

State and parameter estimation of the heat shock response system using Kalman and particle filters

Xin Liu and Mahesan Niranjan*

School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Traditional models of systems biology describe dynamic biological phenomena as solutions to ordinary differential equations, which, when parameters in them are set to correct values, faithfully mimic observations. Often parameter values are tweaked by hand until desired results are achieved, or computed from biochemical experiments carried out *in vitro*. Of interest in this article, is the use of probabilistic modelling tools with which parameters and unobserved variables, modelled as hidden states, can be estimated from limited noisy observations of parts of a dynamical system.

Results: Here we focus on sequential filtering methods and take a detailed look at the capabilities of three members of this family: (i) extended Kalman filter (EKF), (ii) unscented Kalman filter (UKF) and (iii) the particle filter, in estimating parameters and unobserved states of cellular response to sudden temperature elevation of the bacterium *Escherichia coli*. While previous literature has studied this system with the EKF, we show that parameter estimation is only possible with this method when the initial guesses are sufficiently close to the true values. The same turns out to be true for the UKF. In this thorough empirical exploration, we show that the non-parametric method of particle filtering is able to reliably estimate parameters and states, converging from initial distributions relatively far away from the underlying true values.

Availability and implementation: Software implementation of the three filters on this problem can be freely downloaded from <http://users.ecs.soton.ac.uk/mn/HeatShock>

Contact: m.niranjan@southampton.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on September 26, 2011; revised on February 22, 2012; accepted on April 2, 2012

1 INTRODUCTION

In systems biology, sets of ordinary differential equations (ODEs) are often used to characterize biochemical reactions. The differential equations capture our knowledge of the underlying interactions in a quantitative way and solutions to such systems of equations help explain behaviour at an overall systems level. Much of the work in the area deals with deterministic differential equations, often non-linear (e.g. Hill kinetics). Unknown parameter values in their specification, such as reaction rates, are obtained from biochemical

experiments conducted *in vitro*, or are set to specific values by elaborate hand-tuning, so that a set of observations from the system under study are best explained. Sometimes, such parameters may not have direct biological interpretations and are used as approximations (Lillacci and Khammash, 2010). The yeast cell cycle model of (Chen *et al.*, 2000) is a good example of this.

The parameter estimation problem has often been posed as a search and optimization problem in the literature. For example, (Mendes and Kell, 1998) have explored a range of optimization methods including steepest descent gradient search techniques and methods suitable for global optimization, such as simulated annealing and genetic programming, in the parameter estimation of metabolic systems. An alternate method using spline approximation of the solution, using linear and non-linear programming is described in a recent paper (Zhan and Yeung, 2011). These authors consider an enzyme kinetic system and a subset of a cell cycle model. Ashyraliyev *et al.* (2008) show how parameters of a developmental gap gene circuit model may be estimated by extensive search methods.

Such approaches to model-based explanation of biological phenomena, often do not take into account system and measurement noise in the modelling process. They also do not explicitly seek to infer parameter values or unobserved states from noisy measurements. Techniques using Bayesian inference are alternatives to optimization-based approaches, but having the added motivation of being able to capture uncertainties in parameter and state estimates. Examples of work along these lines include (Dewar *et al.*, 2010; Golightly and Wilkinson, 2005) who consider stochastic systems. Barenco *et al.* (2006), Jayawardhana *et al.* (2008), Lawrence *et al.* (2006) and Vyshemirsky and Girolami (2008) address ODE-based systems for which parameter estimation is performed by Markov Chain Monte Carlo (MCMC) methods in a probabilistic inference framework.

Some authors have recently studied the role of sequential estimation methods for state and parameter estimation from systems biology models, formulated as state-space models. Lillacci and Khammash (2010), Nakamura *et al.* (2009), Quach *et al.* (2007), Sun *et al.* (2008) and Yang *et al.* (2007) fall into this category. The first three of these use parametric methods based on Kalman filtering, whereas the latter two use the non-parametric approach of particle filtering. The power of the tools we explore in this article have been demonstrated in other areas of application over many decades. These include the adaptive estimation of neural networks (Kadirkamanathan and Niranjan, 1993), target tracking from bearing-only measurements (Bar-Shalom *et al.*, 2001), modelling futures contracts in computational finance (Niranjan, 1997) and to

*To whom correspondence should be addressed.

find global minima of artificial neural networks (de Freitas *et al.*, 2000). In the context of biology, Dewar *et al.* (2010), Quach *et al.* (2007) and Wilkinson (2009) are examples of using probabilistic dynamical systems models to characterize biological systems and to make parameter estimation and inferences from them. While most literature on the subject focuses on batch-based models, in which parameter estimation and inference are performed on all the available data, our attention is on sequential, or online, methods. The Kalman filter and its variants, and the particle filter (PF) belong to this family.

The formulation of jointly estimating state and parameters by an extended state representation we pursue is due to Sitz *et al.* (2002), who applied unscented Kalman filter (UKF) to analyze two dynamical systems: the Lotka–Volterra system (Hofbauer and Sigmund, 1988) and the Lorenz system (Colin, 1982). Unknown parameters are treated as states with their time variation set to zero. To motivate non-parametric particle filtering, with the argument that computation is sufficiently cheap to be able to do this, Nakamura *et al.* (2009) used an exceedingly large number of Monte Carlo samples to completely cover all the possible states and parameters space for a complex system such as the mammals circadian genetic control model (Matsuno *et al.*, 2005) which has 45 unknown parameters. While this is an ambitious attempt, motivated under the term *peta-computing*, we do not believe the authors make a persuasive case for flooding the space with particles. Instead, as we find in this work, a modest number of particles, of the order of a few thousands, offer very attractive performance in parameter estimation. Additionally, Sun *et al.* (2008) applied extended Kalman filter (EKF) to simultaneously estimate the states and parameters of two real datasets [i.e. JAK/STAT signal transduction pathway (Timmer *et al.*, 2002) and Ras/Raf/MEK/ERK regulatory pathway (Kolch, 2000)].

The remainder of this article is organized as follows. In Section 2, we give the generic formulation of systems biology models cast in state-space form, in Section 3, we describe the three algorithms, in Section 4, we define the heat shock response model we use as our test problem, and in Section 5 we present our results and discuss them. Typical results needed to support our conclusions are presented in the main body of this article, and more detailed results, covering the estimation of various parameter combinations being estimated, are given as online Supplementary Material accompanying this article in the journal's website.

2 NON-LINEAR STATE-SPACE MODELS WITH ODES

Non-linear state-space models of interest here consist of two sets of equations: a set of state equations describing the dynamics of the biological system under study, and a set of output equations that define how observations arise from the state at any time. In the class of systems we consider here, the dynamical system is deterministic, in that there is no process noise, and the observations are assumed to be corrupted by additive noise.

Suppose a biological system consists of n state variables, denoted $\mathbf{x} = (x_1 \dots x_n)$. In every time step, the states of the system are represented by vector $\mathbf{x}(t)$, which might consist of concentrations of chemical species of interest, e.g. mRNA, proteins or metabolites. Outputs, related to the state vector \mathbf{x} through the function $h(\cdot)$, denoted as y , quantify the observations.

Dynamics of the system, characterized by ODEs, are in general written as follow:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \\ \mathbf{x}(t_0) &= \mathbf{x}_0,\end{aligned}\quad (1)$$

where the vector $\boldsymbol{\theta} = [\theta_1 \dots \theta_k]^T$ represents the unknown parameters that are going to be estimated. The extended state representation treats the unknown parameters as constants and imposes zero dynamics on them. With this state extension, (Sitz *et al.*, 2002), the system becomes:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \\ \dot{\boldsymbol{\theta}} &= \mathbf{0} \\ \mathbf{x}(t_0) &= \mathbf{x}_0 \\ \boldsymbol{\theta}(t_0) &= \boldsymbol{\theta}_0\end{aligned}\quad (2)$$

Here, we have a continuous-time underlying dynamical process which is observed at discrete times. Following Sitz *et al.* (2002) and other authors, the way to solve this problem is to numerically integrate the state dynamics between temporal points at which observations are made, using parameter values estimated in the previous point in time:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \int_{t-1}^t \mathbf{f}(\mathbf{x}(\tau), \boldsymbol{\theta}) d\tau \quad (3)$$

By doing so, we are essentially discretizing the continuous-time process. Finally, it is possible to describe the partial observations of the system through the following equation:

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}_t) + \boldsymbol{\omega}_t \quad (4)$$

where $\boldsymbol{\omega}_t$ represents the white Gaussian noise with covariance matrix \mathbf{R} . $h(\cdot)$ is the output function of the system.

While the constant unknown parameters ought to have zero dynamics, it often helps to impose a random walk model on them, in order to enable a tracking behaviour which helps convergence from some arbitrary initial guess to their true values. Several authors including (Kitagawa, 1998) have suggested this. There is extensive discussion in the literature on the ratio between the assumed noise variance of the random walk and the noise variance of the measurements being the determinant of convergence in the Kalman filter setting [e.g. Bar-Shalom *et al.* (2001)].

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^\theta \quad (5)$$

However, this random walk may sometimes result in divergence and posteriors will far away from the true values. To deal with this, the method proposed in Liu and West (2001) is adopted. The posteriors are guaranteed to converge, if the variance of the random walk model decay with time. Liu and West (2001) intended using kernel smoothing with shrinkage for parameter evolution as follows:

$$\begin{aligned}\boldsymbol{\theta}_t &= a\boldsymbol{\theta}_{t-1} + (1-a)\bar{\boldsymbol{\theta}}_{t-1} + \boldsymbol{\omega}_t^\theta \\ a &= \frac{3\delta - 1}{2\delta}\end{aligned}\quad (6)$$

δ is a discount factor that lies in $[0, 1]$, $\bar{\boldsymbol{\theta}}_{t-1}$ is mean of the particles in time instance $t-1$ and $\boldsymbol{\omega}_t^\theta \sim \mathcal{N}(0, m^2 \mathbf{V}_{t-1})$ is the additive noise to parameter evolution. $m^2 = 1 - a^2$, \mathbf{V}_{t-1} is the variance matrix of

the parameters at time $t - 1$. This is a crucial aspect of implementing such sequential methods.

3 SEQUENTIAL ESTIMATION METHODS

In the Bayesian framework, we seek the joint posterior distribution of parameters and states, given the measurements, denoted as $p(\theta, \mathbf{x}_t | \mathbf{y}_t)$. the algorithms of interest here recursively estimate this posterior as data arrives one at a time. We first present the three algorithms of the EKF, UKF and PF.

3.1 Extended Kalman filter

The well-known Kalman filter (Kalman, 1960) is an optimal state estimator in the case of linear, Gaussian systems. When either of these conditions is violated, the resulting probability density over the unknown states is no longer Gaussian. The EKF approximates the non-linearities by Taylor series expansion about the current operating point and truncates. When the truncation is to first- and second-order terms, closed form update equations are obtainable (Bar-Shalom *et al.*, 2001). Truncating to the first-order (gradient) term is the most widely used form of EKF, which we pursue here.

In order to recursively estimate, we assume the initial condition $\hat{\mathbf{x}}_0^+$ and its error covariance matrix \mathbf{P}_0^+ are known. The current prior state distribution termed $\hat{\mathbf{x}}_t^-$, obtained by integrating the continuous-time state process in the time interval $[t-1, t]$ with the previous posterior estimate treated as initial condition. The current prior error covariance distribution termed \mathbf{P}_t^- , obtained by integrating a differential Lyapunov equation and initial condition is the posterior error covariance of previous state (Simon, 2006). We now define an extended state vector consisting of the states and unknown parameters, following Lillacci and Khammash (2010).

$$\mathbf{s} = \begin{bmatrix} \mathbf{x} \\ \theta \end{bmatrix} \quad (7)$$

$$\hat{\mathbf{s}} = f(\hat{\mathbf{s}}) \quad (8)$$

$$\hat{\mathbf{s}}_t^- = \int f(\hat{\mathbf{s}}_{t-1}^+) \quad (9)$$

$$\dot{\mathbf{P}} = \mathbf{A}_t \mathbf{P} + \mathbf{P} \mathbf{A}_t^T + \mathbf{Q} \quad (10)$$

$$\mathbf{P}_t^- = \int \dot{\mathbf{P}}_{t-1}^+ \quad (11)$$

where \mathbf{Q} is the covariance matrix of noise in process model and \mathbf{A}_t is Jacobian matrix of state process f , define as:

$$\mathbf{A}_t = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} & \dots & \frac{\partial f_1}{\partial \theta_d} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial \theta_1} & \frac{\partial f_n}{\partial \theta_2} & \dots & \frac{\partial f_n}{\partial \theta_d} \end{bmatrix} \quad (12)$$

The Kalman gain is

$$\mathbf{L}_t = \mathbf{P}_t^- \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^- \mathbf{H}_t^T + \mathbf{R})^{-1}, \quad (13)$$

where \mathbf{R} is the covariance matrix of noise in measurement and \mathbf{H}_t is Jacobian matrix of the measurement model h . The posterior state and covariances are updated as:

$$\hat{\mathbf{s}}_t^+ = \hat{\mathbf{s}}_t^- + \mathbf{L}_t (\mathbf{y}_t - h(\hat{\mathbf{s}}_t^-)) \quad (14)$$

$$\mathbf{P}_t^+ = (\mathbf{I} - \mathbf{L}_t \mathbf{H}_t) \mathbf{P}_t^- (\mathbf{I} - \mathbf{L}_t \mathbf{H}_t)^T + \mathbf{L}_t \mathbf{R} \mathbf{L}_t^T \quad (15)$$

The EKF algorithm is summarized as Algorithm 1:

Algorithm 1 Extended Kalman filter

```

A. Initialize  $\hat{\mathbf{s}}_0^+$  and  $\mathbf{P}_0^+$ 
B. Updating and Correction
for  $t = 1, \dots, T$ 
  1. Compute  $\hat{\mathbf{s}}_t^- = \int_{t-1}^t f(\hat{\mathbf{s}}_{t-1}^+)$ 
  2. compute  $\mathbf{P}_t^-$  using Eqn.10,11
  3. Calculate the Kalman gain  $\mathbf{L}_t$  according to Eqn. 13
  4. Update the state posterior  $\hat{\mathbf{s}}_t^+$  by Eqn. 14
  5. Update the covariance posterior  $\mathbf{P}_t^+$  according to Eqn. 15
endfor

```

3.2 Unscented Kalman filter

The UKF carries out a different approximation, in that it consists of a recipe to deterministically define a minimum number of state samples in the prior distribution and propagate them through the dynamical system. These propagated points, referred to as sigma points, are used to re-compute a prior distribution and its covariance (Julier *et al.*, 1995) at the next time step. Thus, the UKF also propagates a Gaussian, but the approximation is formulated in a different way and is in fact equivalent to truncating the Taylor expansion of state non-linearity to the second-order term (Julier *et al.*, 2000). For more details on UKF implementations, including pseudocode, see Julier *et al.* (1995) and Quach *et al.* (2007).

3.3 Particle filter

Sequential Monte Carlo methods, more widely known as PFs, offer a more powerful approach to parameter estimation and inference in dynamical systems (Arulampalam *et al.*, 2002; de Freitas *et al.*, 2000; Doucet *et al.*, 2001). The family of Monte Carlo algorithms draw samples according to the probability density from which inference is to be made and approximate difficult integrals by sample averages. The core algorithmic step in particle filtering is importance sampling, that of generating identically and independently distributed samples from a proposal density of convenience, with associated weighting of these samples to account for the fact that they were not drawn from the density of interest. In the sequential setting, PF offers recursive ways of updating these weights in addition to propagating them through the dynamical system. A crucial issue in their implementation is the problem of sample degeneracy, that of all samples collapsing into just one position in the state space, and several tricks to circumvent these exist in the literature. A concise summary of these algorithms can be found in the tutorial paper by Arulampalam *et al.* (2002). A pseudocode description of the PF algorithm is given in Algorithm 2.

4 THE HEAT SHOCK MODEL

We consider cellular response to heat shock as a model system to illustrate the relative performance of the three sequential algorithms. El-Samad *et al.* (2006) describe a heat shock response model in the bacterium *Escherichia coli*, with three differential equations. The model consists of genes encoding molecular chaperones, transcribed in response to a sudden change of environmental temperature, transcription factor σ^{32} , which, via binding to RNA polymerase

Algorithm 2 Particle filter

```

1. Initialization,  $t = 0$ 
for  $i = 1, \dots, N$ 
    sample  $s_0^i \sim p(s_0)$  and set  $t = 1$ 
end for
2. Importance sampling
for  $i = 1, \dots, N$ 
    sample  $\tilde{s}_t^i \sim p(s_t | s_{t-1}^i)$  and set  $\tilde{s}_{0:t}^i = (s_{0:t}^i, \tilde{s}_t^i)$ 
end for
for  $i = 1, \dots, N$ 
    evaluate the importance weights  $\tilde{w}_t^i = p(y_t | \tilde{s}_t^i)$ 
end for
Normalize the importance weights
3. Selection step
Resample with replacement  $N$  particles ( $s_{0:t}^i; i = 1, \dots, N$ ) from the set
( $\tilde{s}_{0:t}^i; i = 1, \dots, N$ ) according to the importance weights
Set  $t \rightarrow t + 1$  and go to step 2.

```

enables the transcriptional regulation of heat shock proteins. σ^{32} , which is rarely present under normal temperatures in the range (30°C–37°C), rapidly accumulates at elevated temperatures activating the transcription of heat shock genes, leading to a different equilibrium.

The variation in total numbers of the relevant proteins are lumped in the model as two terms whose dynamics is described, along with the numbers of unfolded proteins as follows:

$$\begin{aligned}
 \dot{D}_t &= K_d \frac{S_t}{1 + \frac{K_u D_t}{1 + K_u U_f}} - \alpha_d D_t \\
 \dot{S}_t &= \eta(t) - \alpha_0 S_t - \alpha_s \frac{\frac{K_u D_t}{1 + K_u U_f}}{1 + \frac{K_u D_t}{1 + K_u U_f}} S_t \\
 \dot{U}_f &= K(t)[P_t - U_f] - [K(t) + K_{\text{fold}}]D_t
 \end{aligned} \quad (16)$$

Here, D_t represents the number of molecules of chaperones, S_t , the number of molecules of σ^{32} and U_f describes the total number of unfolded proteins. P_t and K_{fold} are the total number of proteins in the cell and a coefficient for the folding process, respectively. $K(t)$ and $\eta(t)$ are constants assuming different values at different steady state temperatures.

We simulated data from the heat shock model, making use of MATLAB's ODE45 function to integrate the differential equations, generating data over a time length of 200 s, sampled at regular intervals of 0.2 s, giving 1000 sample points for representing the observations. The true values of parameters used are given in Supplementary Table S1. We then applied the three filters to infer the unknown parameter and state. Subsets of the six parameters of the model are to be estimated from noisy subsets of the three states. In the evaluations we considered various combinations of knowns and unknowns (observed and estimated), as described in the following sections.

The simplest scenario we considered is one in which two of the three states (D_t and U_f) and five of the six parameters were assumed known. This corresponds to a case in of a well understood biological, some aspects of which are experimentally observable, subject to additive measurement noise, whereas others are inaccessible.

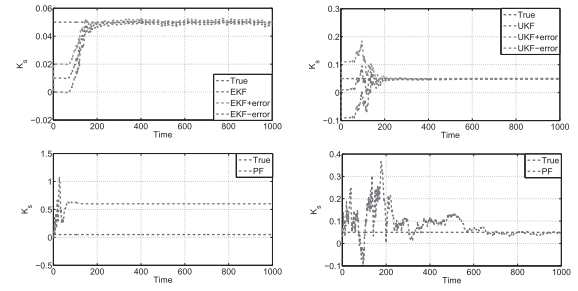


Fig. 1. Estimations of single unknown parameter (K_s) of the heat shock model using three different recursive filters. Top row: EKF and UKF; Bottom row: PFs with 50 and 1000 samples, respectively. For the EKF and UKF confidence levels from posterior variances are also shown.

Implementation details for this scenario, leading to the results in Figure 1 are as follows: of the EKF is set to start from the prior state vector $s = [0, 0, 0, 0.01]$, the last element is the initialization for parameter estimate. The diagonal elements in covariance matrix for measurement noise R are 0.01 times variance of synthetic state data x . The diagonal elements of initial error covariance matrix P_0 are 25, 25, 25 and 1×10^{-2} , respectively. The process noise covariance matrix was set to a small diagonal matrix, 5×10^{-6} .

The UKF is also initialized to start from the prior state vector $s = [0, 0, 0, 0.01]$, the last element being the initial guess of the unknown. The diagonal elements in covariance matrix for measurement noise R are 0.25 times variance of synthetic state data x . The first three diagonal elements (means the error covariance for states) of initial error covariance matrix P_0 are the same, leaving as 0.25 times variance of synthetic state data x . The last diagonal element in P_0 is 0.35, meaning this is for the parameter estimate.

For the PF, the initial particles for states and parameter are generated from $s_0 \sim \mathcal{N}(0, 1)$. The diagonal elements in covariance matrix for measurement noise R were set as 0.001 times the variance of the synthesized time series. With additive Gaussian noise, the importance weight are updated as $\tilde{w}_t^i = p(y_t | \tilde{s}_t^i) \propto \mathcal{N}(y_t - Hs_t, \sigma^2 I)$ where, for example, H , the observation matrix, whose diagonal elements will take the form $[1 \ 0 \ 1]$, if the state S_t is unobserved. The discount factor [Equation (6)] was set to $\delta = 0.98$ and we ran PFs with 50 and 1000 samples. Further, we also investigated a range of operating regimes with data lengths of 80 s to 180 s in steps of 20 s, sampling intervals of 0.1 s, 0.25 s, 0.5 s and 1 s, and noise variance multipliers of 0.0001, 0.001, 0.01, 0.05, 0.1 and 1.

5 RESULTS AND DISCUSSION

We carried out a thorough exploration of the capabilities of the three estimation methods on the heat shock system, at various levels of problem complexity, i.e. the numbers of parameters and states assumed unknown. Typical results obtained are described in the following sections, and results of all the combinations we tested are given as Supplementary Material accompanying this article.

5.1 Estimating a single unknown

Figure 1 shows convergence of the three models on estimating a single parameter (K_s) from two noisy state observations. In these, we have included a simulation of PF with only 50 samples

failing to correctly estimate the parameter. With 1000 samples, PF achieves convergence to the correct estimate. The EKF and UKF also converge to the correct underlying value, as seen in Figure 1, but note that the starting points of these are set to be very close to the true value. When the initial conditions are set to be far away from the true value, however, the EKF and UKF fail to converge whereas the PF is able to find the correct solution. This is shown in Figure 2 where the initial conditions of the EKF and UKF are set to be the closest to the truth among all the particles used as initial samples of the PF. Hence even in the case of a comparison highly favourable to the EKF and UKF, they fail to recover the correct value of the unknown. We have observed similar behaviour for each of the unknown parameters of the heat shock model.

Depending on the form of the interactions, some parameters are easier to identify than the others. Should the conditional posterior probabilities over them be unimodal, and close to Gaussian, parametric models (Kalman filters) have a good chance of finding the correct solutions. When the terms appear in complex non-linear settings, posterior probabilities are often multi-modal and non-parametric PFs are necessary. In Supplementary Table S2, we show in which of the cases of estimating a single unknown parameter from one and two noisy observations the PF clearly outperforms the other two. We explore all the possible combinations and find in 20 of the 42 cases, PF has an advantage over EKF and UKF, whereas in the remaining 22 it performs no worse than them. The difficulty of parameter estimation depends on the complexity of the formulation for the hidden state assumed to be unobserved.

We also consider the case in which all the parameters are known, and a single state was to be inferred from the other two. Results of this are shown in Supplementary Figure S1, and show broadly similar results in the rates of convergence. Other variations of estimating one of the six parameters from one and two observed states are shown in Supplementary Figures S3–S8, and the summary in Supplementary Table S2 is obtained by studying these graphs. In all these simulations, similar to the results in Figure 2, we set the initial guess of the unknown parameter at a value not too close to the true value, and gave the Kalman algorithms an unfair advantage

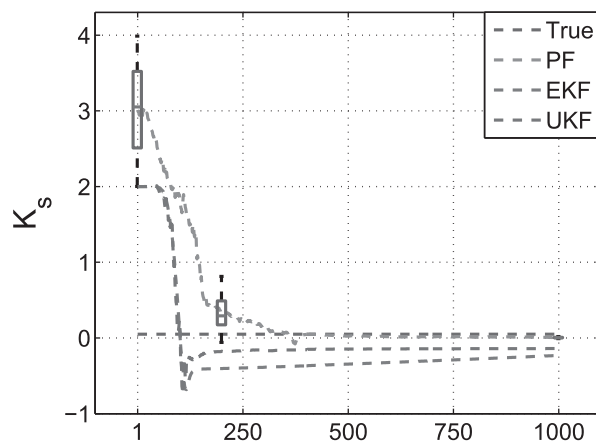


Fig. 2. Estimations of a single parameter K_s from two observed states with different initializations. The boxes show the distribution of samples of the PF at initial, two hundredth and final points in time.

by initializing *all* particles of the PF further away from the initial values to which EKF and UKF were set.

5.2 Estimating multiple parameters

We next examined the performances in multiple unknown parameter spaces, progressively making the problem harder. Results of convergence in estimating two parameters simultaneously, with all possible combinations of which parameters were left unknown, are shown in Supplementary Figures S9–S17. We note the general trend of PF outperforming the other two. Naturally, for the initial conditions chosen, there are cases when all three methods fail to estimate the parameters correctly. This is to be expected because the problem has now been made significantly harder than the case of estimating a single parameter. In exploring different operating regimes, we found the influence due to data lengths to be negligible (Supplementary Figs S19–S21). For the EKF and PF, faster sampling was found to be sometimes advantageous (Supplementary Figs S22–S24).

Figure 3C–E shows the space of initial conditions from which convergence to the true solution of the three algorithms were compared. We find that the PF is more resilient than either EKF or UKF in converging to the true solution. Also see Supplementary Figure S28, which shows the convergence regimes for two other sampling intervals.

For completeness, we also considered the extreme case of all parameters being unknown. Supplementary Figure S18 shows convergence of the three filters in the case of estimating all six parameters from two noisy observations (D_t and U_f). Here, we find that the EKF correctly estimates the parameter α_0 and UKF is able to successfully converge to α_0 and K_s , whereas the PF is able to correctly estimate five of the six unknowns. The PF fails only in the case of the parameter K_d , which is probably because K_d highly influences the temporal evolution of the system via the state S_t , which was assumed unknown in this particular simulation.

5.3 The sequential approach

The Kalman and PF algorithms considered in this work are sequential algorithms. Their use should be considered in the context of Bayesian inference methods that operate on batch data [e.g. Vysheirsky and Girolami (2008); Wilkinson, 2009], as in an example like the heat shock model, all the data are available. In such cases, sequential models, being one-pass algorithms offer a computational advantage. This is illustrated in Figure 3A, where we compare a Metropolis–Hastings sampler with PF. At similar levels of performance, we observed clear computational advantages with the sequential approach. The error bars in Figure 3A are obtained over 20 different runs of the algorithms.

An alternate way of parameter estimation of such models, also considered by El-Samad *et al.* (2006), is via an optimization approach. We compared such an approach with the Bayesian sequential approach, implementing the optimization via MATLAB's `fminsearch` program. This can be viewed as a maximum likelihood estimation, under assumptions of Gaussian noise. We found that with Gaussian noise, the optimization approach often outperformed the sequential Bayesian methods. However, with non-Gaussian noise, the PF gave substantial advantage. We illustrate this in Figure 3B, where in a two parameter estimation task, the

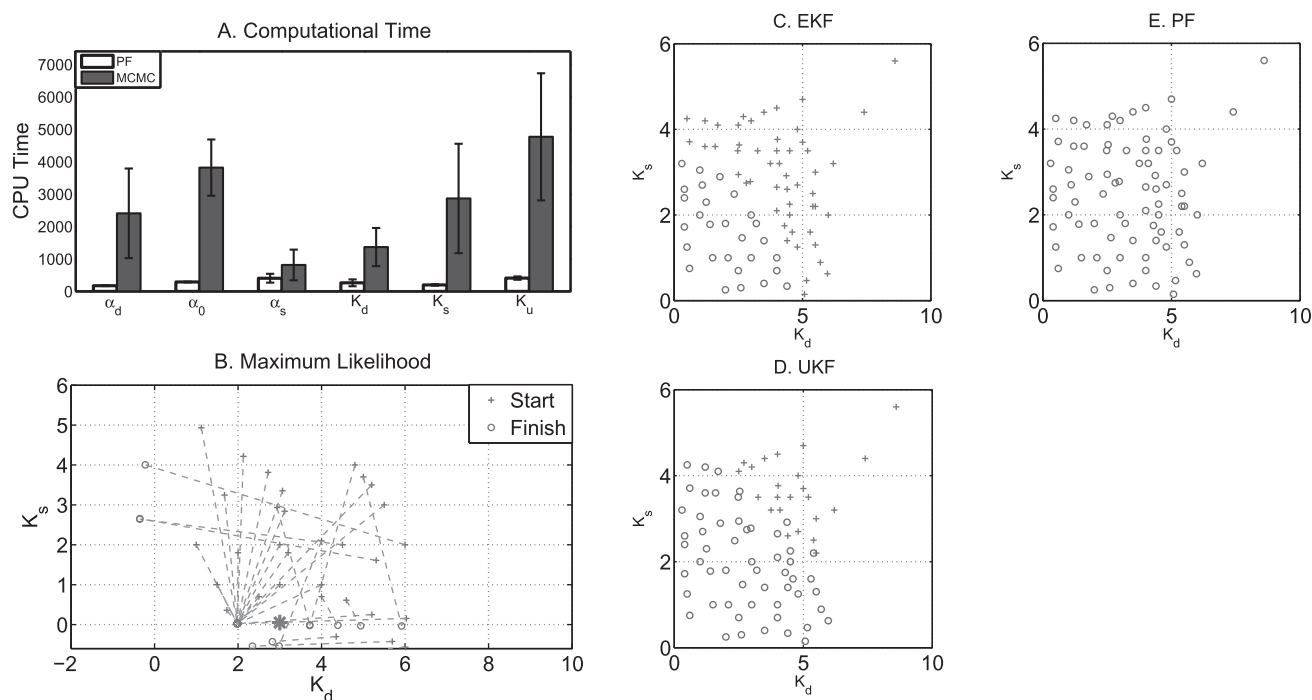


Fig. 3. Computation and performance of sequential models: (A) comparison of computational costs between a batch method and PFs in estimating single unknown parameters; (B) performance of a deterministic optimization approach. From various initial conditions, when simultaneously estimating two unknown parameters with mixture Gaussian observation noise, the optimization method fails to converge (true values denoted by ‘*’) whereas the PF was able to find the correct solution in the posterior mean in about half of these (trajectories not shown); and (C–E) a comparison of initial conditions from which convergence is reached for the three different sequential filters.

optimization approach repeatedly failed to find the correct solution, whereas the PF was successful in over half of the trials.

6 CONCLUSION

This work demonstrates the effectiveness of the method of particle filtering in state and parameter estimation of deterministic systems biological models from noisy observations. We have shown this via a comparative study between extended Kalman, unscented Kalman and particle filtering applied to the heat shock response system using single and multiple parameter estimations. While previous authors have argued that the Kalman filter itself is capable of such estimations, our critical appraisal shows that this is only possible when the initial guess of the state and/or state parameters are very close to the true values. Convergence is not achieved when the initial conditions differ significantly from the corresponding true values. The PF, on the other hand, is able to converge to true values even when all the particles are initially set to values far away from the underlying true values, providing a powerful, yet simple to implement, way of tackling difficult inference problems in systems biology. We further showed in this work that when the complexity of the problem is gradually increased (i.e. the number unknown parameters/states to be inferred), the Kalman filter algorithms failed well before the PF did. Even in the extreme case of all parameters being unknown, the PF manages to find correct estimates of five of the six. This suggests that the non-parametric approach of the PF, by virtue of being able to systematically propagate uncertainties while exploring the space

over a wide range, is a powerful methodology to tackle such difficult problems. Our current work concentrates on stretching such analyses to more difficult higher dimensional problems such as cell-cycle regulation, circadian rhythm and sino-atrial pace making in heart tissues.

Conflict of Interest: none declared.

REFERENCES

- Arulampalam, M.S. *et al.* (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, **50**, 174–188.
- Ashyraliyev, M. *et al.* (2008) Parameter estimation and determinability analysis applied to *Drosophila* gap gene circuits. *BMC Syst. Biol.*, **2**, 83.
- Barenco, M. *et al.* (2006) Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biol.*, **7**, R25.
- Bar-Shalom, Y. *et al.* (2001) *Estimation, Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons, New York.
- Chen, K. *et al.* (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell*, **11**, 369–391.
- Colin, S. (1982). *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors*. Springer, New York.
- Dewar, M. *et al.* (2010) Parameter estimation and inference for stochastic reaction-diffusion systems: application to morphogenesis in *D. melanogaster*. *BMC Syst. Biol.*, **4**, 21.
- de Freitas, J.F.G. *et al.* (2000) Sequential Monte Carlo methods to train neural network models. *Neural Comput.*, **12**, 955–993.
- Doucet, A. *et al.* (2001) *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- El-Samad, H. *et al.* (2006) Advanced methods and algorithms for biological networks analysis. *Proc. IEEE*, **94**, 832–853.

- Golightly,A. and Wilkinson,D. (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, **61**, 781–788.
- Hofbauer,J. and Sigmund,K. (1988) *Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection*. Cambridge University Press, New York.
- Jayawardhana,B. *et al.* (2008) Bayesian inference of the sites of perturbations in metabolic pathways via Markov chain Monte Carlo. *Bioinformatics*, **24**, 1191–1197.
- Julier,S.J. *et al.* (1995) A new approach for filtering nonlinear systems. *Proc. Am. Contr. Conf.*, **D**, 1628–1632.
- Julier,S. *et al.* (2000) A new method for nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Contr.*, **45**, 477–482.
- Kadirkamanathan,V. and Niranjan,M. (1993) A function estimation approach to sequential learning with neural networks. *Neural Comput.*, **5**, 954–975.
- Kalman,R.E. (1960) A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, **82**, 35–45.
- Kitagawa,G. (1998) A self-organizing state-space model. *J. Am. Stat. Assoc.*, **93**, 1203–1215.
- Kolch,W. (2000) Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interaction. *Biochem. J.*, **351**, 289–305.
- Lawrence,N.D. *et al.* (2006) Modelling transcriptional regulation using Gaussian processes. In Schölkopf,B. *et al.* (eds) *Advances in Neural Information Processing Systems 19*. MIT press, Vancouver, pp. 785.
- Lillacci,G. and Khammash,M. (2010) Parameter estimation and model selection in computational biology. *PLoS Comput. Biol.*, **6**, 696–713.
- Liu,J. and West,M. (2001) Combined parameter and state estimation in simulation-based filtering. In Doucet,A. *et al.* (eds) *Sequential Monte Carlo Methods in Practice*. Springer, New York, pp. 197–217.
- Matsuno,H. *et al.* (2005) A new regulatory interactions suggested by simulations for circadian genetic control mechanism in mammals. *J. Bioinform. Comput. Biol.*, **4**, 139–153.
- Mendes,P. and Kell,D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, **14**, 869–883.
- Nakamura,K. *et al.* (2009) Parameter estimation of *in silico* biological pathways with particle filtering towards a petascale computing. *Pac. Symp. Biocomput.*, **14**, 227–238.
- Niranjan,M. (1997) Sequential tracking in pricing financial options using model based and neural network approaches. In Mozer,M. *et al.* (eds) *Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA, pp. 960–966.
- Quach,M. *et al.* (2007) Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. *Bioinformatics*, **23**, 3209–3216.
- Simon,D. (2006) *Optimal State Estimation*. Wiley, New York.
- Sitz,A. *et al.* (2002) Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Phys. Rev. E*, **66**, 016210.
- Sun,X. *et al.* (2008) Extended Kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks. *PLoS ONE*, **3**, e3758.
- Timmer,J. *et al.* (2002) Modeling the nonlinear dynamics of cellular signal transduction. *Int. J. Bifurcat. Chaos*, **14**, 2069–2079.
- Vyshemirsky,V. and Girolami,M.A. (2008) BioBayes: a software package for Bayesian inference in systems biology. *Bioinformatics*, **24**, 1933–1934.
- Wilkinson,D.J. (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.*, **10**, 122–133.
- Yang,J. *et al.* (2007) *In vivo* intracellular metabolite dynamics estimation by sequential Monte Carlo filter. In *IEEE Symposium on Computational Intelligence, Bioinformatics and Computational Biology, 2007 (CIBCB07)*. Hawaii, USA, pp. 387–394.
- Zhan,C. and Yeung,L. (2011) Parameter estimation in systems biology models using spline approximation. *BMC Syst. Biol.*, **5**, 14.