IET Systems Biology

Special Issue based on the 8th International Conference on Systems Biology and the 4th Translational Bioinformatics Conference (ISB/TBC2014)

IET Journals
The Institution of Engineering and Technology

Model reduction and parameter estimation of non-linear dynamical biochemical reaction networks

ISSN 1751-8849
Received on 5th March 2015
Revised on 15th May 2015
Accepted on 25th May 2015
doi: 10.1049/iet-syb.2015.0034
www.ietdl.org

Xiaodian Sun ✉, *Mario Medvedovic*

Department of Environmental Health, Laboratory for Statistical Genomics and Systems Biology, University of Cincinnati College of Medicine, 3223 Eden Ave. ML56, Cincinnati, OH 45267-0056, USA
✉ *E-mail: sunxd@uc.edu*

Abstract: Parameter estimation for high dimension complex dynamic system is a hot topic. However, the current statistical model and inference approach is known as a large *p* small *n* problem. How to reduce the dimension of the dynamic model and improve the accuracy of estimation is more important. To address this question, the authors take some known parameters and structure of system as priori knowledge and incorporate it into dynamic model. At the same time, they decompose the whole dynamic model into subset network modules, based on different modules, and then they apply different estimation approaches. This technique is called Rao-Blackwellised particle filters decomposition methods. To evaluate the performance of this method, the authors apply it to synthetic data generated from repressilator model and experimental data of the JAK-STAT pathway, but this method can be easily extended to large-scale cases.

1 Introduction

Biochemical reaction process is a complex and high-dimension dynamics system, which includes a variety of feedback loop [1, 2] and possesses strongly non-linear kinetic characteristics such as chaos, bifurcation, complex disturbance wave and so on [3–7]. The non-linear complex and high-dimension biochemical reaction can be decomposed into several sets of chemical substances, and then combined to consider the whole mechanism. On the basis of biochemical reaction, dynamic modelling is accordingly divided into many subset network modules. Both subset network modules and the whole dynamic characteristics should be analysed [8, 9]. This is defined as model reduction techniques. Model reduction techniques decompose a critical biochemical reactions and variables according to core dynamical characteristics of the system. There are two kinds of techniques mostly used to partition the state variables. One is fast and slow decompositions, another is linear and non-linear decompositions. The former detailed model reduction approaches have singular perturbation techniques in papers [10, 11], hierarchical approach in paper [12], quasi-steady-states approximations in papers [13, 14], partial-equilibriums in paper [15] and kernel-based manifold learning techniques in paper [16]. The latter includes quasi-steady and quasi-equilibrium in paper [9], hierarchy of coarse grained model in paper [17], distribution state estimation in paper [18] and Rao-Blackwellised particle filters (RBPFs) in paper [19]. In our work, we focus on linear and non-linear decompositions by using RBPFs.

In the past, the dynamic model of the non-linear biochemical reaction is generally based on black-box framework to estimate the parameters and identify the structures of system. Since there exists large p small n problem (number of unknown parameters p is of much larger than sample size n, $p \gg n$) in parameter estimation fields, we take some known parameters and structure of system as priori knowledge and incorporate it into dynamic model. In other words, we estimate the parameters and states in non-linear biochemical reaction network based on grey-box framework [20]. As we know, most of biochemical reaction networks are non-linear and non-Gaussian, however, in which, linear subsystem are still available. Pseudo-monomolecular or monomolecular reaction is the simplest reaction described by a set of first-order reactions. In papers [17, 21, 22], pseudo-monomolecular or monomolecular reaction subsystems are considered as linear subset network modules, based on which, linear and non-linear decompositions are easy to be realised. For linear kinetic models, appearing as pseudo-monomolecular or monomolecular reaction subsystems, we propose to estimate the parameters using conventional Kalman filter algorithm. However, for the remaining non-linear kinetic models, we develop an algorithm to estimate both states and parameters using the particle filter algorithm. It is known as RBPFs [19, 23–27]. Extended Kalman filter (EKF) [28] and unscented Kalman filter (UKF) [29] are the most widely used joint state and parameter estimation algorithm for the non-linear state-space model of biochemical network. In this paper, we will compare the three estimation methods that are RBPF, UKF and EKF in synthetic data generated from repressilator model and experimental data from the JAK-STAT pathway. The results show that RBPF provides a way to handle high-dimensional problems and bears very good accuracy with quite reasonable complexity.

2 Non-linear state-space models

Consider a general non-linear dynamic system

$$\begin{aligned} x_{k+1}^e &= f_k(\boldsymbol{x}_k^e, \theta_k, \boldsymbol{u}_k^e) + w_k \\ \boldsymbol{y}_k^e &= h_k(\boldsymbol{x}_k^e, \theta_k, \boldsymbol{u}_k^e) + v_k, \quad e = 1, \ldots, E; \quad k = 1, \ldots, T_e \end{aligned} \tag{1}$$

where e is the individual and k is the time; $\boldsymbol{x}_k^e$ is the state vector of the e individual at a k time; $\boldsymbol{u}_k^e$ is the input vector of the e individual at a k time; $\boldsymbol{y}_k^e$ is the observation vector of the e individual at a k time; f and h are non-linear functions, $\boldsymbol{\theta}$ is the vector of parameters; The initial state $\boldsymbol{x}_0$ is a Gaussian vector with mean $E[x_0] = \bar{x}_0$ and covariance matrix $\boldsymbol{E}\left[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^{\mathrm{T}}\right] = \sum_0$; $\boldsymbol{w}$ and $\boldsymbol{v}$ are vectors of white noises with zero mean and joint covariance matrix

$$\boldsymbol{E}\left[\begin{pmatrix} w_k \\ v_k \end{pmatrix}\begin{pmatrix} w_k^{\mathrm{T}} & v_k^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}\right] = \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix}$$

Parameters in a non-linear dynamic system (1) can be treated as

additional states in the system. Thus, the state vector $\boldsymbol{x}_k^e$ in (1) is augmented as

$$\xi_k^e = \begin{bmatrix} \boldsymbol{x}_k^e \\ \boldsymbol{\theta} \end{bmatrix}$$

A state-space equation treating parameters as states can be written as

$$\begin{aligned} \xi_{k+1}^e &= \begin{bmatrix} x_{k+1}^e \\ \theta_k \end{bmatrix} = \begin{bmatrix} f_k(\boldsymbol{x}_k^e, \theta_k, \boldsymbol{u}_k^e) \\ \theta_k \end{bmatrix} + \begin{bmatrix} w_k \\ e_k \end{bmatrix} = g_k(\xi_k^e, \boldsymbol{u}_k^e) + \eta_k \\ \boldsymbol{y}_k^e &= h_k(\boldsymbol{x}_k^e, \theta_k, \boldsymbol{u}_k^e) + v_k == h_k(\xi_k^e, \boldsymbol{u}_k^e) + v_k \\ \theta_{k+1} &\sim p(\theta_{k+1}|\theta_k) \end{aligned} \tag{2}$$

where $\eta_k = \left[w_k^{\mathrm{T}}, e_k^{\mathrm{T}}\right]^{\mathrm{T}}$ is a white Gaussian noise with mean zero and covariance matrix

$$\boldsymbol{E}\left[\eta_k \eta_k^{\mathrm{T}}\right] = \begin{bmatrix} Q_k & 0 \\ 0 & \Gamma_k \end{bmatrix} = \varphi_k$$

Supposing a system is divided into two parts: a linear and a non-linear and the noise is additive, then (2) can be expressed as follows

$$\begin{aligned} x_{k+1}^{e\mathrm{n}} &= f_k^{\mathrm{n}}(x_k^{e\mathrm{n}}, \theta_k, \boldsymbol{u}_k^e) + A_k^{\mathrm{n}} x_k^{\mathrm{l}} + B_k^{\mathrm{n}} \boldsymbol{u}_k^e + w_k^{\mathrm{n}} \\ x_{k+1}^{e\mathrm{l}} &= f_k^{\mathrm{l}}(x_k^{e\mathrm{n}}, \theta_k, \boldsymbol{u}_k^e) + A_k^{\mathrm{l}} x_k^{e\mathrm{l}} + B_k^{\mathrm{l}} u_k^e + w_k^{\mathrm{l}} \\ \boldsymbol{y}_k^e &= h_k(x_k^{e\mathrm{n}}, \theta_k, \boldsymbol{u}_k^e) + C x_t^{e\mathrm{l}} + D u_k^e + e_k \\ \theta_{k+1} &\sim p(\theta_{k+1}|\theta_k) \end{aligned} \tag{3}$$

where $x_k^{e\mathrm{n}}$ and $x_k^{e\mathrm{l}}$ denote the non-linear and linear states, respectively, and $\left[x_k^{e\mathrm{n}} \quad x_k^{e\mathrm{l}}\right]^{\mathrm{T}} = \boldsymbol{x}_k^e$, ω_k is the process noise given by

$$w_k = \begin{bmatrix} w_k^{\mathrm{n}} \\ w_k^{\mathrm{l}} \end{bmatrix} \sim N\left(0, \begin{pmatrix} Q_k^{\mathrm{n}} & Q^{\mathrm{ln}}{}_k \\ Q^{\mathrm{ln}\,\mathrm{T}}{}_k & Q_k^{\mathrm{l}} \end{pmatrix}\right), \quad v_k \sim N(0, R_k)$$

where $N(0, \sigma^2)$ denotes the normal distribution with 0 as the mean value and σ^2 the variance. Moreover, $x_0^{e\mathrm{l}} \sim N(\hat{x}_{0|-1}^{e\mathrm{l}}, P_{0|-1}^{\mathrm{l}})$ and v_k, $x_0^{e\mathrm{n}}$ have arbitrary fixed probability density function (pdf).

Assume

$$p(\theta_{k+1}|\theta_k) = N(a\theta_k + (1-a)*\bar{\theta}_k, h^2 \boldsymbol{V}_k)$$

where $a = ((3\delta - 1)/2\delta)$, $h^2 = 1 - a^2$, δ is a discount factor (0, 1], typically around $0.95 \sim 0.99$. $\bar{\theta}_k$ is the Monte Carlo mean of the parameters and $\boldsymbol{V}_k$ being the variance matrix of the parameters at time instant k.

We determine the unknown parameter $\boldsymbol{\theta}$ by estimating the augmented state ξ_k^e with $p(\xi_k^e|Y_k^e)$; The minimum mean-squared errors estimation of ξ_k^e given $Y_k^e = \{y_1^e, y_2^e, \ldots, \boldsymbol{y}_k^e\}$ is

$$\begin{aligned} \hat{\xi}_k^e &= E[\xi_k^e|Y_k^e] = \int \xi_k^e p(\xi_k^e|Y_k^e)\,\mathrm{d}\xi_k^e = \iint (\boldsymbol{x}_k^e, \theta_k) p(\boldsymbol{x}_k^e, \theta_k|Y_k^e)\,\mathrm{d}\boldsymbol{x}_k^e\,\mathrm{d}\theta_k \\ &= \iint (\boldsymbol{x}_k^e, \theta_k) p(\boldsymbol{x}_k^e|\theta_k, Y_k^e) p(\theta_k|Y_k^e)\,\mathrm{d}\boldsymbol{x}_k^e\,\mathrm{d}\theta_k \\ &= \int \left[\int (\boldsymbol{x}_k^e, \theta_k) p(\boldsymbol{x}_k^e|\theta_k, Y_k^e)\,\mathrm{d}\boldsymbol{x}_k^e\right] p(\theta_k|Y_k^e)\,\mathrm{d}\theta_k \end{aligned}$$

$p(\theta_k|Y_k^e)$ is approximated by particle filter, for each given parameter sample, $p(\boldsymbol{x}_k^e|\theta_k, Y_k^e)$ is given by Kalman filter. This will result in each parameter particle being associated with one Kalman filter recursion.

3 RBPF algorithm for dual estimation

RBPF algorithms for dual estimation is summarised in Fig. 1.

The following two applications of the algorithm (see Fig. 1) are from paper [28]; As the length of paper is limited, there is no explanation about the system equation. For detailed explanation, please see paper [28]. As far as signalling pathways, the topological structure of signalling pathways is known. In such case, the most important work is to estimate the parameters of non-linear models based on Hill or mass action kinetics. The topological structure of network is used as prior knowledge and incorporated into the kinetic models and repressilator model.

4 Kinetic models for JAK2-STAT5 signalling pathway

System equation [28, 30]

$$\begin{aligned} \dot{x}_1(t) &= -k_1 x_1(t) u(t) + 2k_4 x_3(t) I(t \geq \tau) + w_1(t) \\ \dot{x}_2(t) &= k_1 x_1(t) u(t) - k_2 x_2^2(t) + w_2(t) \\ \dot{x}_3(t) &= -k_3 x_3(t) + 0.5 k_2 x_2^2(t) + w_3(t) \\ \dot{x}_4(t) &= k_3 x_3(t) - k_4 x_3(t) I(t \geq \tau) + w_4(t) \end{aligned}$$

$$I(t \geq \tau) = \begin{cases} 1 & t \geq \tau \\ 0 & t < \tau \end{cases}$$

Measurement equation

$$\begin{aligned} y_1(t) &= x_2(t) + 2x_3(t) + e_1(t) \\ y_2(t) &= x_1(t) + x_2(t) + 2x_3(t) + e_2(t) \end{aligned}$$

Model reduction

For decomposing the system and dividing the state variables into linear state variable and non-linear state variable, let

$$x_t^{\mathrm{n}} = x_2(t) \quad x_t^{\mathrm{l}} = \begin{bmatrix} x_1(t) \\ x_3(t) \\ x_4(t) \end{bmatrix}$$

where x_t^{l} denotes the state variable with conditional linear dynamics and x_t^{n} denotes the non-linear state variable. The system equation can be rewritten as the following (see equation at the bottom of the next page)

where

$$f_t^{\mathrm{n}} = (x_t^{\mathrm{n}}(t) - k_2 (x_t^{\mathrm{n}}(t))^2), \quad A_t^{\mathrm{n}} = \begin{bmatrix} k_1 u(t) & 0 & 0 \end{bmatrix}, \quad w_t^{\mathrm{n}} = w_2(t)$$

$$f_t^{\mathrm{l}} = \begin{bmatrix} 0 \\ 0.5 k_2 (x_t^{\mathrm{n}}(t))^2 \\ 0 \end{bmatrix},$$

$$A_t^{\mathrm{l}} = \begin{bmatrix} 1 - k_1 u(t) & 2k_4 I(t \geq \tau) & 0 \\ 0 & 1 - k_3 & 0 \\ 0 & k_3 - k_4 I(t \geq \tau) & 1 \end{bmatrix}, \quad w_t^{\mathrm{l}} = \begin{bmatrix} w_1(t) \\ w_3(t) \\ w_4(t) \end{bmatrix}$$

The measurement equation can be rewritten as the following

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_t^{\mathrm{n}}(t) + \begin{bmatrix} 0 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix} x_t^{\mathrm{l}}(t) + e$$

where

$$ht = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x_t^{\mathrm{n}}(t), \quad C = \begin{bmatrix} 0 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}, \quad e = \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix}$$

Rao-Blackwellised Particle Filter (RBPF) Algorithm for Dual Estimation

Rao-Blackwellised Particle Filter (RBPF) algorithms for dual estimation is summarised in the following:

For every individual $e = 1 \cdots E$, repeats the following steps:

1. Initial parameter pdf $p(\theta_0 \mid y_0^e)$ as a uniform distribution over $[a,b]$ and choose the number of particles N. Initial state estimate to be $\hat{x}_0^l$ and initial state covariance matrix to be p_0, Initial the particles $\hat{x}_0^{en}$
2. For $i = 1,2,\cdots,N$, draw initial parameter sample $\theta_{1|0}(i) \sim p(\theta_0 \mid y_0^e)$ and initialise the Kaman filter associated with each parameter particle as $x_{0|0}^{el}(i) = \hat{x}_0^{el}, P_{0|0}(i) = p_0$, Initialise each particles $x_{0|0}^{en}(i) = \hat{x}_0^{en}$
3. For every k $(k = 1,\cdots,T_e)$, repeat the following steps:

(3.1) For every particle $i \in \{1,2,\cdots,N\}$

(3.1.1)Draw $\theta_{k|k-1}(i) \sim p(\theta_k \mid \theta_{k-1}(i))$

$$x_{k|k-1}^{en}(i) \sim p(x_k^{en} \mid x_{k-1}^{en}, Y_k^e) = N(f_k^n + A_k^n \hat{x}_{k-1|k-1}^{el}, A_k^n P_{k-1|k-1}(A_k^n)^T + G_k^n Q_k^n (G_k^n)^T)$$

(3.1.2) From $\theta_{k|k-1}(i)$,obtain the corresponding $A_k^l(\theta_{k-1|k-1}(i))$, $B_k^l(\theta_{k|k-1}(i))$,

$$A_k^n(\theta_{k-1|k-1}(i)), B_k^n(\theta_{k|k-1}(i)), C(\theta_{k|k-1}(i)), D(\theta_{k|k-1}(i))$$

Compute the following:

(a) Kalman Filter time update:

$$\begin{aligned}
z_k^e &= x_{k+1}^{en}(i) - f_k^n \\
\overline{A}_k^l(\theta_{k-1|k-1}(i)) &= A_k^l(\theta_{k-1|k-1}(i)) - G_k^l(Q_k^{ln})^T (G_k^n Q_k^n)^{-1} A_k^n(\theta_{k-1|k-1}(i)) \\
\overline{Q}_k^l &= Q_k^l - (Q_k^{ln})^T (Q_k^n)^{-1} Q_k^{ln} \\
N_k &= A_k^n(\theta_{k-1|k-1}(i)) P_{k|k}(i) (A_k^n(\theta_{k-1|k-1}(i)))^T + G_k^n Q_k^n (G_k^n)^T \\
L_k &= \overline{A}_k^l(\theta_{k-1|k-1}(i)) P_{k|k}(i) (A_k^n(\theta_{k-1|k-1}(i)))^T N_k^{-1} \\
\hat{x}_{k+1|k}^{el}(i) &= \overline{A}_k^l(\theta_{k-1|k-1}(i)) \hat{x}_{k|k}^{el}(i) + G_k^l(Q_k^{ln})^T (G_k^n Q_k^n)^{-1} z_k + f_k^l + L_k(z_k^e - A_k^n \hat{x}_{k|k}^{el}(i)) \\
P_{k|k-1}(i) &= \overline{A}_k^l(\theta_{k-1|k-1}(i)) P_{k-1|k-1}(i) (\overline{A}_k^l(\theta_{k-1|k-1}(i)))^T + G_k^l \overline{Q}_k^l (G_k^l)^T - L_k N_k L_k^T
\end{aligned}$$

(b)Kalman Filter measurement update:

Fig. 1 *RBPF algorithm for dual estimation*

$$x_t^{\mathrm{n}}(t+1) = (x_t^{\mathrm{n}}(t) - k_2(x_t^{\mathrm{n}}(t))^2) + \begin{bmatrix} k_1 u(t) & 0 & 0 \end{bmatrix} x_t^{\mathrm{l}} + w_t^{\mathrm{n}}$$

$$x_t^{\mathrm{l}}(t+1) = \begin{bmatrix} 0 \\ 0.5k_2(x_t^{\mathrm{n}}(t))^2 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 - k_1 u(t) & 2k_4 I(t \geq \tau) & 0 \\ 0 & 1 - k_3 & 0 \\ 0 & k_3 - k_4 I(t \geq \tau) & 1 \end{bmatrix} + x_t^{\mathrm{l}}(t) + w_t^{\mathrm{l}}$$

IET Syst. Biol., 2016, Vol. 10, Iss. 1, pp. 10–16

$$\hat{z}^{e}_{k|k-1}(i) = C(\theta_{k|k-1}(i))\hat{x}^{e}_{k|k-1}(i) + D(\theta_{k|k-1}(i))u^{e}_{k-1}(i)$$
$$R_k(i) = C(\theta_{k|k-1}(i))P_{k|k-1}(i)C^T(\theta_{k|k-1}(i)) + Q_v$$
$$K_k(i) = P_{k|k-1}(i)C^T(\theta_{k|k}(i)R_k^{-1}(i)$$

$$\begin{aligned}\hat{x}^{el}_{k|k} &= \hat{x}^{el}_{k|k-1}(i) + K_k(i)(z^{e}_{k} - h_k(x^{en}_k) - \hat{z}^{e}_{k|k-1}(i)) \\ &= \hat{x}^{el}_{k|k-1}(i) + K_k(i)(z^{e}_{k} - h_k(x^{en}_k) - C(\theta_{k|k-1}(i))\,\hat{x}^{el}_{k|k-1} - D(\theta_{k|k-1}(i))u^{e}_{k-1}(i))\end{aligned}$$
$$P_{k|k}(i) = P_{k|k-1}(i) - K_k(i)R_k(i)K_k^T(i)$$

(3.1.3) $\tilde{\alpha}_k(i) = p(y^{i}_{k} \mid X^{en}_k, Y^{e}_{k-1}, \theta_{k|k-1}(i)) \sim N((h_k(x^{en}_k) + \hat{z}^{e}_{k|k-1}(i)), R_k(i))$

(3.2) Compute the normalizing factor $\sum_{j=1}^{N}\tilde{\alpha}_k(j)$

(3.3) Normalize the weights, $\alpha_k(i) = \dfrac{\tilde{\alpha}_k(i)}{\sum_{j=1}^{N}\tilde{\alpha}_k(j)}$, $i = 1, \cdots, N$

(3.4) Using the weights, resample the parameter particles

$$P_r(\theta_{k|k}(i) = \theta_{k|k-1}(j)) = \alpha_k(j)$$

Complete resample algorithm is summarized in the following:

Generate a uniformly distributed random point $\lambda_1 \in [0, N^{-1}]$

And let $i = 1, \alpha_k(0) = 0$

For $j = 1 : N$

let $\lambda_j = \lambda_1 + N^{-1}(j-1)$

If $((\lambda_j \geq \sum_{l=0}^{i-1}\alpha_k(l)) \quad or \quad (\lambda_j > \sum_{l=0}^{i}\alpha_k(l)))$

$$i = i + 1$$

Else if $(\sum_{l=0}^{i-1}\alpha_k(l) < \lambda_j \leq \sum_{l=0}^{i}\alpha_k(l)))$

Set $\theta_k(j) = \theta_{k|k-1}(i)$

Endif

(3.5) Obtain the parameter estimates: $\hat{\theta}_k = \sum_{i=1}^{N}\alpha_k(i)\theta_{k|k}(i)$

Fig. 1 *Continued*

5 Repressilator

System and measurement equation [31]

$$\frac{dR_1}{dt} = \frac{V_{1\max}K_{12}^n}{K_{12}^n + P_2^n} - k_{1m}R_1 + w_1$$
$$\frac{dR_2}{dt} = \frac{V_{2\max}K_{23}^n}{K_{23}^n + P_3^n} - k_{2m}R_2 + w_2$$
$$\frac{dR_3}{dt} = \frac{V_{3\max}K_{31}^n}{K_{31}^n + P_1^n} - k_{3m}R_3 + w_3$$
$$\frac{dP_1}{dt} = k_1R_1 - k_{1p}P_1 + w_4$$
$$\frac{dP_2}{dt} = k_2R_2 - k_{2p}P_2 + w_5$$
$$\frac{dP_3}{dt} = k_3R_3 - k_{3p}P_3 + w_6$$

$$y_1(t) = R_1(t) + e_1$$
$$y_2(t) = R_2(t) + e_2$$
$$y_3(t) = R_3(t) + e_3$$
$$y_4(t) = P_1(t) + e_4$$
$$y_5(t) = P_2(t) + e_5$$
$$y_6(t) = P_3(t) + e_6$$

where R_i is the concentration of mRNA transcript from gene i and $P_i$is the concentration of proteins translated from R_i. Estimated parameters: $V_{1\max}$, $V_{2\max}$, $V_{3\max}$, k_{12}, k_{23}, k_{31}.

Model reduction

For decomposing the system and dividing the state variables into linear state variable and non-linear state variable, let

$$x_t^n(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{bmatrix}, \quad x_t^l(t) = \begin{bmatrix} R_1(t) \\ R_2(t) \\ R_3(t) \end{bmatrix}$$

where x_t^l denotes the state variable with conditional linear dynamics and x_t^n denotes the non-linear state variable. Then, the system equation can be rewritten as the following

$$x_t^l(t+1) = f_t^l(x_t^n) = \begin{bmatrix} \frac{V_{1\max}k_{12}^3}{k_{12}^3 + p_2^3(t)} \\ \frac{V_{2\max}k_{23}^3}{k_{23}^3 + p_3^3(t)} \\ \frac{V_{3\max}k_{31}^3}{k_{31}^3 + p_1^3(t)} \end{bmatrix} + w_t^l,$$
$$x_t^n(t+1) = A_t^n x_t^l = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} x_t^l + w_t^n$$

where

$$f_t^l(x_t^n) = \begin{bmatrix} \frac{V_{1\max}k_{12}^3}{k_{12}^3 + p_2^3(t)} \\ \frac{V_{2\max}k_{23}^3}{k_{23}^3 + p_3^3(t)} \\ \frac{V_{3\max}k_{31}^3}{k_{31}^3 + p_1^3(t)} \end{bmatrix}, \quad A_t^l = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_t^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$f_t^n(x_t^n) = 0, \quad w_t^n = \begin{bmatrix} w_4 \\ w_5 \\ w_6 \end{bmatrix}, \quad w_t^l = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

The measurement equation can be rewritten as the following

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \\ y_5(t) \\ y_6(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x_t^n + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x_t^l + e$$

where

$$h_t(x_t^n) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

6 Results

6.1 JAK-STAT pathway

The initial values of the linear state variables of every particle are assumed as: $x_1^0 = 0.5$, $x_2^0 = 0$, $x_3^0 = 0$ and $x_4^0 = 0$. The initial values of the non-linear state variables of every particle are random number between 0 and 1. The initial values of parameters k_1, k_2, k_3 and k_4 are random number generated from the following intervals $k_1 \in (-0.299,\ 0.421)$, $k_2 \in (2.16,\ 2.76)$, $k_3 \in (-0.2534,\ 0.3466)$ and $k_4 \in (-0.14342,\ 0.30658)$. The estimates by the RBPF, the EKF and maximum likelihood (ML) method [32] are close, but significantly different from the estimates by UKF [29] as shown in Table 1. Under the given above initial values, using the concentration of $EpoR_A$ as input, Fig. 2 plot the predicted and observed concentrations of tyrosine phosphorylated STAT5 in the cytoplasm and total STAT5 in the cytoplasm (y_1 and y_2) by the RBPF method, the EKF method and the UKF method. The observed data were from experiment data 1. The results of experiment data 2–4 and the estimated parameters were listed in additional files AFigures 1–3 and ATable 1.

6.2 Synthetic data generated from repressilator model

Let $k_{1m} = 1$, $k_{2m} = 1$, $k_{3m} = 1$, $k_{1p} = 1$, $k_{2p} = 1$, $k_{3p} = 1$, $K_1 = 1$, $K_2 = 2$, $K_3 = 3$, $n = 3$ and $V_{1\max} = 150$, $V_{2\max} = 80$, $V_{3\max} = 100$, $K_{12} = 50$, $K_{23} = 40$, $K_{31} = 60$, the initial values of linear and non-linear state variables are random numbers between 0 and 1. The initial values of parameters are random number generated from the following intervals: $V_{1\max} \in (140,\ 160)$, $V_{2\max} \in (70,\ 90)$, $V_{3\max} \in (90,\ 110)$, $K_{12} \in (40,\ 60)$, $K_{23} \in (20,\ 40)$, $K_{31} \in (30,\ 50)$. The estimated parameters as a function of time k are shown in Figs. 3 and 4. From Figs. 3 and 4, we can see that at the beginning the estimated parameters quickly converge to the true parameters. This example demonstrates that although the parameters are treated as the states of the systems and hence may change over time, they can reach stable values. The estimated parameters over time k are summarised in ATable 2 in additional files, which demonstrated

Table 1 Comparison of estimated parameters in the non-linear state-space model for the JAK-STAT pathway using RBPF, EKF, ML approach and UKF

Study	k_1	k_2	k_3	k_4	τ
RBPF (our study)	0.022916	2.343347	0.117178	0.102687	6.1
EKF	0.0211	2.2788	0.1064	0.1057	6 min
ML	0.0210	2.4600	0.1066	0.1066	6.4 min
UKF	0.0515		3.3900	0.3500	

IET Syst. Biol., 2016, Vol. 10, Iss. 1, pp. 10–16

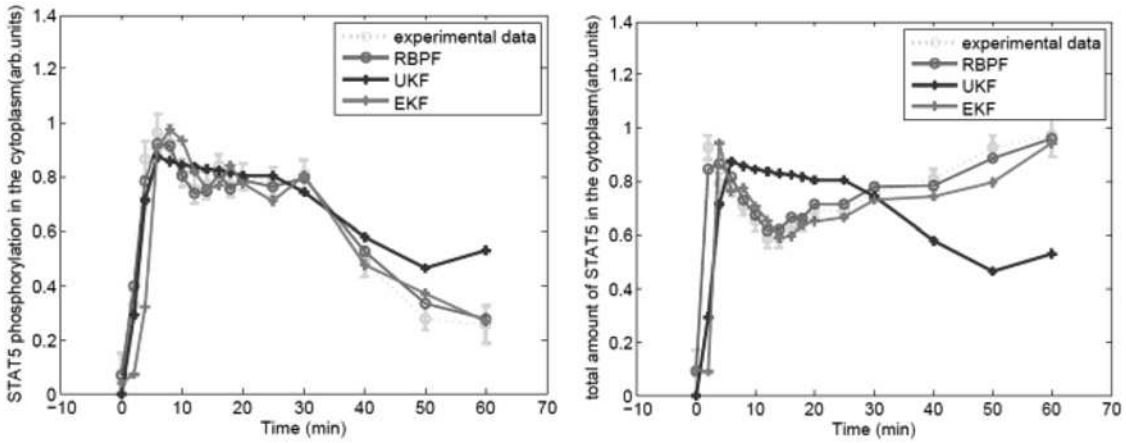

Fig. 2 *Predicted and observed concentrations of tyrosine phosphorylated STAT5 and total STAT5 in the cytoplasm for experiment 1*

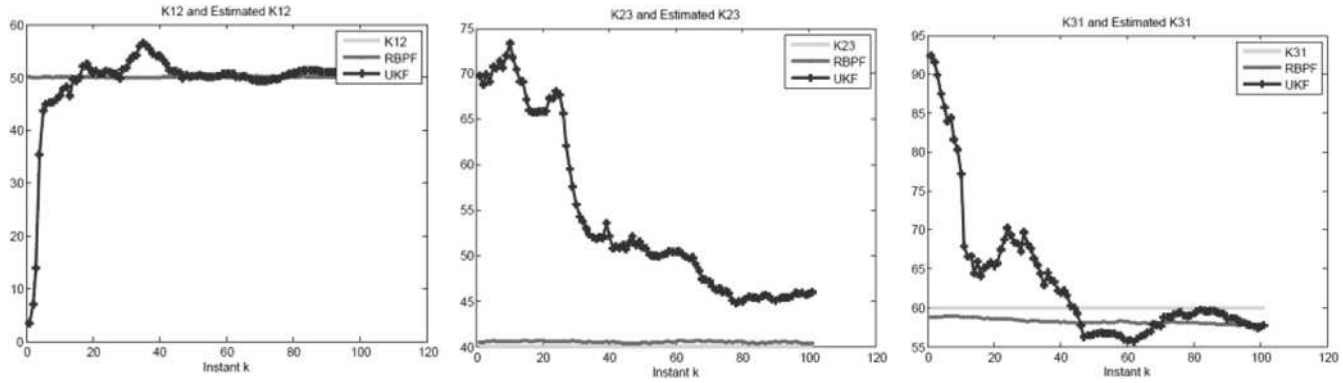

Fig. 3 *Estimated parameters of K_{12}, K_{23} and K_{31} as a function of the time in the non-linear state-space model for the synthetic data generated from the repressilator model*

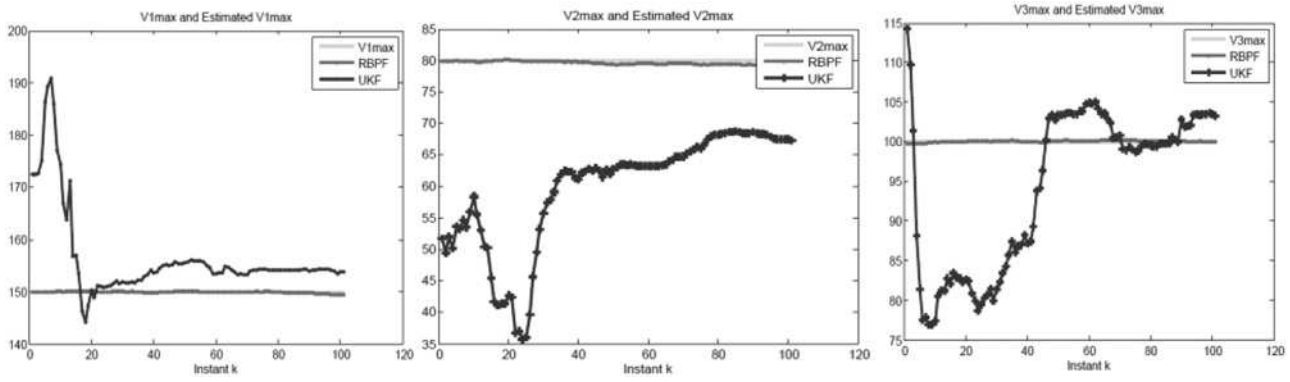

Fig. 4 *Estimated parameters of V_{1max}, V_{2max} and V_{3max} as a function of the time in the non-linear state-space model for the synthetic data generated from the repressilator model*

that the estimates of the parameters were very close to set the value of parameters. In this example, EKF does not converge, high non-linearity of the repressilator model makes EKF a failure to converge to an optimum. Therefore, we only compare the two methods of RBPF and UKF.

7 Conclusions

To evaluate the performance of our new methods, we have applied it to both synthetic data generated from repressilator model and experimental data of the JAK-STAT pathway [31, 32]. The structure of both the above examples is known in modelling literature [29, 33]. Therefore, we use structure information and partly known parameters as priori knowledge and then conduct the identification of biochemical reaction networks based on grey box [34–37]. We consider the pseudo-monomolecular or monomolecular reaction subsystems as linear subset network modules, then the whole dynamic model are decomposed into linear and non-linear subset modules dynamic model. For linear subset modules dynamic model, we use Kalman filter algorithm to estimate both states and parameters, however, for non-linear subset

modules dynamic model, we adopt the particle filter algorithm to estimate both states and parameters. This model reduction technique is called RBPF. The results show that RBPF method perform well as a new model reduction techniques for high dimension non-linear dynamic model. As future work, we will apply our algorithms to a high-dimensional biochemical network in order to improve and validate it.

8 Acknowledgments

Xiaodian Sun and Mario Medvedovic are funded by the NIH Data Coordination and Integration Center for LINCS-BD2 K grant U54HG008230.

9 References

1 Sobieszczanski-Sobieski, J.: 'Sensitivity of complex, internally coupled systems', *AIAA J.*, 1990, **28**, (1), pp. 153–160
2 Hurty, W.C.: 'Dynamic analysis of structural systems using component modes', *AIAA J.*, 1965, **3**, (4), pp. 678–685
3 Roesky, P.W., Doumbouya, S.I., Schneider, F.W.: 'Chaos induced by delayed feedback', *J. Phys. Chem.*, 1993, **97**, (2), pp. 398–402
4 Nitzan, A., Ortoleva, P., Deutch, J., Ross, J.: 'Fluctuations and transitions at chemical instabilities: the analogy to phase transitions', *J. Chem. Phys.*, 1974, **61**, (3), pp. 1056–1074
5 Matheson, I., Walls, D., Gardiner, C.: 'Stochastic models of firstorder nonequilibrium phase transitions in chemical reactions', *J. Stat. Phys.*, 1975, **12**, (1), pp. 21–34
6 Lee, K.J., McCormick, W., Ouyang, Q., Swinney, H.L.: 'Pattern formation by interacting chemical fronts', *Science*, 1993, **261**, (5118), pp. 192–194
7 McNeil, K., Walls, D.: 'Nonequilibrium phase transitions in chemical reactions', *J. Stat. Phys.*, 1974, **10**, (6), pp. 439–448
8 Matthews, M.L., Williams, C.: 'Region of attraction estimation of biological continuous Boolean models', in (Eds.): 'Book region of attraction estimation of biological continuous Boolean models' (IEEE, 2012), pp. 1700–1705
9 Radulescu, O., Gorban, A.N., Zinovyev, A., Noel, V.: 'Reduction of dynamical biochemical reaction networks in computational biology', arXiv preprint arXiv:1205.2851, 2012
10 Prescott, T.P., Papachristodoulou, A.: 'Layered decomposition for the model order reduction of timescale separated biochemical reaction networks', *J. Theor. Biol.*, 2014, **356**, pp. 113–122
11 Kourdis, P.D., Palasantza, A.G., Goussis, D.A.: 'Algorithmic asymptotic analysis of the NF-κB signaling system', *Comput. Math. Appl.*, 2013, **65**, (10), pp. 1516–1534
12 Radulescu, O., Gorban, A.N., Vakulenko, S., Zinovyev, A.: 'Hierarchies and modules in complex biological systems'. Proc. ECCS'06, 2006
13 Schneider, K.R., Wilhelm, T.: 'Model reduction by extended quasi-steady-state approximation', *J. Math. Biol.*, 2000, **40**, (5), pp. 443–450
14 Segel, L.A., Slemrod, M.: 'The quasi-steady-state assumption: a case study in perturbation', *SIAM Rev.*, 1989, **31**, (3), pp. 446–477
15 Gallagher, P.M., Athayde, A.L., Ivory, C.F.: 'The combined flux technique for diffusion – reaction problems in partial equilibrium: application to the facilitated transport of carbon dioxide in aqueous bicarbonate solutions', *Chem. Eng. Sci.*, 1986, **41**, (3), pp. 567–578
16 Dsilva, C.J., Talmon, R., Gear, C.W., Coifman, R.R., Kevrekidis, I.G.: 'Data-driven reduction for multiscale stochastic dynamical systems', arXiv preprint arXiv:1501.05195, 2015
17 Radulescu, O., Gorban, A.N., Zinovyev, A., Lilienbaum, A.: 'Robust simplifications of multiscale biochemical networks', *BMC Syst. Biol.*, 2008, **2**, (1), p. 86
18 Guo, Y., Wu, W., Zhang, B., Sun, H.: 'A distributed state estimation method for power systems incorporating linear and nonlinear models', *Int. J. Electr. Power Energy Syst.*, 2015, **64**, pp. 608–616
19 Schon, T., Gustafsson, F., Nordlund, P.-J.: 'Marginalized particle filters for mixed linear/nonlinear state-space models', *IEEE Trans. Signal Process.*, 2005, **53**, (7), pp. 2279–2289
20 Kristensen, N.R., Madsen, H., Jørgensen, S.B.: 'Parameter estimation in stochastic grey-box models', *Automatica*, 2004, **40**, (2), pp. 225–237
21 Gorban, A.N., Radulescu, O.: 'Dynamic and static limitation in multiscale reaction networks, revisited', *Adv. Chem. Eng.*, 2008, **34**, pp. 103–173
22 Gorban, A., Radulescu, O., Zinovyev, A.Y.: 'Asymptotology of chemical reaction networks', *Chem. Eng. Sci.*, 2010, **65**, (7), pp. 2310–2324
23 Schön, T., Gustafsson, F.: 'Particle filters for system identification of state-space models linear in either parameters or states', 2003
24 Li, P., Goodall, R., Kadirkamanathan, V.: 'Parameter estimation of railway vehicle dynamic model using Rao-Blackwellised particle filter', in (Eds.): 'Book parameter estimation of railway vehicle dynamic model using Rao-Blackwellised particle filter' (2003)
25 Daly, M.J., Reilly, J.P., Morelande, M.R.: 'Rao-Blackwellised particle filtering for blind system identification', in (Eds.): 'Book Rao-Blackwellised particle filtering for blind system identification' (IEEE, 2005), vol. 324, pp. iv/321–iv/324
26 Karlsson, R., Schön, T., Gustafsson, F.: 'Complexity analysis of the marginalized particle filter', 2004
27 Schön, T., Karlsson, R., Gustafsson, F.: 'The marginalized particle filter in practice', 2005
28 Sun, X., Jin, L., Xiong, M.: 'Extended Kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks', *PloS One*, 2008, **3**, (11), p. e3758
29 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*, 2007, **23**, (23), pp. 3209–3216
30 Kisseleva, T., Bhattacharya, S., Braunstein, J., Schindler, C.: 'Signaling through the JAK/STAT pathway, recent advances and future challenges', *Gene*, 2002, **285**, (1), pp. 1–24
31 Elowitz, M.B., Leibler, S.: 'A synthetic oscillatory network of transcriptional regulators', *Nature*, 2000, **403**, (6767), pp. 335–338
32 Swameye, I., Müller, T., Timmer, J., Sandra, O., Klingmüller, U.: 'Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling', *Proc. Natl. Acad. Sci.*, 2003, **100**, (3), pp. 1028–1033
33 Klipp, E., Liebermeister, W.: 'Mathematical modeling of intracellular signaling pathways', *BMC Neurosci.*, 2006, **7**, (Suppl 1), p. S10
34 Holst, J., Holst, U., Madsen, H., Melgaard, H.: 'Validation of grey box models', in (Eds.): 'Book validation of grey box models' (Elsevier, 2014), p. 53
35 Liu, Z.-P.: 'Reverse engineering of genome-wide gene regulatory networks from gene expression data', *Curr. Genomics*, 2015, **16**, (1), pp. 3–22
36 Liu, Z.-P., Wu, H., Zhu, J., Miao, H.: 'Systematic identification of transcriptional and post-transcriptional regulations in human respiratory epithelial cells during influenza A virus infection', *BMC Bioinf.*, 2014, **15**, (1), p. 336
37 Liu, Z.-P., Zhang, W., Horimoto, K., Chen, L.: 'Gaussian graphical model for identifying significantly responsive regulatory networks from time course high-throughput data', *IET Syst. Biol.*, 2013, **7**, (5), pp. 143–152