \frac{1}{n}\sum_{j=1}^{n}\theta_n^{(j)}$ and

$$V_n = \frac{1}{n}\sum_{j=1}^{n}(\theta_n^{(j)} - \hat{\theta}_n)(\theta_n^{(j)} - \bar{\theta}_n)^T$$

The values generated $\{\theta_n^j\}_{j=1}^N$ they serve to approximate the posterior distribution of the parameters and allow us to obtain estimated quantities such as $\mathcal{E}\left(\theta|y_{1:n}^a\right)$, α is a contraction factor associated with h, where $h = \sqrt{1-\alpha^2}$ and represents a smoothness factor.

Analysis step: Once a new observation is available, all members $x_{n,i}^f$ y θ_{n-1}^j, are updated using the Kalman filter

$$y_{n,i}^f = \tilde{\mathcal{H}}_n(x_{n,i}^f) + \epsilon_{n,i} \quad , \quad \epsilon_{n,i} \sim N(0, R_n)$$

now the predicted states of each member of the set can be updated as follows

$$x_{n,i}^a = x_{n,i}^f + \mathcal{K}\left(y_{n,i}^f - \tilde{\mathcal{H}}_n(x_{n,i}^f)\right)$$

where $x_{n,i}^a$ denotes an order vector $k \times 1$ with the updated estimated values of the variable states, and $\mathcal{K}$ denotes an order matrix $k \times m$, called the gain matrix, which is calculated

$$\mathcal{K} = CH_n^T\left(H_nCH_n^T + R_n\right)^{-1} \tag{12}$$

The equation (12) offers more flexibility in dealing with operators. observational H_n nonlinear ([20]).

Let $x^* = \begin{pmatrix} x \\ \theta \end{pmatrix}$ with $\mathcal{H}^* = (\mathcal{H}_n, \mathbf{0})^T$, the covariance matrix C^* is determined by

$$C^* = \begin{pmatrix} \mathbf{v}ar(x) & Cov(x,\theta) \\ Cov(\theta,x) & \mathbf{v}ar(\theta) \end{pmatrix}$$

then the gain matrix is modified

$$\mathcal{K}^* = C^*(H^*)^T\left(H^*C^*(H^*)^T + R_n\right)^{-1}$$

that is

$$\mathcal{K}^* = \begin{pmatrix} \mathbf{v}ar(x) & Cov(x,\theta) \\ Cov(\theta,x) & \mathbf{v}ar(\theta) \end{pmatrix}\begin{pmatrix} \mathcal{H}_n^T \\ \mathbf{0} \end{pmatrix}\left((\mathcal{H}_n \quad \mathbf{0})\begin{pmatrix} \mathbf{v}ar(x) & Cov(x,\theta) \\ Cov(\theta,x) & \mathbf{v}ar(\theta) \end{pmatrix}\begin{pmatrix} \mathcal{H}_n^T \\ \mathbf{0} \end{pmatrix}\right)^{-1}$$

Solving, is obtained

$$\mathcal{K}^* = \begin{pmatrix} \mathbf{v}ar(x)\mathcal{H}_n^T\left(\mathcal{H}_n\mathbf{v}ar(x)\mathcal{H}_n^T + R\right)^{-1} \\ Cov(\theta,x)\mathcal{H}_n^T\left(\mathcal{H}_n\mathbf{v}ar(x)\mathcal{H}_n^T + R\right)^{-1} \end{pmatrix} = \begin{pmatrix} \mathcal{K}_x^* \\ \mathcal{K}_\theta^* \end{pmatrix}$$

This last equation allows updating the states and parameters as follows

$$\begin{pmatrix} x_n^a \\ \theta_n^a \end{pmatrix} = \begin{pmatrix} x_n^f \\ \theta_n^f \end{pmatrix} + \begin{pmatrix} \mathcal{K}_x^* \left(y_{n,i}^f - \tilde{\mathcal{H}}_n(x_{n,i}^f) \right) \\ \mathcal{K}_\theta^* \left(y_{n,i}^f - \tilde{\mathcal{H}}_n(x_{n,i}^f) \right) \end{pmatrix}$$

In summary, the EnKF implementation is as follows

Step 1. Creation of the initial ensemble: the algorithm is initialized in time $n = 0$ and an initial state is generated

$$x_{0,i}^a \sim N(\bar{m}_0, C_0) \quad , \quad i = 1, \ldots, n$$

This is done in the following way, first factoring the covariance $C_0 = \Sigma_0 \Sigma_0^T$, and define

$$x_{0,i}^a = \bar{m}_0 + \Sigma_0 n_0^{(i)} \quad , \quad n_0^{(i)} \sim N(0, I)$$

The sample mean is

$$\bar{m}_0^f = \frac{1}{n} \sum_{i=1}^{n} x_{0,i}^a$$

and the sample covariance is

$$C_0^f = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_{0,i}^a - \bar{m}_0^f \right) \left(x_{0,i}^a - \bar{m}_0^f \right)^T$$

The parameters θ_0^i required for initialization can be generated according to the Liu and West algorithm (2001).

Step 2. Ensemble forecast: consider that for an instant of time $n = t$, the mean and the covariance are known $\left(\bar{m}_n, \tilde{C}_n \right)$, let $\tilde{C}_n = \tilde{\Sigma}_n \tilde{\Sigma}_n^T$, then the ensemble is created

$$x_{n,i}^a = \bar{m}_n + \tilde{\Sigma}_n \varepsilon_n^{(i)} \quad , \quad \varepsilon_n^{(i)} \sim N(0, I)$$

The n members of the ensemble forecast at the time $n = t$ are generated as follows

$$\mathbf{x}_{n,i}^f = \tilde{\mathcal{M}}_{n-1} \left(x_{n,i}^a, \theta \right) + \tilde{\mathbf{u}}_n^i \quad , \quad \tilde{\mathbf{u}}_n^i \sim N(0, Q_n)$$

An unbiased estimator of the sample mean is given by

$$\bar{m}_n^f = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_{n,i}^f$$

An unbiased sample covariance estimator is given by

$$C_n^f = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{x}_{n,i}^f - \bar{m}_n^f \right) \left(\mathbf{x}_{n,i}^f - \bar{m}_n^f \right)^T + Q_n$$

In practice, it is not common to approximate C_n^f and instead estimate

$$C_n^{cr} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_{n,i}^f - \bar{m}_n^f\right)\left(\tilde{\mathcal{H}}_n(\mathbf{x}_{n,i}^f) - \tilde{\mathcal{H}}_n(\bar{m}_n^f)\right)^T$$

and

$$C_n^{pr} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{\mathcal{H}}_n(\mathbf{x}_{n,i}^f) - \tilde{\mathcal{H}}_n(\bar{m}_n^f)\right)\left(\tilde{\mathcal{H}}_n(\mathbf{x}_{n,i}^f) - \tilde{\mathcal{H}}_n(\bar{m}_n^f)\right)^T$$

where C_n^{cr} is the cross-sampled covariance between the previous ensemble and the predicted projection in the observations space and C_n^{pr} is the sample covariance of the predicted projection of the previous ensemble in the observations space.
Step 3. Data assimilation: If no observations are available over time n, the step of update is the following, it is done that

$$\left\{x_{n,i}^a = x_{n,i}^f \quad , \quad i = 1, \ldots, n\right\}$$

If observations are available in time $n = t$, then the ensemble is updated. using the disturbed observation algorithm ([**?**]). First, the following are generated synthetic observations from the observation equation:

$$\mathbf{y}_{n,i}^f = \tilde{\mathcal{H}}_n(x_{n,i}^f) + \tilde{\mathbf{v}}_n^i \quad , \quad i = 1, \ldots, n \quad , \quad \tilde{\mathbf{v}}_n^i \sim N(0, R_n) \tag{13}$$

The update is as follows

- The states are updated

$$x_n^a = x_n^f + \mathcal{K}_x^*\left(y_{n,i}^f - \tilde{\mathcal{H}}_n(x_{n,i}^f)\right)$$

- The parameters are updated

$$\theta_n^a = \theta_n^f + \mathcal{K}_\theta^*\left(y_{n,i}^f - \tilde{\mathcal{H}}_n(x_{n,i}^f)\right)$$

4 Results

In particular, a mixed-effect model defined by a SDE modeled by an O-U process and an addictive error model defined in [11].

$$\begin{aligned} y_{ij} &= x_{ij} + \epsilon_{ij} \quad , \quad \epsilon_{ij} \sim N(0.\sigma^2) \\ dx_{it} &= -\left(\tfrac{x_{it}}{\tau_i} - \kappa_j\right) dt + \gamma dB_{it} \quad , \quad x_0 = 0 \end{aligned} \tag{14}$$

where $\kappa_i \in \mathrm{I\!R}$, $\tau_i > 0$. In this case, consider the following parameterization $\phi_i = (\log(\tau_i), \kappa_i)$ the vector of individual parameters, and suppose that: $\log(\tau_i) \overset{i.i.d}{\sim} N(0, \sigma^2)$

and $\kappa_i \overset{i.i.d}{\sim} N(k, \omega_k^2)$ vector of parameters is $\theta = (\log(\tau), \kappa, \omega_\tau, \omega_\kappa, \gamma, \sigma)$. The discretized model in the times t_{ij} is

$$\begin{aligned} y_{ij} &= x_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0.\sigma^2) \\ x_{ij} &= x_{ij-1}\exp(-\Delta_{ij}\tau_i) - \kappa_i\tau_i\left(1 - \exp(-\Delta_{ij}\tau_i)\right) + \eta_{ij},\ \eta_{ij} \sim N\left(0, \nu_\eta\right) \end{aligned} \tag{15}$$

where $\nu_\eta = \gamma^2\tau_i(1 - \exp(-\Delta_{ij}\tau_i)$, $\Delta_{ij} = t_{ij} - t_{ij-1}$.
The vector of conditional states in random effects is given by

$$X_i = (x_{i1}, \ldots, x_{iJ}|\phi) \sim N\left(\mu_{ix}, \Sigma_{ix}\right)$$

with $\mu_{ix} = (\tau_i\kappa_i(1 - \exp(-t_{i1})), \ldots, \tau_i\kappa_i(1 - \exp(-t_{iJ})))^T$ and

$$\Sigma_{ix} = \left[\frac{\tau_i\gamma^2}{2}\left(1 - \exp(-\frac{2\min(t_{ij}, t_{ik})}{\tau_i})\right)\exp\left(-\frac{|t_{ij} - t_{ik}|}{\tau_i}\right)\right]_{1\leq j,k\leq J}$$

Although the stochastic differential equation is linear, the transition density $p(X_{ij}, \Delta_{ij}|X_{ij-1}, \phi_i; \theta)$ is a nonlinear function of ϕ_i. Then the likelihood does not have a closed form and consequently an exact estimator of θ. Using Software R, a series of simulations were carried out, the results of which are presented below,we apply the EKF to equation (14). To estimate the X_i states and θ parameters, the EKF algorithm and the EnKF will be implemented, use the model defined by the equation (15). To initialize the algorithms a partition was taken $I = [0, 20]$, $\Delta t_i = t_i - t_{i-1}$, $t_0 = 0 < t_1 < \cdots < t_{20}$, $n = 200$ iterations, $x = 10$, and $\kappa = -4$. The results obtained are presented in the following summary table where I.V denote initial values.

	$\hat{x}$	$\log(\tau)$	κ	γ	σ	ω_t	ω_κ
I.V.	0	0.6	1	0.1	0.1	0.5	0.5
Mean	1.2	0.65445668	1.13230340	0.03880077	0.02745250	0.48106387	0.45670812
SD	2.88	0.03812852	0.08503318	0.0300471	0.0894444	0.0356457	0.04869007

Table 1. Mean and standard deviations for $\hat{x}$ and θ obtained by the EKF.

To determine the estimation quality of the EKF and EnKF algorithms, it was used as a measure of goodness-of-fit, the square root of the mean quadratic error is used, which is defined as follows $RMSE = \sqrt{\sum_{r=1}^{N}\frac{(\hat{\theta}_r - \theta)^2}{N}}$. In this case the values of $RMSE = 0.006923648$ obtained, which represents a good approximation, so we conclude that the method of estimating states and parameters using the EKF works correctly. Next we present some of the obtained graphs, where we compare the observed data, the origin signal and the estimated signal by means of the EKF, for the equation (14). To estimate states and parameters using EnKF, in this case, size 100 ensembles were used. Below is the summary table of the values obtained a s well as the graphs corresponding to the observed data and the estimates of the states obtained.

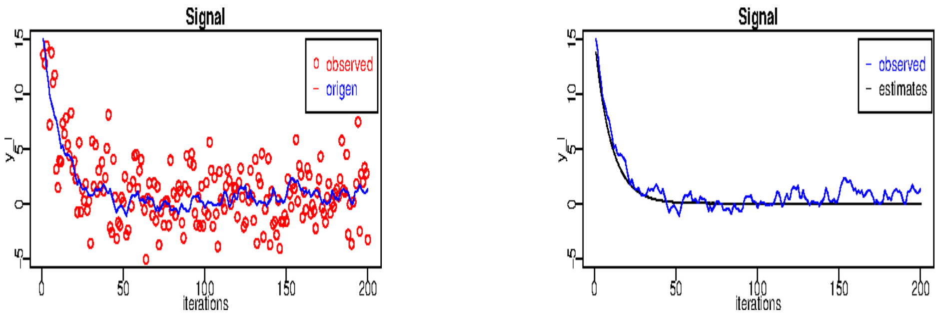

Fig. 1. signal observed and signal estimates using EKF

	$\hat{x}$	$\log(\tau)$	κ	γ	σ	ω_t	ω_κ
I.V.	0	4.9346678	-0.0007751	4.0	1.65	2.0	0.07
Mean	-1.7	-1.774401	2.186047	1.696948	0.834189	0.566036	-0.012332
SD	0.15	0.1518524	0.1617901	0.1524445	0.4045653	0.1639408	0.704994

Table 2. Mean and standard deviations for $\hat{x}$ and θ obtained by the EnKF.

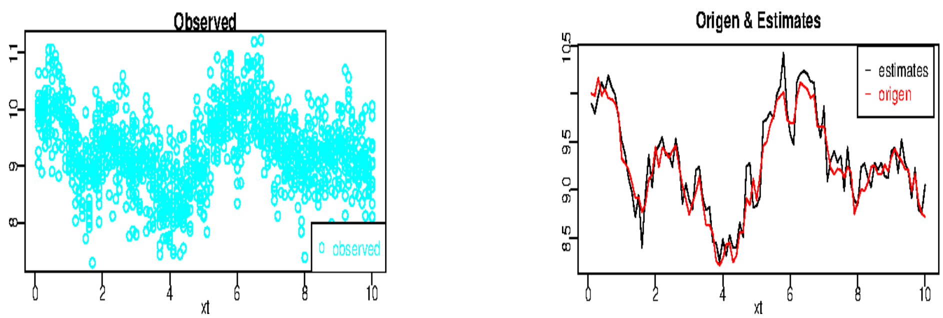

Fig. 2. signal observed and signal estimates using EnKF

??

5 Discussions and Conclusions

In this article we present two computational algorithms known as the EKF and the EnKF for the joint estimation of states and parameters in stochastic differential mixed-effect models including observations noise. In the absence of efficient and computationally and reasonable estimation methods, these algorithms are easy to code, the implementation is automatic, they require less calculation time than other similar methods such as the SAEM algorithm; in addition, they allow us to obtain accurate estimates when comparing true data against estimated data, random samples can also be generated when there is a problem of missing data or partially observed processes. To illustrate the methodology, the solution states and parameters of the Ornstein-Uhlembeck mixed-effects model are estimated, obtaining results similar to those obtained by the SAEM algorithm ([11]), but with smaller errors. In order to measure the estimation quality of the algorithms, the square root of the mean quadratic error was used as a measure of goodness of adjustment, obtaining very small errors.

References

1. Bacher, P. and Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. Energy and Buildings 43; 1511-1522.
2. Evensen, G. (2003): The ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics, 53:343-367.
3. Evensen G (2009): Data assimilation: the ensemble Kalman Filter. Second Edition, Springer.
4. Dahlin, J; Kohn, R; Sochön, T. (2016). Bayesian inference for mixed effects models with heterogeneity. Technical reports from the Automatic Control group in Linköping are available from http://www.control.isy.liu.se/publications.
5. Delattre, M and Lavielle, M. (2013). Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. Statistics and Its Interface Volume 6, 519 -532. 10.
6. Ditlevsen, S; De Gaetano, A. (2005). Mixed effects in Stochastic Differential Equation Models. Statistical Journal, Volume 3, Number 2, 137(153).
7. Donnet, S and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. ESAIM: Probability and Statistics, Vol. 12, p. $196-218$. www.esaim-ps.org. DOI: 10.1051/ps:2007045.
8. Donnet, S and Samson, A. (2014). Using PMCMC in EM algorithm for stochastic mixed models; theorietical and practical issues. Journal de la Société Francaise de Statistique. Vol. 155, 1.
9. , S and Samson, A. (2013). A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamics models. Advanced Drug Delivery Reviews, 65(7): 929(939)939.
10. Donnet, S; Foulley, J; Samson, A. (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differentials equations. Biometrics, Wiley, 66(3), pp.733-741.
11. Donnet, S; Foulley, J; Samson, A. (2014). Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. Journal de la Société Française de Statistique, Vol. 155 No. 1 49-72
12. Faugeras, O., Touboul, J. and Cessac, B. (2009). A constructive mean field analysis of multi population neural networks with random synaptic weights and stochastic inputs. Frontiers in Computational Neuroscience 3; 1-28.
13. Goodall, Li; and Kadirkamanathan, V. (2004) Estimation of parameters in a linear state space model using a Rao-Blackwellised particle filter. IEE Proc Control Theory Appl 151: 727-738.
14. Hansen, A., Dun-Henriksen, A., Juhl, R., Schmidt, S., Norgaard, K., Jorgensen, J. and Madsen, H. (2014). Predicting plasma glucose from interstitial glucose observations using Bayesian methods. Journal of diabetes science and technology 8; 321-330.
15. Hermann, S., Ickstadt, K. and Müller, C. (2016). Bayesian Prediction of Crack Growth Based on a Hierarchical Diffusion Model. to appear in: Applied Stochastic Models in Business and Industry.

16. Liu, J., West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In: Doucet, A, de Freitas, J., Gordon, N.J. (eds.), Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York.
17. Infante, S; Sánchez, L; y Cedeño, F. (2012). Filtros para Predecir Incertidumbre de Lluvia y Clima. Revista de climatología, Vol. 12 (2012): 33-48 ISSN 1578-8768. Vol.
18. Infante, S; Sánchez, L; and Hernández, A. (2019). Stochastic Models to Estimate Population Dynamics. Stat., Optim. Inf. Comput., Vol. 7, pp $311 - 328$. Published online in International Academic Press (www.IAPress.org).
19. Iversen, E., Morales, J., Moller, J. and Madsen, H. (2014). Probabilistic forecasts of solar irradiance using stochastic differential equations. Environmetrics 25; 152-164.
20. Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Adv. Water Resour., 28, 135-147.
21. Sánchez, L; Infante, S; Marcano, J; Griffin, V. (2015). Polynomial chaos based on the parallelized ensemble kalman filter to estimate precipitation states. Statistics, Optimization & Information Computing, 3(1):79-95.
22. Sánchez, L; Infante, S; Griffin, V; Rey, D. (2016). Spatio-temporal dynamic model and parallelized ensemble kalman filter for precipitation data. Brazilian Journal of Probability and Statistics, 30(4):653-675.
23. Simon D. (2006). optimal state estimation-Kalman, H‘, and nonlinear approaches. New Jersey: A John Wiley & Sons, Inc. Hoboken. 552 p.
24. Soto, J; Infante, S; Camaho, F, Amaro, I. (2019). Estimación de un modelo de efectos mixtos usando un proceso de difusión parcialmente observado. Revista de Matemática: teoría y aplicaciones. Vol 26(1):83-98 CIMPA-UCR ISSN: 1409-2433 (PRINT),2215-3373 (ONLINE). DOI: https://doi.org/10.15517/rmta.v26i1.35527.
25. Sun, X; Jin, L; Xiong, M. (2008). Extended Kalman Filter for Estimation of Parameters in Nonlinear State-Space Models of Biochemical Networks. PLoS ONE 3(11): e3758. doi:10.1371/journal.pone.0003758.
26. Overgaard, R. V., Jonsson, N., Tornøe, C. W. and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. Journal of Pharmacokinetics and Pharmacodynamics 32 85-107.
27. Picchini, U, Forman, J. (2016). Stochastic differential equation mixed effects models for tumor growth and response to treatment. arXiv : 1607;02633v2[stat:AP].
28. van Leeuwen P, Evensen G (1996): Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. Monthly Weather Rev., 124:2898-2913.
29. Whitaker, G. A; Golightly, A; Boys, R, Sherlock, C. (2016). Bayesian inference for diffusion-driven mixed-effects models. Bayesian Analysis. DOI: http : //dx:doi:org/10.1214/16-BA1009.