



HAL
open science

Flow-based Bayesian estimation of nonlinear differential equations for modeling biological networks

Nicolas Brunel, Florence d'Alché-Buc

► **To cite this version:**

Nicolas Brunel, Florence d'Alché-Buc. Flow-based Bayesian estimation of nonlinear differential equations for modeling biological networks. 5th IAPR international conference on Pattern recognition in bioinformatics (PRIB'10), Sep 2010, Nijmegen, Netherlands. pp.443–454, 10.1007/978-3-642-16001-1_38 . hal-00654125

HAL Id: hal-00654125

<https://hal.science/hal-00654125>

Submitted on 17 Feb 2014

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.

Flow-based Bayesian estimation of nonlinear differential equations for modeling biological networks

Nicolas Brunel⁽¹⁾ and Florence d'Alché-Buc^(1,2)

(1) Laboratoire IBISC, Université d'Evry, FRANCE

(2) URA 2171, Institut Pasteur, FRANCE

`nicolas.brunel,florence.dalche@ibisc.fr`

Abstract. We consider the problem of estimating parameters and unobserved trajectories in nonlinear ordinary differential equations (ODEs) from noisy and partially observed data. We focus on a class of state-space models defined from the integration of the differential equation in the evolution equation. Within a Bayesian framework, we derive a non-sequential estimation procedure that infers the parameters and the initial condition of the ODE, taking into account that both are required to fully characterize the solution of the ODE. This point of view, new in the context of state-space models, modifies the learning problem. To evaluate the relevance of this approach, we use an Adaptive Importance Sampling in a population Monte Carlo scheme to approximate the posterior probability distribution. We compare this approach to recursive estimation via Unscented Kalman Filtering on two reverse-modeling problems in systems biology. On both problems, our method improves on classical smoothing methods used in state space models for the estimation of unobserved trajectories.

1 Introduction

1.1 Context

In recent years, there has been a growing interest in identifying complex dynamical systems in biochemistry and biology [15]. In this context, Ordinary Differential Equations (ODEs) have been widely studied for analyzing the dynamics of gene regulatory and signaling networks [11, 14]. They also appear as good candidates for the reverse-modeling task. In the present work, we consider the problem of estimating parameters and unobserved trajectories in differential equations from experimental data. Nowadays, parameter estimation in differential equations is still considered as a challenging problem when the dynamical system is only partially observed through noisy measurements and exhibit nonlinear dynamics. This is usually the case in reverse-modeling of regulatory and signaling networks [2, 16]. Some approaches address the estimation problem based on a Bayesian estimation of state-space models that integrate the ODE in the evolution equation. This framework has shown to be relevant in producing

efficient algorithms [20, 16, 21]. However, they suffer from two drawbacks: first they largely neglect the role of the initial condition and second, they assume the gaussianity of the posterior probability distribution of the parameters. In the present work, we are mainly interested in eliminating the first drawback by taking into account that the initial condition is a key parameter of the ODE solution. This means that we search the system parameters and the initial conditions that fit the observed data and also provide a proper solution to the ODE. Then, as a secondary contribution, we also improve the Bayesian approach derived in [16] and in [21] by a better approximation of the posterior probability distribution.

1.2 Strategy

We first define the estimation task by introducing into the equations the flow of the ordinary differential equation. The flow of an ODE puts emphasis on the sensitivity of its solution with respect to the initial conditions. Then, we use an augmented approach that encapsulates the initial conditions and the ODEs parameters into the same augmented initial condition vector. Within this framework, the deterministic nature of the hidden process provides a non-recursive definition of hidden states from the augmented initial condition, with an integration of the ODE in the whole time interval of observation.

At this stage, we propose to address the problem with a Bayesian approach, searching for the posterior probability distribution of the augmented initial condition. The solutions previously proposed for recursive estimation in the case of nonlinear systems are based on nonlinear extensions of Kalman filtering and smoothing. We notice that procedures like Unscented Transform methods used for computing the posterior probability of the states make a strong assumption about the Gaussianity of the posterior distribution. Whereas particle filters do not make this assumptions, they do not deal correctly with deterministic processes as pointed out by the work of [13].

The idea of approximating the posterior probability by a weighted sample is computationally attractive while being a versatile approach adapted to a large variety of distributions. With respect to these considerations, we investigate the use of Monte-Carlo methods [9] for the approximation of the posterior distribution by a weighted sample built from an iterative importance sampling resampling scheme. As recently shown by [7] and [3], this approach consists in an adaptive selection of the importance distribution, which is crucial in high-dimensional sampling. The updating mechanism of the importance distribution consists in moving the population with a transition kernel (D-kernel [7]). The non-recursive estimation of the augmented initial condition is applied on two typical systems biology models: the α -pinene network [19] and the Repressilator network ([8]).

The paper is organized as follows. In section 2, we introduce the new setting of parameter estimation in terms of augmented initial condition estimation and exploit it in the context of Bayesian estimation. In section 3, we recall the main features of Population Monte-Carlo schemes and focus on an adaptive algorithm

that corrects the importance distribution. Section 4 is devoted to numerical experiments. Finally, we draw a conclusion and perspectives to this work in the section 5.

2 The initial condition learning problem

2.1 Flow of an ODE and statistical modeling

We consider a biological dynamical system, for instance a gene regulatory network, modeled by the following ordinary differential equation:

$$\dot{x}(t) = f(t, x(t), \theta) \quad (1)$$

defined on the time interval $[0, T]$ ($T > 0$). $x(t)$ is the state vector of dimension d : in the case of a regulatory network, it corresponds to the vector of the expression levels of d genes. f is a (time-dependent) vector field from \mathbb{R}^d to \mathbb{R}^d , indexed by a parameter $\theta \in \Theta \subset \mathbb{R}^p$. Examples of functions f abound in the literature of systems biology [15]: Hill kinetics, law of mass equations, ...

A relevant way to characterize the differential equation under study is to define its flow $\phi_\theta : (t, x_0) \mapsto \phi_\theta(t, x_0)$ which represents the influence of $x(0) = x_0$ on the solution, i.e. $t \mapsto \phi_\theta(t, x_0)$ is the solution to (1) starting from x_0 . Hence, the flow puts emphasis on the sensitivity of a solution of (1) with respect to the initial conditions. It can also be seen as a map defined in the phase space that shows how the points are transported by the vector field.

Now, let us introduce N noisy measurements, $y_n \in \mathbb{R}^m, n = 0 \dots N - 1$, that are acquired from a smooth observation function $h : \mathbb{R}^d \rightarrow \mathbb{R}^m$ ($m \geq 1$) at N times $t_0 = 0 < t_1 < \dots < t_{N-1} = T$:

$$y_n = h(\phi_\theta(t_n, x_0)) + \epsilon_n \quad (2)$$

where the noise ϵ_n is supposed to be Gaussian and homoscedastic.

If we want to fully identify the ODE, we must estimate both the parameter θ and the initial condition $x(0)$ so that the solution $\phi_{\hat{\theta}}(\cdot, \hat{x}_0)$ of the system fits the observations $y_{0:N-1} = (y_0, \dots, y_{N-1})$. The estimation of ODE parameters by classical approaches (such as least squares [12]) is standard but gives rise to difficult global optimization problem [1]. To solve this kind of problem, variants of least square methods have been recently developed and use approximations of the solution in a spline basis (in the spirit of functional data analysis) as the generalized smoothing proposed by Ramsay et al [17], or two-step estimators [6]. When some states are hidden (typically $m < d$), the estimation (optimization step) is particularly difficult and alternative approaches have been proposed, building on the state-space model interpretation of the couple of equations (1-2). Indeed, Sitz et al. [20] first introduced a state-space model that encapsulates a differential equation in the hidden process and make use of filtering algorithms for deriving an estimate of θ . Subsequent works have exploited the same framework [16, 21, 6], but the initial condition of the system is estimated as a by-product of

the filtering/smoothing steps, and the estimated states are also approximations of the solution of the ODE.

In this work, we keep the same state-space model, and we develop a Bayesian estimator which is a quite natural in state-space models, and permits the use of prior information for ameliorating the estimation. Moreover, our aim is to modify the iterative approach and to show that there is a benefit in jointly estimating θ and the initial condition x_0 in this framework. A classical evaluation of the estimated system $\dot{x}(t) = f(t, x(t), \hat{\theta})$ is to measure the quality of the fit between the true sequence $y_{0:N-1}$ and the predicted sequence $\hat{y}_n = \phi(t_n, (\hat{\theta}, \hat{x}_0))$. Now this simple evaluation requires to know the initial value x_0 of the system, due to the one-to-one relationship between the solution of an Initial Value Problem (IVP) and an initial value x_0 . Hence, despite the little interest of x_0 in general applications, it is in fact fundamental to estimate correctly x_0 in order to disentangle the influence of the parameter from the one of the initial value. Therefore, we suppose that the initial condition x_0 is unknown, so that we are also interested in its estimation. Finally, we want to estimate the augmented initial condition $z_0 = (x_0, \theta) \in R^{d+p}$ of the augmented state ODE model:

$$\begin{cases} \dot{x}(t) = f(t, x(t), \theta(t)) \\ \dot{\theta}(t) = 0 \end{cases} \quad (3)$$

with initial condition $z_0 = (x_0, \theta)$. The solution is the function $t \mapsto \phi(t, z_0)$ from $[0, T]$ to R^{p+d} . For sake of notational simplicity, we will note again this augmented ODE in R^{p+d} with the same vector field f :

$$\dot{z}(t) = f(t, z(t)) \quad (4)$$

and $z(0) = z_0$. Hence, the estimation of z_0 consists only in estimating the initial condition z_0 in (4), from $y_{0:N-1}$:

$$y_n = h(z(t_n)) + \epsilon_n$$

where we keep the notation h for the observation function from R^{p+d} to R^m $h : z = (x, \theta) \mapsto h(x)$. Now, the observed and discretized differential equation (4) fits itself in the (discrete-time) framework of state-space models in R^{p+d} with a deterministic hidden state evolution:

$$\begin{cases} z_{n+1} = z_n + \int_{t_n}^{t_{n+1}} f(\tau, z(\tau), \theta) d\tau \\ y_n = h(z_n) + \sigma \epsilon_n \end{cases} \quad (5)$$

The state-space representation is usually exploited for deriving recursive estimation either in Maximum Likelihood approaches or in Bayesian setting as described in [5]. However, we notice an important feature of the last setting (5): the deterministic evolution of the states implies that we can compute exactly the states at each time from the initial condition (parameter) z_0 , and in particular the hidden part of z_n . In equation (5), the evolution equation can be replaced by the following non-recursive definition of z_n :

$$z_n = \phi(t_n, z_0) = z_0 + \int_0^{t_n} f(\tau, z(\tau), \theta) d\tau \quad (6)$$

ϕ , as a function of the initial state z_0 is the flow of the ODE and describes the way the differential equation move the points in the phase space: the Bayesian estimation of z_0 consists in retrieving the starting point when we observe imperfectly the flow at different times.

2.2 Flow-based Bayesian Estimation

We consider the Bayesian inference framework for the estimation of the augmented initial condition. We call Flow-based Bayesian Estimation (FBE), the Bayesian approach that consists in estimating the augmented initial condition. Since ϵ_n is Gaussian, the likelihood can be written as follows:

$$L(y_{0:N-1}; z_0) \propto \exp(-e(y_{0:N-1}, z_0)) \quad (7)$$

where $e(y_{0:N-1}, z_0) = \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \|y_n - h(\phi(t_n, z_0))\|^2$ is the classical squared error term. In the Bayesian framework, we complete the information on the parameter by a prior distribution for z_0 whose density is π_{-1} , which gives the following posterior distribution

$$\pi_{N-1}(z_0) = p(z_0|y_{0:N-1}) \propto \exp(-e(y_{0:N-1}, z_0)) \pi_{-1}(z_0) \quad (8)$$

As usual, the normalizing constant of the posterior distribution is unknown. Moreover, we have the additional computational complexity due to the absence of closed-form for the flow ϕ . Hence the Bayesian inference relies on the computation of a reliable approximation of $\pi_{N-1}(z_0)$, from which we can derive Bayesian estimators such as the posterior mean $E(Z_0|y_{0:N-1})$ or the Maximum *A Posteriori* (MAP) estimate $\arg \max_{z_0} \pi_{N-1}(z_0)$. In the MAP case, as the flow is highly nonlinear and produces wiggly likelihood functions ([17]), the direct computation of (8) is difficult and the corresponding global optimization algorithm requires intensive computations. This motivates the use of fast approximate optimization or computation of the posterior distribution, which are widely developed for non-linear state-space models. Indeed, the computation of this posterior probability can be done efficiently by recursive smoothing algorithms [5], such as Extended Kalman Filtering/Smoothing (EKF/EKS) [21], Unscented Kalman Filtering/Smoothing (UKF/UKS [20, 16] and more generally sequential Monte Carlo methods (particle filters). These classical algorithms are based on recursive computations of the filtering probabilities $p(z_n|y_{0:n})$ and several versions do exist for the computation of the smoothing probabilities $p(z_n|y_{0:N-1})$. However, in these algorithms, the initial condition is estimated as an initial state and not as a parameter of the flow. The filtering probability is even characterized by the forgetting of the initial condition as the number of observations N tends to infinity. This implies that the more data we observe, the less information we get on the parameter z_0 . As a consequence, the use of refined smoothing strategies remains problematic and calls for careful adaptations ([13, 10]). Moreover, the estimated (smoothed) trajectories of the hidden states are yet not solutions of the ODE on the whole time interval $[0, T]$: the simulated trajectories may differ significantly

from the smoothed trajectories (obtained by Kalman recursions or Particle filtering). Therefore, it might be preferable to turn to a non-recursive estimation of the augmented initial condition based on the non-recursive definition pointed out in (6). Taking into account the flow of the differential equation, we maximize the likelihood of exact solutions starting from different initial conditions, instead of selecting parameters admissible for describing the local transitions (based only the parameter θ).

3 Posterior probability estimation using Population Monte Carlo

To test our hypothesis about the potential interest of a better estimation of the initial conditions in a Bayesian setting, we need to estimate the posterior distribution probability defined in (8). The intractability of the posterior distribution is a well-known problem in Bayesian estimation. Several general simulation methods have been developed such as Markov Chain Monte Carlo (MCMC), Importance Sampling (IS) and variants [18] are commonly used and both are well-suited to the Bayesian setting. However, one difficulty of this Monte Carlo methods is that they can be very (computationally) intensive: this is typically the case for general Hastings-Metropolis algorithms, even if some optimization can be performed. A challenging difficulty of ODE learning is that the evaluation of the likelihood is costly due to the integration of the ODE. This point motivates us to focus on importance sampling algorithms. These methods require only a "reasonable" amount of likelihood evaluations if the importance distribution is not too far from the true posterior distribution. The pitfalls of this method are well-known and are recalled in the next section, but they can be reduced by using some recently introduced adaptive schemes that we will recall.

3.1 Adaptive Importance Sampling and Population Monte-Carlo algorithm

The principle of importance sampling is to use a Monte Carlo approximation derived thanks to a proposal (or importance) distribution q easier to simulate than π_{N-1} and to make a change of measure by introducing the weight function $w = \frac{\pi_{N-1}}{q}$:

$$E_{\pi_{N-1}}(h(Z_0)) = E_q(h(Z_0)w(Z_0)) \simeq \frac{1}{M} \sum_{i=1}^M h(\xi_i)w(\xi_i) \quad (9)$$

where ξ_i are i.i.d. realizations of the distribution q . Hence, the importance sampling estimators are expressed as weighted means $\frac{1}{M} \sum_{i=1}^M \omega_i h(\xi_i)$. Since the posterior distribution is known only up to a normalizing constant, self-normalized importance sampling estimators are rather used i.e. $\sum_{i=1}^M \tilde{\omega}_i h(\xi_i)$, where $\tilde{\omega}_i = \frac{\omega_i}{\sum_{i=1}^M \omega_i}$. The values $\tilde{\omega}_i$ are called the (normalized) importance weight,

and one can interpret identity (9) as the approximation of π_{N-1} by the weighted empirical measure $\hat{\mu}_M(z) = \sum_{i=1}^M \tilde{\omega}_i \delta_{\xi_i}(z)$. Nevertheless, applications of IS can be very delicate as a "good" proposal distribution depends on the unknown target distribution. If the distribution q has a weak overlap with π_{N-1} (i.e. high variance of the so-called importance weights $w(\xi_i) = \omega_i$), then the IS estimators can be very poorly behaved. In that case, we have a so-called weight degeneracy, which means that all the weights vanish except on. This situation can be detected by checking the Shannon entropy of the weighted population. Another pitfall, harder to detect, is when the samples ξ_i have explored insufficiently the tails of the target distribution.

Hence a reasonable prior knowledge of the target distribution is needed to avoid these pitfalls is needed but it remains hard to have especially when dealing with posterior distribution. In order to come up with these limitations, we propose to use the Population Monte Carlo framework developed by Cappé et al [4] for deriving an adaptive Importance Sampling algorithm for dynamical systems.

Population Monte Carlo algorithm Population Monte Carlo (PMC) is a sequential Monte Carlo method, i.e. it is an iterated Importance Sampling Resampling algorithm (ISR) which sequentially moves and re-weights a population of weighted particles $(\xi_i, \tilde{\omega}_i)$, $i = 1, \dots, M$. An essential feature of this algorithm is the resampling step that enables to discard particles with low weights, and to duplicate particles with high weights: this mechanism prevents then the degeneracy of the weights (i.e. all the weights vanish except one), as it is commonly used in particle filtering for instance. In all generality, a PMC scheme is defined for $t = 0, 1, \dots, T$ and a sequence of proposal distributions q_t defined on (R^{d+p})

1. Generate $(\xi_{i,0})_{1 \leq i \leq M} \sim q_t$ (i.i.d sampling) and compute normalized weights $\tilde{\omega}_{i,t}$,
2. Resample $(\tilde{\xi}_{i,0})_{1 \leq i \leq M}$ by multinomial sampling with weights $\tilde{\omega}_{i,t}$, $i = 1, \dots, M$
3. Construct q_{t+1} from $((\tilde{\xi}_{i,t'}, \tilde{\omega}_{i,t'}))_{1 \leq i \leq M, 0 \leq t' \leq t}$

The essential interest of PMC is to introduce a sequence of proposal distributions that are allowed to depend on all the past which enables to consider adaptive IS procedure based on the performance of the previous populations. PMC offers then a great versatility through the construction of the sequence of distribution q_t . In that case, the PMC estimator is still unbiased and the weights depends of step t , i.e. $w_{i,t} = \frac{\pi_{N-1}(\xi_{i,t})}{q_{i,t}(\xi_{i,t})}$. Next, we present a possible construction of a sequence of proposal distributions.

3.2 Markovian transition and adaptive kernels

A simple way to randomly perturb a population is to add an independent noise to each particle $\xi_{i,t-1}$, i.e. to modify independently each particle $\xi_{i,t} = \xi_{i,t-1} + \epsilon_{i,t}$ with $\epsilon_{i,t} \sim N(0, \Sigma_t)$ (usually $\Sigma_t = \sigma_t^2 I_{d+p}$). Then, at each iteration t , we have $\xi_{i,t} \sim N(\xi_{i,t-1}, \Sigma_t)$. General moves from $\xi_{i,t-1}$ to $\xi_{i,t}$ are described with a

(Markov) transition kernel $K_{i,t}(\xi_{i,t-1}, \cdot)$. Through the resampling mechanism, particles moving in good regions are duplicated and particles moving to low credibility regions do vanish which permits a global amelioration of the population. This evolution rule described above is a simple random walk, and the mean size of the jumps is controlled by σ_t . The variance of the proposal is related to the speed at which we do move from an uninteresting region a space to an interesting one. This move is very basic, and it is interesting to propose at least several size of jumps by using a mixture of D Gaussian transition kernels: $\epsilon_{i,t} \sim \sum_{d=1}^D \alpha_d N(0, \Sigma_{d,t})$. With such a D-kernel, the population is moved at each iteration t at different speed $\Sigma_{d,t}$ selected with probability α_d . The D-kernel used in [4] can behave in a satisfying manner, but this algorithm is not fully adaptive as the evolution rule is not updated, whatever the success of the proposed move. Hence a better adaptive kernel is to change the move according to the survival rate of a given move. This problem of determining the weights of the mixture of kernel proposals can be seen as an estimation problem where the weights α_d used are chosen for minimizing a Kullback-Leibler divergence with an EM-like algorithm [7].

4 Experimental results and discussion

In this section, we compare the results provided by Unscented Kalman Smoothing (UKS) and Flow-based Bayesian Estimation using the Importance Sampling scheme (FBE-IS) and the adaptive Population Monte Carlo (FBE-PMC). We measure the quality of the approximation (estimates of the posterior covariance matrices) and also the quality of the reconstruction of the hidden states. In both cases, we consider a relatively small number of observations so that the posterior distribution is far from being approximately Gaussian. Then, we need to use approximation that can take into account multi-modality. In the cases worked out, we have used a so-called multi-start UKS based on 50 random initializations of z_0 for initiating the smoothing algorithm as described in [6]. We select the solution with the smallest quadratic error (along a trajectory) among the 50 different approximated posterior means and the corresponding posterior covariances. We first present results on a nonlinear dynamical biochemical system fully observed with noise. In this case, the state-space model reduces to the discretization of a system of ODEs, observed with some additional gaussian noise. We use this model to test the relevance of the FBE algorithm in a simple case. Then, we turn to the Repressilator which is a partially observed and nonlinear model of a gene regulatory network .

4.1 α -pinene

The α -pinene model presented here is a biochemical system of 5 interacting chemical species. It describes the isomerization of α -pinene, and the dynamics of the concentrations of the 5 species involved is described through the following

time homogeneous linear ODE:

$$\begin{cases} \dot{x}_1 = -(p_1 + p_2)x_1 \\ \dot{x}_2 = p_1x_1 \\ \dot{x}_3 = p_2x_1 - (p_3 + p_4)x_3 + p_5x_5 \\ \dot{x}_4 = p_3x_3 \\ \dot{x}_5 = p_4x_3 + p_5x_5 \end{cases} \quad (10)$$

The evolution of the system is controlled by 5 rate constants $\theta = (p_1, \dots, p_5)$ that we wish to estimate from noisy time series. This estimation problem is relatively classical and it has been introduced as a benchmark for the estimation of ODE, [19]. The system is completely observed and the number of observations is $N = 8$. In [19], the parameter θ has been estimated by global optimization of the least squares criterion. We use their estimate as a reference value (see reference value θ^{ref} in table 1) as it provides a good fit to the data. In this case, the situation is relatively simple as the initial condition x_0 is known and equals to $[100, 0, 0, 0, 0]^T$ and the system is completely observed. For the Bayesian estimation, we use a non-informative uniform distribution for θ and defined on $[\theta^{ref} + 10^{-3}]$. We compare the 3 methods (UKS, FBE-IS and FBE-PMC D-kernel) only for the estimation of the parameters, see table 1, since we can set x_0 to its true value directly in FBE-IS and FBE-PMC. Finally, we use $M = 5000$ particles, and the resampled population of FBE-IS is used as the starting population for FBE-PMC. The proposal for FBE-IS is Gaussian (not centered on θ^{ref}) and is homoskedastic with standard deviation equals to 3×10^{-6} . We use $D = 7$ kernels with different variances: $\sigma_1 = 10^{-11}$, $\sigma_2 = 10^{-10}$, $\sigma_3 = 10^{-9}$, $\sigma_4 = 10^{-8}$, $\sigma_5 = 10^{-7}$, $\sigma_6 = 10^{-6}$, $\sigma_7 = 10^{-5}$. The results in table 1 show that FBE-PMC improves on the other estimates, as it is closer to the reference value, and it gives also a smaller standard deviation than UKS and FBE-IS. Moreover, results provided by FBE-PMC are more reliable than FBE-IS (or UKS), as the entropy of the PMC population $S^{PMC} = 4.3$ is bigger than the entropy of IS population $S^{IS} = 0.15$, which indicates that FBE-PMC enables to avoid degeneracy of the population, and the weights are well scattered. Finally, at iteration $t = 20$ of FBE-PMC, it remains only 1 component with variance 10^{-9} and a population with entropy equals to 4.5. Finally, the reconstructed trajectories obtained by UKS and FBE-PMC are presented in figure ??, which shows that the data best fitting and predicting model is provided by FBE-PMC, thanks to the initial value parametrization.

4.2 An example of a partially observed system: the Repressilator network in *E. coli*

The Repressilator network was proposed in 2000 [8] to describe sustained oscillations observed in a small system in the bacterium *E. coli*, composed of three genes that code for 3 proteins. The first repressor protein, LacI from *E. coli*, inhibits the transcription of the second repressor gene, TetR, from the tetracycline-resistance transposon Tn10, whose protein product in turn inhibits the expression of a third

	Reference ($\times 10^{-5}$)	UKS ($\times 10^{-5}$)	FBE-IS ($\times 10^{-5}$)	FBE-PMC (D-Kernels) ($\times 10^{-5}$)
p_1	5.9259	3.66 ± 5.6	$5.98 \pm 1.3 \times 10^{-2}$	$5.93 \pm 4.7 \times 10^{-2}$
p_2	2.9634	2.5 ± 4.8	$2.92 \pm 1.3 \times 10^{-2}$	$2.96 \pm 5 \times 10^{-2}$
p_3	2.0473	1.78 ± 20.4	$2.05 \pm 5.69 \times 10^{-2}$	$2.06 \pm 2 \times 10^{-2}$
p_4	27.4490	27.3 ± 31.1	$26.7 \pm 5.69 \times 10^{-2}$	$27.89 \pm 10 \times 10^{-2}$
p_5	3.9980	4.24 ± 26	$3.53 \pm 13.1 \times 10^{-2}$	$4.11 \pm 5.2 \times 10^{-2}$
$\ \hat{\theta} - \theta^{ref}\ $	0	2.3×10^{-5}	8.2×10^{-6}	4.5×10^{-6}

Table 1. Estimated parameter values with UKS, FBE-IS and FBE-PMC with standard deviation.

gene, CI from λ phage. Finally, CI inhibits LacI expression, completing the cycle. Hill kinetics are used to model the dynamics. For sake of simplicity in notations, x_1, x_2, x_3 denote respectively the expression of genes LacI, TetR1, CI and x_4, x_5, x_6 the concentrations of corresponding proteins. The network evolution is described by the following ODE:

$$\begin{cases} \dot{x}_1 = v_1^{max} \frac{k_{12}^n}{k_{12}^n + x_5^n} - k_1^{mRNA} x_1 \\ \dot{x}_2 = v_2^{max} \frac{k_{23}^n}{k_{23}^n + x_6^n} - k_2^{mRNA} x_2 \\ \dot{x}_3 = v_3^{max} \frac{k_{31}^n}{k_{31}^n + x_4^n} - k_3^{mRNA} x_3 \\ \dot{x}_4 = k_1 x_1 - k_1^{protein} x_4 \\ \dot{x}_5 = k_2 x_2 - k_2^{protein} x_5 \\ \dot{x}_6 = k_3 x_3 - k_3^{protein} x_6 \end{cases} \quad (11)$$

In the simulations, the RNA concentrations x_1, x_2, x_3 are supposed to be observed through a noisy measurement process (modeled by a Gaussian with $\sigma = 3$) while the proteins concentrations x_4, x_5, x_6 are not measured. The initial values are also supposed to be unknown, and need then to be estimated. The true parameter values and initial conditions are available in tables (2, 3) respectively. We use a Gaussian distribution centered at z_0^{true} as a prior, and the proposal distribution q is a Gaussian distribution with a shifted mean ($z_0^{true} + 2$) and homoskedastic covariance (with standard deviation = 5). For measuring the performance of the 3 estimators, we perform a Monte Carlo study with $N^{MC} = 100$ independent replicates (and we use populations of size $M = 1000$). The mean results in tables (2,3) show that FBE-PMC is unbiased and gives reliable confidence results whereas UKS provide significantly different estimates, and important standard deviation. In the case of Importance Sampling the mean estimates are correct, but the standard deviation are very small. This come from the weights degeneracy of FBE-IS: indeed, in 84% of simulations, a single particle has a weight greater than 90% (the mean entropy of the weights of the IS population is 0.12). This is not the case for FBE-PMC which avoids this curse with $T = 10$ steps starting from the population used by Importance Sampling (the mean entropy is 2.24). In particular from table 3 one can see that FBE-PMC enables to gives more credible values for the hidden states, and guarantees also that the corresponding solution with estimated z_0 is close to the data.

Parameter	True Parameter	UKS	FBE-IS	FBE-PMC
v_1^{max}	150	147.3 ± 0.9	150.2 ± 0.09	150.0 ± 0.46
v_2^{max}	80	81.9 ± 1.7	80.7 ± 0.49	80.2 ± 0.66
v_3^{max}	100	102.2 ± 1.7	100.7 ± 0.25	100.1 ± 0.91
k_1	50	53.0 ± 0.9	50.7 ± 0.05	50.1 ± 0.35
k_2	30	37.1 ± 0.94	30.9 ± 0.08	29.9 ± 0.38
k_3	40	47.6 ± 0.8	40.72 ± 0.03	40.0 ± 0.36

Table 2. Estimated Parameters using UKS, FBE-IS and FBE-PMC (D-kernel) $T = 25$ observations. Average Means and Standard Deviations computed on 100 samples.

Parameter	True Parameter	UKS estimation	FBE-IS	FBE-PMC
$p_1(0)$	1	97.8 ± 5.9	3.11 ± 0.21	2.86 ± 0.09
$p_2(0)$	2	143.6 ± 3.0	3.83 ± 0.21	3.51 ± 0.10
$p_3(0)$	3	148.5 ± 8.6	4.76 ± 0.17	4.75 ± 0.27

Table 3. Estimated Initial Conditions of Hidden States using UKS, FBE-IS and FBE-PMC (D-kernel) approach $T = 25$ observations. Average Means and Standard Deviations computed on 100 samples.

5 Conclusion and perspective

We have proposed to learn both the initial condition and the parameters in such a way that they convey a proper solution of the ODE. As in biological or biochemical experiments, the initial condition vector can be fully observed, we turn the ODE estimation problem into a state-space model estimation task where the only parameter to estimate is an augmented initial condition. A Bayesian approach to this problem, called FBE, has been derived using an Importance Sampling scheme (FBE-IS) and a Population Monte Carlo scheme (FBE-PMC) for the approximation of the posterior probability. The FBE-PMC approach overcomes classical limitations of standard estimation methods in state-space models. The versatility of the PMC schemes gives new estimation methods, based on the learning of proposal distribution q_t that permits a better exploration of the space. The engineering of proposal distributions adapted to the dynamical systems remains quite unexplored and links with particle filters might be pointed out. Finally, a promising research direction for reverse-modelling of biological networks is to combine the augmented initial condition estimation with the graph structure estimation in the Bayesian framework.

Acknowledgement This work was supported by the ANR Projects GD2GS (2006–2009) and ODESSA (2009–2012).

References

- [1] Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R.: Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics* 7(483) (2006)

- [2] Calderhead, B., Girolami, M., Lawrence, N.D.: Accelerating bayesian inference over nonlinear differential equations with gaussian processes. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) *Advances in Neural Information Processing Systems* 21, pp. 217–224. MIT Press (2009)
- [3] Cappé, O., Douc, R., Guillin, A., Marin, J.M., Robert, C.P.: Adaptive importance sampling in general mixture classes. *Statistics and Computing* 18(4), 447–459 (2008)
- [4] Cappé, O., Guillin, A., Marin, J.M., Robert, C.P.: Population monte carlo. *Journal of Computational and Graphical Statistics* 13(4), 907–929 (2004)
- [5] Cappé, O., Moulines, E., Rydén, T.: *Inference in Hidden Markov Models*. Springer-Verlag (2005)
- [6] d’Alché-Buc, F., Brunel, N.J-B.: Learning and inference in computational systems biology. chap. Estimation of parametric nonlinear ODEs for biological networks identification. MIT Press (2010)
- [7] Douc, R., Guillin, A., Marin, J.M., Robert, C.: Convergence of adaptive mixtures of importance sampling schemes. *Annals of Statistics* 35(1), 420–448 (2007)
- [8] Elowitz, M., Leibler, S.: A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338 (2000)
- [9] Gentle, J.E., Hardle, W., Mori, Y.: *Handbook of computational statistics: concepts and methods*. Springer (2004)
- [10] Ionides, E., Breto, C., King, A.: Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 103, 18438–18443 (2006)
- [11] de Jong, H.: Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology* 9(1), 67 – 103 (2002)
- [12] Li, Z., Osborne, M.R., Prvan, T.: Parameter estimation of ordinary differential equations. *IMA Journal of Numerical Analysis* 25, pp. 264–285 (2005)
- [13] Liu, J., West, M.: Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (eds.) *Sequential Monte Carlo Methods in Practice*, pp. 197–217. Springer-Verlag (2001)
- [14] Mendes, P.: Learning and inference in computational systems biology. chap. Comparative Assessment of Parameter Estimation and Inference Methods. MIT Press (2010)
- [15] N.Lawrence, Girolami, M., Rattray, M., Sanguinetti, G.: *Learning and Inference in Computational Systems Biology*. MIT Press (2010)
- [16] Quach, M., Brunel, N., d’Alché-Buc, F.: Estimating parameters and hidden variables in non-linear state-space models based on odes for biological networks inference. *Bioinformatics* 23(23), 3209–3216 (2007)
- [17] Ramsay, J.O., Hooker, G., Campbell, D., Cao, J.: Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society, Series B* 69, 741–796 (2007)
- [18] Robert, C.P., Casella, G.: *Monte Carlo Statistical Methods*. Springer (2004)
- [19] Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R.: Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics* 7(483) (2006)
- [20] Sitz, A., Schwarz, U., Kurths, J., Voss, H.: Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Physical review E* 66, 016210 (2002)
- [21] Sun, X., Jin, L., Xiong, M.: Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks. *PLoS ONE* 3(11), e3758+ November (2008).