

Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference

Minh Quach, Nicolas Brunel and Florence d'Alché-Buc*

IBISC FRE CNRS 2873, University of Evry and Genopole® 523, place des terrasses 91025 Evry, France

Received on June 15, 2007; revised on October 6, 2007; accepted on October 7, 2007

Associate Editor: Chris Stoecker

ABSTRACT

Motivation: Statistical inference of biological networks such as gene regulatory networks, signaling pathways and metabolic networks can contribute to build a picture of complex interactions that take place in the cell. However, biological systems considered as dynamical, non-linear and generally partially observed processes may be difficult to estimate even if the structure of interactions is given.

Results: Using the same approach as Sitz *et al.* proposed in another context, we derive non-linear state-space models from ODEs describing biological networks. In this framework, we apply Unscented Kalman Filtering (UKF) to the estimation of both parameters and hidden variables of non-linear state-space models. We instantiate the method on a transcriptional regulatory model based on Hill kinetics and a signaling pathway model based on mass action kinetics. We successfully use synthetic data and experimental data to test our approach.

Conclusion: This approach covers a large set of biological networks models and gives rise to simple and fast estimation algorithms. Moreover, the Bayesian tool used here directly provides uncertainty estimates on parameters and hidden states. Let us also emphasize that it can be coupled with structure inference methods used in Graphical Probabilistic Models.

Availability: Matlab code available on demand.

Contact: florence.dalche@ibisc.univ-evry.fr

Supplementary information: Supplementary data are available from <http://amisbio.ibisc.fr/dm>

1 INTRODUCTION

Cellular networks (Elowitz and Leibler, 2000; Klipp and Liebermeister, 2006) implement complex mechanisms that enable the cell to respond through time to input signals. Identifying the structure and the parameters of these networks from experimental data is undoubtedly one of the most important challenges in systems biology. Recently, several directions have been simultaneously but independently explored in reverse engineering of metabolic networks, signaling

pathways and transcriptional regulatory. This diversity of approaches is essentially due to the kind of available data.

In the case of metabolic pathways, various modeling frameworks from Flux Balance Analysis (FBA) to Ordinary Differential Equations (ODEs) have been developed. As modelers in this domain often benefit from an important background knowledge, much of the current work is focused on model refinement (Herrgard *et al.*, 2006) using perturbation data. Only few works concern parameter estimation of ODEs with a strong assumption about the network structure.

Regarding signaling pathways, most of the related work in the modeling literature (Klipp and Liebermeister, 2006) also considers that the structure of signaling pathways is known. In this case, the most important task becomes the estimation of (generally) non-linear models based on Hill or mass action kinetics.

On the contrary, in modeling transcriptional regulatory networks, the availability of mRNA concentrations has led researchers to develop algorithms for the estimation of both parameters and structure from prior knowledge and experimental data (Nachman *et al.*, 2004; Perrin *et al.*, 2003; Rangel *et al.*, 2004).

While differences can be stressed between gene regulatory networks, signaling pathways and metabolic networks (Klipp and Liebermeister, 2006), we adopt here a transversal point of view and propose to solve in a unique framework the parameter estimation task when the structure of the network is known. We notice that whichever biological network is under study (metabolic, signaling or regulatory), some of its variables may not be observed, increasing the difficulty of the estimation task. We thus suggest the use of a general framework based on state-space models that accounts not only for non-linear dynamics but also for partially observed systems. Similarly to Sitz *et al.*'s (2002) work developed in a non-biological context, we derive such models from well-grounded Ordinary Differential Equations used in systems biology. This broadens our ability to cover a large variety of biological systems and establishes a bridge between dynamical graphical models and ODEs used in Systems Biology. In non-linear systems, the statistical learning problem is no longer solved in closed form as opposed to linear systems and this raises computational difficulties. In this work, we have chosen to use Unscented Kalman Filtering to tackle

*To whom correspondence should be addressed.

non-linearities. Among other Bayesian approaches, this one is fast and relatively easy to implement. It can be considered as a first step towards more sophisticated approaches such as particle filtering. We illustrate the efficiency of this framework by estimating the parameters and hidden variables of two different systems. The first system is the Repressilator, a synthetic transcriptional regulatory network proposed by Elowitz and Leibler (2000) in order to exhibit how sustained oscillations can be obtained through a simple system of three repressors. We use Michaelis–Menten kinetics with Hill curves to describe it.

The second system under consideration is the JAK-STAT signaling pathway (Swameye *et al.*, 2003; Zi and Klipp, 2006) which takes part in the regulation of cellular responses to cytokines and growth factors. Following the setting introduced by Zi and Klipp (2006), we study the parameter and hidden variables estimation problem when using mass action dynamics.

We show by our derivations that the same state-space model can encompass these two kinds of modeling. Both on artificial and experimental data, the estimation method performs successfully.

2 MODELING BIOLOGICAL NETWORKS WITH NON-LINEAR STATE-SPACE MODELS

Let us consider a biological network composed of p_x variables evolving with time, denoted $x = (x_1, \dots, x_{p_x})$. The vector $\mathbf{x}(t)$ is supposed to represent the state of the network at time t , which is observed at $N+1$ times $t_0=0 < t_1 < \dots < t_N$. The graph structure of the network is known. Thus the functional nature of interactions is contained in a parameter θ and the state evolves in the following way for $n \geq 0$

$$\mathbf{x}_{n+1} = \mathbf{F}_n(\mathbf{x}_n; \theta). \quad (1)$$

The function \mathbf{F}_n has to be chosen according to the kind of network considered (metabolic, signaling or regulatory pathways). We assume that, for biological or experimental reasons, the states \mathbf{x}_n may not be accessible and we can only observe the variables $\mathbf{y} = (y_1, \dots, y_{p_y})$ through the observation functions \mathbf{O}_n , $n=0, \dots, N$

$$\mathbf{y}_n = \mathbf{O}_n(\mathbf{x}_n; \theta) + \epsilon_n^o. \quad (2)$$

ϵ_n^o is a measurement noise chosen as a centered Gaussian noise with covariance R_n . The model defined by Equations (1) and (2) is a state-space model, frequently encountered in engineering science. In general state-space models, \mathbf{x} can have a stochastic evolution, so equation (1) may be replaced by the more general one $\mathbf{x}_{n+1} = \mathbf{F}_n(\mathbf{x}_n, \epsilon_n^x; \theta)$, with $(\epsilon_n^x)_{n \geq 0}$ being a white noise. This assumption also has a biological motivation; for instance, McAdams and Arkin (1999) have shown the intrinsic randomness of gene regulatory networks, where \mathbf{x} represents gene expression levels and concentrations of transcription factors in the cell. Our goal is to provide a general learning framework in which parameters and hidden variables can be estimated from a time series $\mathbf{y}_{0:N} = (\mathbf{y}_0, \dots, \mathbf{y}_N)$.

2.1 Deriving non-linear state-space models from ODEs

Quantitative models of biological networks are usually based on Ordinary Differential Equations (ODE), which means that

the state of the networks is supposed to satisfy the following ODE

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), u(t); \theta). \quad (3)$$

There exist several ways to link state-space models with ODEs, for instance by discretizing time (Perrin *et al.*, 2003). Contrary to the use of integration through time, this implies several limitations for the sampling interval of the observations. Similarly to Sitz *et al.*'s approach (Sitz *et al.*, 2002), we notice that when the state is observed at finite times ($t_{n0} \leq n \leq N$), its evolution can be cast into (1) with functions \mathbf{F}_n and $\mathbf{x}(t_n) = \mathbf{x}_n$ for $n \geq 0$

$$\mathbf{F}_n(\mathbf{x}_n; \theta) = \mathbf{x}_n + \int_{t_n}^{t_{n+1}} \mathbf{f}(\mathbf{x}(\tau), u(\tau); \theta) d\tau. \quad (4)$$

In general, the transition from \mathbf{x}_n to \mathbf{x}_{n+1} is time-dependent: first, because of uneven sampling times [even if the ODE (3) is autonomous] and second, because of the presence of a time-dependent input variable. Finally, the network can be partially observed and the observation process is usually described by Equation (2) with $\mathbf{O}_n(\cdot, \theta) = \mathbf{o}(\cdot)$ being a function independent of n and θ . In the following, we show that this framework encompasses both models of transcriptional regulatory networks and models of signaling pathways.

2.2 Modeling transcriptional regulatory networks with Hill equations

In transcriptional regulatory networks, variables of interest are mRNA and protein concentrations, denoted respectively by r_i and p_i , $i=1, \dots, d$. Let us make the assumption here that one gene can only produce one protein. We consider transcription and translation as dynamical processes, in which the production of mRNAs depends on the concentrations of protein transcription factors (TFs) and the production of proteins depends on the concentrations of mRNAs. Hence, we have $\mathbf{x}(t) = (\mathbf{r}(t)^\top, \mathbf{p}(t)^\top)^\top$ with $\mathbf{r}(t) = (r_1(t), \dots, r_d(t))^\top$ and $\mathbf{p}(t) = (p_1(t), \dots, p_d(t))^\top$. Equation (3) can be split into the following equations:

$$\dot{r}_i = g_i(\mathbf{p}) - k_i^g r_i, \quad (5)$$

$$\dot{p}_i = k_i r_i - k_i^p p_i. \quad (6)$$

Where k_i^g and k_i^p are respectively the degradation rates of mRNA i and protein i . The function g_i describes how TFs regulate the transcription of gene i and Equation (6) describes the production and the degradation of protein i as linear functions where k_i is the translational constant for gene i (Chen *et al.*, 1999).

Various forms have been proposed to model $g_i(\mathbf{p})$, ranging from linear (Chen *et al.*, 1999) to non-linear approaches (de Jong, 2002; Elowitz and Leibler, 2000; Nachman *et al.*, 2004; Smolen *et al.*, 2000). Experimental evidence has suggested that the response of mRNA to TFs concentrations has a *Hill curve* form (de Jong, 2002; Elowitz and Leibler, 2000). The regulation function of transcription factor p_j on its target gene i can be described by $g^+(p_j; v_i^{\max}, k_{ij}, n) = v_i^{\max} (p_j^n / (k_{ij}^n + p_j^n))$ for the activation case and $g^-(p_j; v_i^{\max}, k_{ij}, n) = v_i^{\max} (k_{ij}^n / (k_{ij}^n + p_j^n))$ for the inhibition case. v_i^{\max} is the maximum rate of transcription of

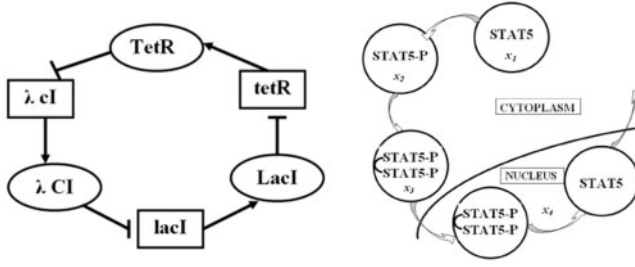


Fig. 1. Left: repressilator. The first repressor protein, LacI inhibits the transcription of the second repressor gene TetR whose protein product in turn inhibits the expression of a third gene *cl*. Finally, CI inhibits *lacI* expression, completing the cycle. Right: JAK-STAT signaling pathway. JAK protein binds to the Erythropoietin Receptor (EpoR) and causes the phosphorylation of STAT5 protein. Phosphorylated STAT5 protein then forms a dimer and moves into the nucleus. In the nucleus, phosphorylated STAT5 dimer is dephosphorylated and forms a STAT5 monomer, which finally goes back to the cytoplasm.

gene i , k_{ij} is the concentration of protein p_j at which gene i reaches half of its maximum transcription rate and n is a steepness parameter describing the shape of sigmoid responses. The parameter $\theta = (v_i^{\max}, k_{ij}, k_i^g, k_i, k_i^g, n)$ for $i, j = 1, \dots, d$ is the set of kinetic constants to be estimated. Note that if a gene has several regulators, the regulatory part of the Equation (5) can be extended into a product of functions g^+ and g^- that expresses the combined effect of regulators. However, we consider here examples where the genes have only one regulator.

Elowitz and Leibler (2000) introduced the Repressilator, a synthetic network based on three transcriptional repressors in order to implement the desired dynamical behavior (sustained oscillations) as illustrated in Figure 1. The system was also built experimentally by genetic engineering with mutated *Escherichia coli* strains. Despite the simplicity of the transcriptional regulation model, the negative feedback loop leads to oscillating concentrations confirmed by experiments. The kinetics of the system can be described by six coupled ODEs, which exactly fit the framework previously described and can be translated into a discrete-time state-space model of form (1). The hidden variables are the protein concentrations with evolution as in Equation (6), and the observations are the (noisily) observed mRNA concentrations r_i , $i = 1, \dots, 3$ which satisfies:

$$\dot{r}_i(t) = g_i^-(p_{[i+1]}; v_i^{\max}, k_{i[i+1]}, n) - k_i^g r_i(t)$$

where $[i+1]$ equals $i+1$ modulo 3.

2.3 Modeling signaling pathways: the JAK-STAT example

Signaling pathways usually involve numerous and various intermediate products in a complex sequence of transformations hence it is quite difficult to describe them in a general way. Depending on the types of signals and intermediary components, and the localization of the pathways, there exist several relevant types of ODEs. Consequently, it seems rather difficult to give the same wide picture as for transcriptional regulatory networks, but most of the time we can say that the ODEs

system involves non-linear reaction rates derived from mass action law and Michaelis–Menten (or Hill) kinetics. We focus here on the JAK-STAT signaling pathway involved in the cellular response to cytokines and growth factors, which involves Janus kinases (JAKs) and Signal Transducers and Activators of Transcription (STATs), see the graph on the right of Figure 1. This pathway transduces the signal carried by these extracellular polypeptides to the cell nucleus, where activated STAT proteins modify gene expression. In both cases, there may be some difficulties to observe the variables of the pathway, and this gives rise to different observation functions \mathbf{o} from gene regulatory networks. This is particularly emphasized in the JAK-STAT pathway, for which it is difficult to discriminate between several intermediates in the pathway. Swameye *et al.* (2003) have suggested an ODE linking the Erythropoietin receptor (EpoR) to the various forms of the STAT5 protein: dephosphorylated STAT5 monomer (x_1) and phosphorylated STAT5 monomer (x_2) in the cytoplasm, phosphorylated STAT5 dimer (x_3) and STAT5 monomer (x_4) in the nucleus. In Swameye *et al.* (2003), the concentration of EpoR is considered as an exogenous variable of the system. The evolution of this network can be described by the following system of coupled differential equations with an input variable $u(\cdot)$ (EpoR), which is an adaptation of the system proposed in Swameye *et al.* (2003) by Zi and Klipp (2006):

$$\begin{cases} \dot{x}_1(t) &= -a_1 x_1(t) u(t) + 2a_4 x_4(t) 1_{\{t \geq \tau\}} \\ \dot{x}_2(t) &= a_1 x_1(t) u(t) - 2a_4 x_2^2(t) \\ \dot{x}_3(t) &= -a_3 x_3(t) + x_2^2(t) \\ \dot{x}_4(t) &= a_3 x_3(t) - a_4 x_4(t) 1_{\{t \geq \tau\}} \end{cases} \quad (7)$$

where $1_{\{t \geq \tau\}}$ denotes the indicator function, equal to 0 for $t \leq \tau$ and equal to 1 otherwise. The concentrations and constants a_i , $i = 1, 3, 4$ in (7) stand for normalized quantities (Zi, 2006). $\theta = (a_1, a_3, a_4)^T$ is the parameter to be estimated. As pointed out by Swameye *et al.*, the individual STAT5 population is difficult to access experimentally, and only the following variables could be measured: $y_1 = (x_2 + 2x_3)$, the concentration of phosphorylated STAT5 in the cytoplasm and $y_2 = (x_1 + x_2 + 2x_3)$, the total amount of STAT5 in the cytoplasm. Thus, the model and data obtained fit the framework of the state space model described in Section 2.1.

3 ESTIMATION WITH UNSCENTED KALMAN FILTERING

3.1 Bayesian estimation

In a Bayesian framework, parameters as well as hidden states are random variables. The goal of inference is to compute the posterior distribution of parameters and initial state, $p(\theta, \mathbf{x}_0 | \mathbf{y}_{0:N})$, given a prior distribution $\pi(\theta, \mathbf{x}_0)$. It is then possible to estimate variances for parameters. Moreover, we benefit from the large family of algorithms developed for this framework: the state-space form given by equations (1) and (2) is exploited for deriving a sequential estimation procedure based on filtering. We then use an extension of Kalman filtering that computes an approximation of the posterior probability and which also gives an approximation of the Minimum Mean Squared Error estimator (MMSE). This method makes it

possible to deal with hidden variables and to estimate them quite simply.

We first recall the general principle of filtering and describe the adaptation of the Kalman filter to the case of non-linear evolution equations, using the Unscented Transformation (UT). We then derive the estimation method for unknown parameters and partially observed states and describe it for the biological models considered in the previous section.

3.2 Filtering

In this section, we remove for sake of clarity parameter θ in Equations (1) and (2), and we describe only the estimation of hidden states. Filtering is the sequential computation of the *posterior* (or filtering) probability $\alpha_n(\mathbf{x}) = p(\mathbf{x}_n | \mathbf{y}_{0:n})$ for $n = 0, \dots, N$ (Cappé, *et al.*, 2005). Without loss of generality, the complete process $\mathbf{x} = (\mathbf{x}_n)_{0 \leq n \leq N}$ may be a Markov (non-deterministic) chain, with values in \mathcal{X} (here $\mathcal{X} \subset \mathbb{R}^{p_x}$). The computation of the filtering probability consists of the alternate and sequential computation of the prediction probability $p(\mathbf{x}_n | \mathbf{y}_{0:n-1})$, $n \geq 0$, the so-called prediction step:

$$p(\mathbf{x}_n | \mathbf{y}_{0:n-1}) = \int_{\mathcal{X}} p(\mathbf{x}_n | \mathbf{x}_{n-1}) \alpha_{n-1}(\mathbf{x}_{n-1}) d\mathbf{x}_{n-1} \quad (8)$$

and its ‘correction’ into $\alpha_n(\cdot)$ (the so-called correction step) by:

$$\alpha_n(\mathbf{x}_n) = \frac{p(\mathbf{y}_n | \mathbf{x}_n) p(\mathbf{x}_n | \mathbf{y}_{0:n-1})}{\int_{\mathcal{X}} p(\mathbf{y}_n | \mathbf{x}_n) p(\mathbf{x}_n | \mathbf{y}_{0:n-1}) d\mathbf{x}_n} \quad (9)$$

We can then derive the sequence of most likely current states characterized by $\hat{\mathbf{x}}_n = \arg \max_{\mathbf{x} \in \mathcal{X}} \alpha_n(\mathbf{x})$, $n = 0, \dots, N$. Note that at $n=0$, the prediction step is replaced by setting $p(\mathbf{x}_0 | \mathbf{y}_{0:-1}) \triangleq \pi(\mathbf{x}_0)$, where $\pi(\mathbf{x}_0)$ is our prior distribution on the initial state.

3.3 Approximate filtering

When \mathbf{X} is a Gaussian and linear Markov process (\mathbf{F} and \mathbf{o} are linear), the prediction-correction algorithm is the well-known Kalman filter that consists of a recursive computation of the mean and covariance of the (Gaussian) distribution $\alpha_n(\cdot)$. This algorithm is no longer valid when the process \mathbf{X} is not Gaussian, nor when the function \mathbf{F} is non-linear, but several extensions have been proposed to tackle non-linearity or non-gaussianity: Extended Kalman Filtering (EKF, Wan and van der Merwe, 2001), Kernel Kalman Filtering (KKF, Ralaivola and d’Alché-Buc, 2004), Unscented Kalman Filtering (UKF, Julier *et al.*, 2000; Wan and van der Merwe, 2001) and Particle Filtering (PF, Doucet *et al.*, 2001). Among the previously published methods, we have focused on UKF that is an approximation of the posterior distributions $\alpha_n(\cdot)$ by Gaussian distributions with mean $\mathbf{m}_n(\mathbf{x})$ and covariance $\Sigma_n(\mathbf{x})$. The UKF has the same computational complexity as EKF but offers a better approximation of the true covariance $\Sigma_n(\mathbf{x})$ and does not require the derivatives of \mathbf{F} to be computed. Compared to the other filtering methods, KKF requires us to define a kernel function that may be difficult to choose if one wants to stick to classical equations of kinetics. PF involves several hyperparameters and, unlike UKF, is based on clouds of randomly generated points of important size that induces numerous ODE

integrations in prediction-correction steps and leads to a higher computational cost. UKF relies on small deterministic sets of appropriately chosen points used in order to mimic the non-linear evolution of the state variable: the so-called sigma points $\xi_{0,n}, \dots, \xi_{2p_x,n}$ (p_x is the dimension of \mathbf{x}). The key idea in UKF lies in the prediction step, where the ‘unscented transformation’ allows one to compute an approximation of the mean $\mathbf{m}_{n+1|n}(\mathbf{x})$ and covariance $\Sigma_{n+1|n}(\mathbf{x})$ of the prediction probability. The mean and covariance of the transformed variable $\mathbf{F}(\mathbf{x}_n)$ (when \mathbf{x}_n has the posterior distribution α_n) can indeed be approximated simply by using the first empirical moments of transformed sigma points chosen as $\xi_{0,n} = \mathbf{m}_n(\mathbf{x})$, $\xi_{i,n} = \mathbf{m}_n(\mathbf{x}) + Q_{i,n}$ and $\xi_{i+p_x,n} = \mathbf{m}_n(\mathbf{x}) - Q_{i,n}$, where Q_n is a square root matrix of $(2p_x + 1/2) \Sigma_n(\mathbf{x})$. Other interesting choices of sigma points are given in Julier *et al.* (2000).

Then, the correction step is carried out in a way similar to Kalman filtering using the classical Kalman gain matrix K_{n+1} and the (approximate) covariance $\Sigma_{n+1|n}(\mathbf{y})$ of the pdf $p(\mathbf{y}_{n+1} | \mathbf{y}_{0:n})$ (see Section 1 in the Supplementary Material for a complete description of the algorithm). The sequence of estimates (*filtered process*) $\hat{\mathbf{x}}_n$ of the hidden variables is the sequence of means $\mathbf{m}_n(\mathbf{x})$.

3.4 Bayesian estimation of parameters

We adapt here the previous general setting to the joint estimation of hidden states and parameters. This can be accomplished using the augmented state vector approach, which consists in rewriting the dynamical system (1), (2):

$$\theta_{n+1} = \theta_n \quad (10)$$

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n; \theta_n) \quad (11)$$

$$\mathbf{y}_n = \mathbf{o}(\mathbf{x}_n; \theta_n) + \epsilon_n^o \quad (12)$$

i.e. the parameter is considered as a hidden state without any temporal evolution. We can use the previous UKF in order to compute the approximation of $p(\theta_n, \mathbf{x}_n | \mathbf{y}_{0:n})$ and we can derive a sequence of (improving) estimates $(\hat{\theta}_n, \hat{\mathbf{x}}_n)$.

The minimizer of the squared error is approximated by $\mathbf{m}_n(\theta)$. Nevertheless, in practice $\mathbf{m}_n(\theta)$ can be a spurious minimizer, and it is recommended to perform several sweeps of the algorithm on the data, i.e. the parameters estimated in the previous sweep are chosen as the initial value for the next sweep.

As described for hidden states, at $n=0$, the prediction step is replaced by setting $p(\theta_0, \mathbf{x}_0 | \mathbf{y}_{0:-1}) = \pi(\theta, \mathbf{x}_0)$. We propose using the rather non-informative hierarchical prior $\pi(\theta, \mathbf{x}_0) = \pi(\theta)\pi(\mathbf{x}_0)$, where $\pi((x_i)_i) = \prod_i \mathcal{N}(\mu_{x_i}, \sigma_{x_i}^2)$, with $\mu_{x_i} \sim \mathcal{U}([0, \lambda_i])$, i.e. all of the components of the vector are independent and Gaussian, with a mean drawn according to a uniform distribution whose support is determined by a hyperparameter λ_i computed from the data, and the variance $\sigma_{x_i}^2$ is a fixed value (depending on the data). However, if a certain constraint concerning the initial value is made available, more informative prior could be used.

The estimation procedure for the Repressilator and the JAK-STAT model is done by considering the stacked state variable $\tilde{\mathbf{x}}_n = (\theta_n^\top, \mathbf{x}_n^\top)^\top$. The state components of the points $\xi_{i,n}$ are computed using by $F_n(\cdot, \theta_n)$, i.e. by integration of the ODE described in Section 2.1. For the repressilator, the $\chi_{i,n}$ are

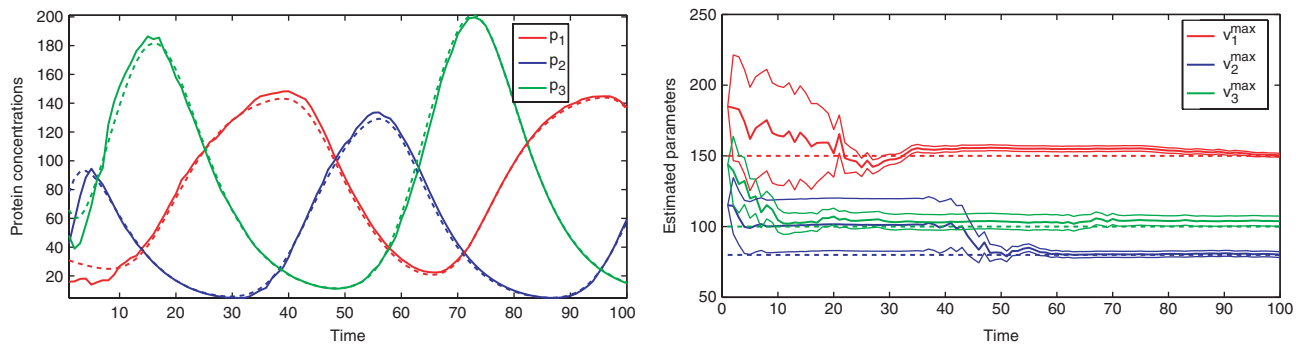


Fig. 2. Left: the evolution of the true (dashed) and estimated (solid) protein concentrations. Right: recursive estimation of the maximal rate of Michaelis–Menten kinetics through time for the sampling interval $\Delta_t = 0.2$ (corresponds to 100 data points). Dash lines: true parameters. Solid lines: estimated parameters. For each parameter, the bold solid line shows the mean of the filtering distribution and the thin curves show the confidence interval with one SD.

vectors of dimension 3 corresponding to the gene expression levels. For the JAK-STAT model, the $\chi_{i,n}$ are vectors of dimension 2 computed from $\xi_{i,n}$ by the linear combinations $x_2 + 2x_3$ and $x_1 + x_2 + 2x_3$.

4 RESULTS

We first illustrate our approach on artificial data generated from the Repressilator model and second, on experimental data of the JAK-STAT pathway. Simulation studies are useful in providing insights into the strengths and weaknesses of learning algorithms, such as robustness against numerous choices of settings, including the quantity of observed data, the sampling interval for observing the data, the number of time points in the observed time series.

The size of the systems used in the experiments are quite representative of the size of the models that the proposed method (based on UKF) can efficiently handle, i.e. around 10 variables and parameters. In higher dimensions, the recursive optimization using the UKF approximation can lead to spurious minimizers, but such a limitation can be partly overcome by using a better approximation, such as the Particle Filter. The main limitation of the approach depends on the respective sizes of the observed and hidden parts of the system.

4.1 Parameter estimation of the Repressilator

4.1.1 Simulated data We start from the equations given in (2.2) and fix the following values of the parameters according to the stability study presented in Elowitz and Leibler (2000): $k_1^p = 1$, $k_2^p = 2$, $k_3^p = 3$, $k_1^g = k_2^g = k_3^g = 1$, $v_1^{\max} = 50$, $v_2^{\max} = 80$, $v_3^{\max} = 100$, $k_{12} = 50$, $k_{23} = 30$, $k_{31} = 40$ and $n = 3$. The components of the initial state are drawn independently from a uniform distribution on $[0, 100]$ (arbitrary units). Simulations are performed using the MATLAB numerical integrator *ode45* over the time interval $[0, T]$, with $T = 20$. The observation noises ϵ_n^o are added to three observed variables to mimic gene expression data and the SD of ϵ_n^o shown in the experiments is chosen to be equal to 20% of the SD of the states. The robustness of the method has been tested with respect to a higher noise level (30%, 40%), and similar results for the estimation for the states and parameters have been obtained.

The estimated predicted variance and the variance of the estimators increase, although no systematic divergence of the method has been detected.

During the simulation, measurements are sampled at a fixed interval Δ_t , so that for each experiment a time series containing T/Δ_t time points is collected. We assume that the learning problem consists in identifying the following six parameters: v_1^{\max} , v_2^{\max} , v_3 , k_{12} , k_{23} and k_{31} while the degradation rates for proteins and mRNAs are known. In order to learn the true parameters, we use a multi-start approach by sampling $I = 50$ different initial states and parameters from our prior $\pi(\theta, \mathbf{x}_0)$, so that we compute 50 filters in parallel. Our final state and parameter estimates are simply the mean of the prediction of the 50 different filters (an alternative way to combine the different filters would be to select the filter with the lowest prediction error). The Gaussian priors for the parameter are such that $\lambda_i = 2 \times \theta_i^*$ and $\sigma_{\theta_i} = 0.2 \times \theta_i^*$, and for the unobserved variables $\lambda_i = 2 \times \mathbf{x}_0^i$ and $\sigma_{\mathbf{x}_0^i}$ is set to 20% of the SD of the state \mathbf{x}_i . For the observed variables, the prior is also Gaussian with mean $\mu_{\mathbf{x}_0^i} = y_0^i$ and the same formula as for the unobserved variables is used for the SD.

4.1.2 Evaluation of parameter estimation The filtered protein concentrations and parameters using UKF are shown in Figure 2 for $\Delta_t = 0.2$, (see also Fig. 1 in Supplementary Material for the case of one experiment: one time series corresponding to one initial condition). The filtered concentrations of proteins quickly adjust to the true protein trajectories. The SDs of the estimators are estimated using the square-root of the diagonal of the matrix $\Sigma_n(\mathbf{x})$. Parameter estimates start far from the true value with high SDs, but they gradually converge to the true parameters. The small values of the final SDs of the estimates point out the convergence of the learning algorithm. Finally, since a single sweep of UKF on a time series containing 100 observations takes less 5 s, our multi-start approach gives an estimation within ~ 250 s.

4.1.3 Dependency between the prediction error of the hidden states and the sampling interval In order to analyze the dependency between the prediction error and the sampling interval, we used different sampling intervals (respectively) $\Delta_t = 0.2, 0.4, 0.8, 1$ and 2 , corresponding to (respectively) 100,

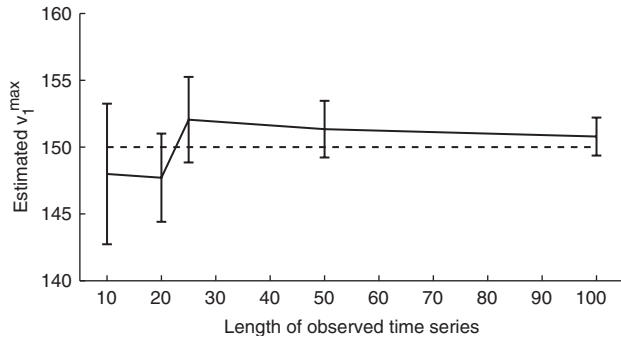


Fig. 3. Estimation of parameter v_1^{\max} versus the length of observed time series. The dash line shows the true parameter. The bars show means and SDs of estimated parameter.

50, 25, 20 and 10 time points. The estimates for all parameters are reported in Table 1 of the Supplementary Material. We plot the estimation results for parameter v_1^{\max} in Figure 3. Obviously, the estimates are closer to the true values with smaller SD when there are more time points available. In order to compare errors for the different cases, we introduce the normalized Mean Squared Error between the true and estimated trajectories:

$$\text{MSE} = \frac{1}{N+1} \sum_{n=0}^N (\hat{\mathbf{x}}_n - \mathbf{x}_n^{\text{true}})^{\top} \text{diag}(\Sigma_n^{-1}) (\hat{\mathbf{x}}_n - \mathbf{x}_n^{\text{true}})$$

where Σ_n is the covariance matrix of the UKF estimate at time n and $\text{diag}(A)$ is the diagonal matrix equal to the diagonal of a square matrix A . The errors for each component of the state variable are rescaled in order for them to be comparable (though the covariance between the components is not taken into account). The MSE is plotted in Figure 2 of the Supplementary Material and increases when there are less time points. However, a relevant result is still obtained for only 10 observed time points. As we can see in the final column of Table 1 of the Supplementary Material, the true parameters stands within $\pm\sigma$ confidence interval of the estimated parameters.

4.1.4 Dependency between the prediction error of the parameters and the number of repeated experiments We show here that the influence of the number of different experiments (i.e. time series corresponding to the observation of the same system but with different initial conditions). The learning algorithm can be adapted to this setting in a straightforward manner, and we show that it is possible to deal with more difficult situations. As an illustration, we assume that the parameters k_1, k_2, k_3 are also unknown, so that we have to estimate nine parameters from several time series (with 50 observations each). The MSE's obtained for a number of experiments varying from 1 to 6 are reported in Figure 4 of the Supplementary material. The MSE decreases when more experiments are available. One notes that three different experiments provides the same mean level of MSE as six experiments but with higher variance. This may help in designing experiments and gives some hints for obtaining accurate estimates of the parameters from only a

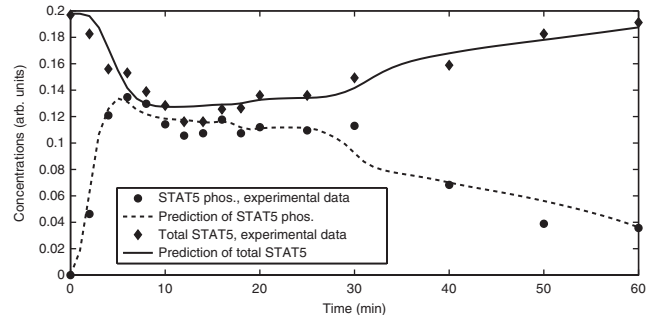


Fig. 4. Prediction of phosphorylated STAT5 and total amount of STAT5.

few experiments. We also plot the estimation of parameter v_1^{\max} versus the number of repeated experiments in Figure 3 of the Supplementary Material. The estimated parameter tends to the true value with smaller SDs when the number of experiment increases.

4.2 Parameter estimation for the JAK-STAT pathway model using experimental data

Experimental data of JAK-STAT pathways from Swameye *et al.* (2003) was used. Time series of two observed variables y_1 (the total concentration of phosphorylated STAT5) and y_2 (the total concentration of STAT5 in the cytoplasm) are measurable. Each time series contains 16 time points sampled at $t=[0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 40, 50, 60]$ min. Data for the input EPoR phosphorylation is also available. Here, we use a linear interpolation in order to obtain a continuous time input. We initialize the parameters a_1, a_3, a_4 and the initial condition x_1 independently with a uniform distribution on $[0, 5]$. We plot the normalized MSE between the predicted time series and the data in Figure 4 of the Supplementary Material. The convergence of this curve shows the stability of the learning algorithm, and ensures that we have reached a local minima. Eventually, the parameter estimates (with SDs) are $\hat{a}_1 = 0.0515 \pm 0.0055$, $\hat{a}_3 = 3.39 \pm 0.45$ and $\hat{a}_4 = 0.35 \pm 0.047$, and the prediction for the observed variables y_1 and y_2 are shown in Figure 5, which shows a good fit of the learned model. We also check the coherence of the estimation by simulating the JAK-STAT pathway with these estimates. A new time series \mathbf{x}^* is simulated from (7) with initial conditions $x_1=0.2, x_2=x_3=x_4=0$ and the estimated parameters. The result in Figure 5 showed that the learned model is able to predict well the four unobserved components of \mathbf{x}^* , so we may have a higher confidence in the prediction of the unobserved variables.

5 DISCUSSION

The approach we derived for non-linear biological systems has already been proposed in another context by Sitz *et al.* (2002). However, the present work is a novel application of this approach in Systems Biology, which opens new perspectives in estimating non-linear biological systems. So far, linear state-space models have been mainly used for transcriptional

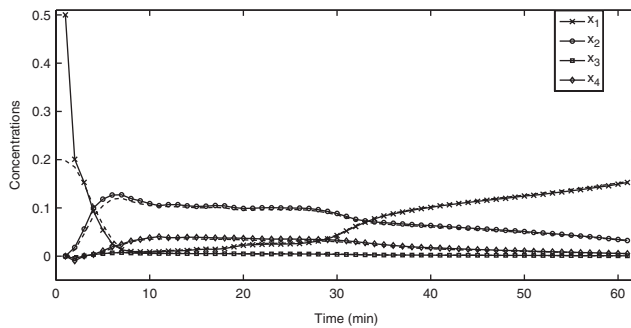


Fig. 5. The evolution of the simulated (dashed) and estimated (solid) concentrations of the four unobserved variables.

regulatory networks. Perrin *et al.* (2003) proposed IDBN (Inertial Dynamic Bayesian Network), a second-order linear ODE model that accounts for inertia and allows us to represent damped oscillations: in this model hidden variables are the true mRNA concentrations and their derivative. Parameters and hidden variables have been estimated in this context by linear Kalman filtering and smoothing (d'Alché-Buc *et al.*, 2005). Rangel *et al.* (2004) introduced also a linear dynamical probabilistic model for which the biological interpretation is less explicit but this work has been interestingly followed by a study on hidden variable estimation in Beal *et al.* (2005). However, a higher level of detail is often required in order to improve the biological relevance of the models. This is why Nachman *et al.* (2004) suggested a non-linear state-space model based on Michaelis–Menten equations but only to cope with transcription factors. In their model as well as in Rogers *et al.* (2007), the equation taking into account protein production and degradation is not used. Our model in the case of regulatory networks can thus be seen as an extension to a full modeling of protein and mRNA concentrations through time.

Another interest in the state-space interpretation of ODE model lies in the access to new estimation methods, which could be faster than classical ones and could also be able to incorporate priors on the system. Indeed, most of the methods proposed so far for signaling pathway consist of the minimization of a least squares criterion (Li *et al.*, 2005; van Riel and Sontag, 2006) without any regularization term. Even when the model is slightly complex (around 10 variables), the minimization step requires special care in order to reach the true maxima, because classical local optimization methods (gradient-based or Newton-like) are doomed to reach local minima. In the end, global optimization techniques have been proposed in order to solve these problems such as simulated annealing, evolutionary algorithms, (Koh *et al.*, 2006; Moles *et al.*, 2003; Polisetty *et al.*, 2006). These techniques are batch methods that do not use the recursive structure induced by the ODE model. In contrast to the former, the alternate regression (Chou *et al.*, 2006) or the system perturbation method (Schmidt *et al.*, 2005) exploit the particular structure of the learning problem in order to derive relatively simple algorithms. Our method takes advantage of the dynamical nature of the model by implementing a recursive optimization. Though the UKF is only able to

reach a local minimum of the posterior probability, it delivers a sequence of estimated states \hat{x}_n and parameters $\hat{\theta}_n$ which are likely for all the intermediate estimation problems with observations $y_{0:n}$. This sequence of estimates remains plausible, at least when the initial conditions are correctly chosen (which can be done in some case with the literature). We have used in our experiments simple priors (flat or Gaussian priors), but Bayesian estimation may benefit from more elaborated prior distributions in order to favor meaningful regions of the parameter and state spaces. Moreover, the most striking property of this estimator is its ease of implementation and above all its speed, which is an advantage over global optimization methods. One should also note that the variance of the estimator ($\hat{x}_N, \hat{\theta}$) is simultaneously computed, whereas it is not straightforward to compute it in batch methods since the model can be too complex to be done analytically or approximately. Moreover, our estimation method can still be enhanced by the use of smoothing probabilities and the promising results of UKF calls for more sophisticated filtering approaches such as particle filtering.

6 CONCLUSION AND PERSPECTIVES

We have presented non-linear state-space models for describing biological networks and non-linear filtering approaches to estimate both parameters and hidden variables. As the models are built up from ODEs, they benefit from all the existing background in biological modeling with ODEs and thus are ensured to exhibit high biological relevance. This point was illustrated on two different kinds of networks models: a transcriptional regulatory model based on Hill kinetics and a signaling pathway model based on mass action kinetics. Let us notice that, given the type of equations we already dealt with, there is no reason not to apply this approach to model metabolic networks and to estimate their parameters as soon as experimental data are available. Moreover, this work raises several issues that encourage further works. First, our estimation algorithm like others in literature requires time series of sufficient length to be efficient. In this case, we need to reduce the complexity of the parameter space by introducing relevant biological priors. The Bayesian framework we use is appropriate for this. Second, large networks with large number of parameters may not be identifiable. In order to overcome this limitation, we suggest applying the decompositional scheme developed in Koh *et al.* (2006) in order to work only on subnetworks. Third, it should be emphasized that our parameter estimation method can be coupled with any of the classical structure learning schemes used in graphical probabilistic models (MCMC, evolutionary approaches) in order to fully reverse-engineer biological networks. Finally, it should also be stressed that this framework could account for joint modeling of metabolic, signaling and regulatory networks if one can deal with the various time scales and has access to appropriately observed time series.

ACKNOWLEDGEMENTS

This research has been mainly funded by Genopole® through an ATIGE grant and a postdoc grant. ANR (National Agency

for Research) has also contributed to fund the project through the ARA 2005 call for projects.

Conflict of Interest: none declared.

REFERENCES

- Beal, M. et al. (2005) A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, **21**, 349–356.
- Cappé, O. et al. (2005) *Inference in Hidden Markov Models*. Springer-Verlag.
- Chen, T. et al. (1999) Modeling gene expression with differential equations. In *Pacific Symposium of Biocomputing*.
- Chou, I.-C. et al. (2006) Parameter estimation in biochemical systems models with alternating regression. *Theor. Biol. Med. Model.*, **3**, 1–15.
- d'Alché-Buc, F. et al. (2005) Dynamic model of gene regulatory networks based on inertia principle. *Bioinformatics using Computational Intelligence Paradigms*, volume 176 of *Studies in Fuzziness and Soft Computing*. Springer, 93–117.
- de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, **9**, 67–103.
- Doucet, A. et al. (2001) *Sequential Monte Carlo Methods in Practice*. Springer-Verlag.
- Elowitz, M. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.
- Herrgard, M. J. et al. (2006) Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS Comput. Biol.*, **2**, 676–686.
- Julier, S. et al. (2000) A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Automat. Contr.*, **45**, 477–482.
- Klipp, E. and Liebermeister, W. (2006) Mathematical modeling of intracellular signaling pathways. *BMC Neurosci.*, **7** (Suppl. 1).
- Koh, G. et al. (2006) A compositional approach to parameter estimation in pathway modeling: a case study of the Akt and MAPK pathways and their crosstalk. *Bioinformatics*, **22**, 271–280.
- Li, Z. et al. (2005) Parameter estimation of ordinary differential equations. *IMA J. Numer. Anal.*, **25**, 264–285.
- McAdams, H. and Arkin, A. (1999) It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.*, **15**, 65–69.
- Moles, C. et al. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, **13**, 2467–2474.
- Nachman, I. et al. (2004) Inferring quantitative models of regulatory networks from expression data. *Bioinformatics*, **20** (Suppl. 1), 248–256.
- Perrin, B.-E. et al. (2003) Inference of gene regulatory network with Dynamic Bayesian Network. *Bioinformatics*, **19**, i386–i49.
- Polisetty, P. K. et al. (2006) Identification of metabolic system parameters using global optimization methods. *Theor. Biol. Med. Model.*, **3**, 1–15.
- Ralaivola, L. and d'Alché-Buc, F. (2004) Dynamical modeling with kernels for nonlinear time series prediction. In *Advances in Neural Information Processing Systems (NIPS'04)*.
- Rangel, C. et al. (2004) Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics*, **20**, 1361–1372.
- Rogers, S. et al. (2007) Bayesian model-based inference of transcription factor activity. *BMC Bioinformatics (special issue from PMSB'06)*.
- Schmidt, H. et al. (2005) Identification of small scale biochemical networks based on general type system perturbations. *FEBS J.*, **272**, 2141–2151.
- Sitz, A. et al. (2002) Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Phys. Rev. E*, **66**, 016210.
- Smolen, P. et al. (2000) Modeling transcriptional control in gene networks - methods, recent results, and future directions. *Bull. Math. Biol.*, **62**, 247–292.
- Swameye, I. et al. (2003) Identification of nucleocytoplasmic cycling as remote sensor in signaling by databased modeling. *Proc. Natl. Acad. Sci. USA*, **100**, 1028–1033.
- van Riel, N. and Sontag, E. (2006) Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach. *IEEE Proc. Syst. Biol.*, **153**, 263–274.
- Wan, E. A. and van der Merwe, R. (2001) The Unscented Kalman Filter. Series in adaptive and Learning Systems for Signal Processing, Communications, and Control. In Haykin, S. (ed.), *Kalman Filtering and Neural Networks*. Wiley, pp. 221–283.
- Zi, Z. (2006) *User Guide: SBML-PET, A Systems Biology Markup Language Based Estimation Parameter Tool*. Max Planck Institute for Molecular Genetics, Available from www.sysbio.molgen.mpg.de/SBML-PET/.
- Zi, Z. and Klipp, E. (2006) SBML-PET: a systems biology markup language-based parameter estimation tool. *Bioinformatics*, **22**, 2704–2705.