An Extended Kalman Filtering Approach to Modeling Nonlinear Dynamic Gene Regulatory Networks via Short Gene Expression Time Series

Zidong Wang, Xiaohui Liu, Yurong Liu, Jinling Liang, and Veronica Vinciotti

Abstract—In this paper, the extended Kalman filter (EKF) algorithm is applied to model the gene regulatory network from gene time series data. The gene regulatory network is considered as a nonlinear dynamic stochastic model that consists of the gene measurement equation and the gene regulation equation. After specifying the model structure, we apply the EKF algorithm for identifying both the model parameters and the actual value of gene expression levels. It is shown that the EKF algorithm is an online estimation algorithm that can identify a large number of parameters (including parameters of nonlinear functions) through iterative procedure by using a small number of observations. Four real-world gene expression data sets are employed to demonstrate the effectiveness of the EKF algorithm, and the obtained models are evaluated from the viewpoint of bioinformatics.

Index Terms—Modeling, clustering, DNA microarray technology, extended Kalman filtering, gene expression, time series data.

1 Introduction

The DNA microarray technology has made it possible to conduct simultaneous expression measurements from tens of thousands of genes, and a global view on the expression levels of all genes is enabled when the cell undergoes specific conditions or processes. Measuring gene expression levels in different conditions may prove useful in medical diagnosis, treatment, and drug design. Many gene expression experiments produce *short* time series data with only a few time points due to the high measurement costs. The time series usually represents the dynamic response of an organism to a change in conditions, e.g., application of some drug or other treatment. Therefore, it is highly desired to extract the functional information from gene expression time series data, and the modeling of gene expression time series has become an area attracting increasing research attention.

Among many methods for modeling gene expression data, the clustering approach has gained a particular research focus. For example, cluster analysis of the gene expression data was studied as early as in [12] and, since then, many clustering algorithms have been applied to gene expression data, such as hierarchical clustering [12], self-organizing map [38], *k*-means [39], and Gaussian model-based clustering [33], [46]; see [23] for an overview. It has now been recognized that the cluster analysis is based on the assumption that there exists the correlation similarity between genes, which is essentially a static approach. Recently, dynamic modeling of gene regulatory networks from time series data has received more and more research interest [10], [35]. A number of dynamical models have been put forward with examples including Boolean network model [1], [21], [27], [36], linear differential equation model [5], [9], [11], [20], [43], Bayesian model [19], [24], [28], [31], state-space model [3], [34], [44], and stochastic model [7], [41], [42].

Since it is well known that the gene expression is an inherently stochastic phenomenon [25], [29], [31], [40], the network should be of a "stochastic" nature. Recently, stochastic modeling of gene expression time series has been paid a great deal of research attention. For example, dynamic Bayesian networks (DBNs) have been proposed to model gene expression time series data [24], [28], [31] because DBNs can model stochasticity and handle noisy/hidden variables. However, DBNs need more complex algorithms such as the genetic algorithm [24], [37] to infer gene regulatory networks. The state-space model [3], [34], [44] assumes that the gene expression value depends not only on the current internal state variables but also on the external inputs, which reflects the nature of a dynamic network. Unfortunately, most results reported on state-space models have been focused on linear systems, and therefore, the nonlinear phenomenon of the gene networks cannot be taken into account.

In addition to the stochastic behavior, there are still two typical features that contribute to the complexity of a gene regulatory network: 1) the network is essentially nonlinear

- *Z. Wang, X. Liu, and V. Vinciotti are with the Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex UB8 3PH, UK. E-mail: {zidong.wang, xiaohui.liu, veronica.vinciotti}@brunel.ac.uk.*
- *Y. Liu is with the Department of Mathematics, Yangzhou University, Yangzhou 225002, P.R. China. E-mail: liuyurong@gmail.com.*
- *J. Liang is with the Department of Mathematics, Southeast University, Nanjing 210096, P.R. China. E-mail: jinlliang@gmail.com.*

Manuscript received 3 Sept. 2008; revised 6 Jan. 2009; accepted 8 Jan. 2009; published online 12 Jan. 2009.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2008-09-0158. Digital Object Identifier no. 10.1109/TCBB.2009.5.

(for example, the activation function regulating the gene activity profile is highly nonlinear) and 2) the available gene expression time series usually consists of a large number of variables but with a small number of observations. However, most available literature concerning modeling gene expression time series has not explicitly dealt with these two features, and therefore, there is a need to seek alternative approaches to identifying parameters of a nonlinear stochastic gene regulatory network through real-time gene expression time series. In search of such an approach, the extended Kalman filtering (EKF) approach appears to be an appropriate candidate.

The traditional Kalman filter addresses the general problem of trying to estimate the state of a discrete-time system governed by a linear stochastic difference equation, see, e.g., [13], [14], [15], [16], [17], [18]. EKF linearizes about the current mean and covariance, and can therefore handle nonlinearities that can be associated either with the process model or with the observation model or with both. On the other hand, EKF is known as an effective online (recursive) estimator of process variables, which can be suitable for identifying large number of parameters using a short time series [8]. However, despite its potential in modeling gene regulatory networks (GRNs), so far, the EKF approach with applications to identify *nonlinear dynamic GRNs via short gene expression time series* has received little research attention. Note that the unscented Kalman filtering (UKF) has been utilized in [32] to the estimation of both parameters and hidden variables of nonlinear state-space models, but the UKF approach would require a sufficiently long time series for the *statistical* inference. Hence, it is our intention in this paper to fill the gap by investigating the use of EKF in the presence of nonlinearities, short time series, and high dimension of the variables.

In this paper, the gene regulatory network is considered as a nonlinear dynamic stochastic model that consists of the gene measurement equation and the gene regulation equation. In order to reflect the reality, we consider the gene measurement from microarray as noisy, and assume that the gene regulation equation is a nonlinar autoregressive stochastic dynamic process, where the nonlinearity stems from the inherently nonlinear regulatory relationship and degree among genes. After specifying the model structure, we apply the EKF algorithm for identifying both the model parameters and the actual value of gene expression levels. Note that the EKF algorithm is an online estimation algorithm that can identify a large number of parameters (including parameters of nonlinear functions) through iterative procedure by using a small number of observations. Four real-world gene expression data sets are employed to demonstrate the effectiveness of the EKF algorithm, and the obtained models are evaluated from the viewpoint of bioinformatics.

2 Gene Model and Problem Formulation

The measured gene expression levels can be modeled as

$$y_i(k) = x_i(k) + v_i(k), \quad i = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, m, \quad \text{(I)}$$

where $y(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T$ is the measurement data from microarray experiments at time k with $y_i(k)$ describing the ith gene expression levels at time k, $x_i(k)$ is the ith actual gene expression levels which stand for mRNA concentrations and/or protein concentrations at time k, $v_i(k)$ is the measurement noise, n is the number of the genes, and m is the number of the measurement time points. Here, $v(k) = [v_1(k), v_2(k), \ldots, v_n(k)]^T$ is assumed to be a zero-mean Gaussian white noise sequence with constant covariance $R > 0$, i.e., $v(k) \sim \mathcal{N}(0, R)$.

The gene regulatory network containing n genes is described by the following discrete-time nonlinear stochastic dynamical system [6]:

$$x_i(k+1) = \sum_{j=1}^{n} a_{ij} x_j(k) + \sum_{j=1}^{n} b_{ij} f_j(x_j(k), \mu_j) + I_{0i} + \xi_i(k), \quad i = 1, 2, \ldots, n, \quad k = 0, 1, 2, \ldots, m-1, \quad \text{(II)}$$

where $A = (a_{ij})_{n\times n}$ is the linear regulatory relationship and degree among genes; $B = (b_{ij})_{n\times n}$ represents the nonlinear regulatory relationship and degree among genes; $I_0 = [I_{01}, I_{02}, \ldots, I_{0n}]^T$ is the constant vector with I_{0i} standing for the external bias on the ith gene; $\xi(k) = [\xi_1(k), \xi_2(k), \ldots, \xi_n(k)]^T \sim \mathcal{N}(0, Q_0)$; and the nonlinear function $f_j(x_j, \mu_j)$ is given by

$$f_j(x_j, \mu_j) = \frac{1}{1 + e^{-\mu_j x_j}},$$

with μ_j being a parameter to be identified.

Setting $\mu = [\mu_1, \mu_2, \ldots, \mu_n]^T$ and $f(x(k), \mu) = [f_1(x_1(k), \mu_1), f_2(x_2(k), \mu_2), \ldots, f_n(x_n(k), \mu_n)]^T$, we can rewrite (I) and (II) in the following vector form:

$$x(k+1) = Ax(k) + Bf(x(k), \mu) + I_0 + \xi(k), \quad (1)$$

$$y(k) = x(k) + v(k). \quad (2)$$

Letting

$$A_e = [a_{11}, a_{21}, \ldots, a_{n1}, a_{12}, a_{22}, \ldots, a_{n2}, \ldots, a_{1n}, a_{2n}, \ldots, a_{nn}]^T, \quad (3)$$

$$B_e = [b_{11}, b_{21}, \ldots, b_{n1}, b_{12}, b_{22}, \ldots, b_{n2}, \ldots, b_{1n}, b_{2n}, \ldots, b_{nn}]^T, \quad (4)$$

$$\mu = [\mu_1, \mu_2, \ldots, \mu_n]^T, \quad (5)$$

$$\theta = [A_e^T \quad B_e^T \quad \mu^T \quad I_0^T]^T, \quad (6)$$

all the parameters to be estimated are denoted by

$$\theta = [A_e^T \quad B_e^T \quad \mu^T \quad I_0^T]^T. \quad (7)$$

In order to establish the gene expression model (II), it is necessary to identify the parameter vector θ. In this paper, we aim at estimating the parameters of the model (II) via the EKF method from the measurement data.

Remark 1. In many papers, dealing with modeling problems of genetic regulatory networks, Hill functions are used as

the regulatory functions instead of the logistic function. It is worth mentioning that the developed EKF algorithm could also estimate the parameters of the Hill functions since these parameters can be arranged in a similar form to (5). On the other hand, gene regulatory networks are sometimes more appropriate to include delays when describing the characteristics of transcription and translation. In the case that the time delay is small, the state augmentation can be used to convert the model into a delay-free one and the EKF approach can be applied. For the large time delays, traditional EKF approach needs to be modified that constitutes one of the future research topics.

3 The Extended Kalman Filter Approach to Parameter Estimation

In this section, we introduce the Kalman filter, the EKF approach to parameter identification, respectively; see, e.g., [8], [45].

3.1 The Discrete Kalman Filter

The Kalman filter addresses the general problem of trying to estimate the state $x(k) \in \mathbb{R}^n$ of a discrete-time process that is governed by the linear stochastic difference equation

$$x(k+1) = G_{k+1,k}x(k) + H_{k+1,k}u(k) + \Gamma_{k+1,k}w(k), \quad (8)$$

with the measurement $z(k) \in \mathbb{R}^m$ given by

$$z(k) = C_k x(k) + D_k u(k) + v(k), \quad (9)$$

where the random variables $w(k)$ and $v(k)$ represent the process and measurement noise, respectively. They are assumed to be independent (of each other), white, and with normal probability distributions

$$E[w(k)] = E[v(k)] = 0, \quad E[w(k)w^T(j)] = Q_k\delta_{kj},$$
$$E[v(k)v^T(j)] = R_k\delta_{kj}, \quad E[v(k)w^T(j)] = 0,$$

where

$$\delta_{kj} = \begin{cases} 1, & k = j, \\ 0, & k \neq j. \end{cases}$$

We define $\hat{x}(k+1|k)$ to be our a priori state estimate at step $k+1$ given knowledge of the process prior to step $k+1$, and $\hat{x}(k+1|k+1) \in \mathbb{R}^n$ to be our a posteriori state estimate at step $k+1$ given measurement $z(k+1)$.

Discrete-time Kalman filter algorithm.

Initialization. For $k = 0$, set

$$\hat{x}(0|0) = E[x(0)] = x_0,$$
$$P(0|0) = E[(x(0) - x_0)(x(0) - x_0)^T] = P_{x_0}.$$

Step 1 (Time update). Given $x(k|k)$ and $P(k|k)$ apply *the time update* (effect of system dynamics)

$$\hat{x}(k+1|k) = G_{k+1,k}\hat{x}(k|k) + H_{k+1,k}u(k),$$
$$P(k+1|k) = G_{k+1,k}P(k|k)G_{k+1,k}^T + \Gamma_{k+1,k}Q_k\Gamma_{k+1,k}^T,$$

to obtain $\hat{x}(k+1|k), P(k+1|k)$.

Step 2. Then, after obtaining the new measurement $z(k+1)$, apply the following *measurement update* (effect of measurement):

$$K_{k+1} = P(k+1|k)C_{k+1}^T(C_{k+1}P(k+1|k)C_{k+1}^T + R_{k+1})^{-1},$$
$$P(k+1|k+1) = (I - K_{k+1}C_{k+1})P(k+1|k),$$
$$\hat{x}(k+1|k+1) = G_{k+1,k}\hat{x}(k|k) + H_{k+1,k}u(k) + K_{k+1}\big[Z(k+1) - D_{k+1}u(k+1) - C_{k+1}(G_{k+1,k}\hat{x}(k|k) + H_{k+1,k}u(k))\big]$$

to obtain the optimal estimates $P(k+1|k+1), \hat{x}(k+1|k+1)$. K_{k+1} is called the Kalman gain.

Set $k = k+1$ and go to step 1.

3.2 Extended Kalman Filtering

The Kalman filter is the optimum state estimator for a linear system with the assumptions as described. If the system is nonlinear, then we may use a linearization process at every time step to approximate the nonlinear system with a linear time varying (LTV) system. This LTV system is then used in the Kalman filter, resulting in an EKF on the true nonlinear system. Note that although EKF is not necessarily optimal, it often works very well.

Consider the following nonlinear system:

$$x(k+1) = f(x(k)) + w(k), \quad (10)$$

$$y(k) = g(x(k)) + v(k), \quad (11)$$

where k is a nonnegative integer, $x(k) \in \mathbb{R}^n$ is the system state vector, $y(k) \in \mathbb{R}^r$ is the observation vector, $w(k)$ and $v(k)$ are the system noise and the measurement noise, respectively, and $w(k)$ and $v(k)$ are the zero-mean white Gaussian stochastic processes with covariance matrices Q_k and R_k, respectively. Here, $f: \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear state transition function and $g: \mathbb{R}^n \to \mathbb{R}^r$ is a nonlinear measurement function.

The EKF is implemented by the following consecutive steps:

1. Consider the last filtered state estimate $\hat{x}(k|k)$.
2. Linearize the system dynamics (10) around $\hat{x}(k|k)$.
3. Apply the prediction step of the Kalman filter to the linearized system dynamics just obtained, yielding $\hat{x}(k+1|k)$ and $P(k+1|k)$.
4. Linearize the observation equation (11) around $\hat{x}(k|k)$.
5. Apply the filtering or update cycle of the Kalman filter to the linearized observation dynamics, yielding $\hat{x}(k+1|k+1)$ and $P(k+1|k+1)$.

Let

$$\hat{A}(k) = \left.\frac{\partial f(x(k))}{\partial x(k)}\right|_{x(k)=\hat{x}(k|k)}, \quad \hat{C}(k) = \left.\frac{\partial g(x(k))}{\partial x(k)}\right|_{x(k)=\hat{x}(k|k-1)}. \quad (12)$$

Assume that $x(0) \sim \mathcal{N}(x_0, P_{x_0})$, $w(k) \sim \mathcal{N}(0, Q_k)$, $v(k) \sim \mathcal{N}(0, R_k)$ with $R_k > 0$, and that $\{w(k)\}$ and $\{v(k)\}$ are the white noise processes uncorrelated with $x(0)$ and with each other. Then, the EKF algorithm can be stated as follows:

Initialization.

For $k = 0$, set

$$\hat{x}(0|0) = E[x(0)] = x_0,$$
$$P(0|0) = E[(x(0) - x_0)(x(0) - x_0)^T] = P_{x_0}.$$

For $k = 1, 2, 3, \ldots$ compute.

Time update ("Predict")

State estimate time update: $\hat{x}(k|k-1) = f(\hat{x}(k-1|k-1))$.

Error covariance time update: $P(k|k-1) = \hat{A}(k-1)P(k-1|k-1)\hat{A}(k-1)^T + Q_{k-1}$.

Measurement update ("Correct")

Compute the Kalman gain matrix: $K_k = P(k|k-1)\hat{C}^T(k)[\hat{C}(k)P(k|k-1)C(k)^T + R_k]^{-1}$.

Update the estimate with measurement $y(k)$: $\hat{x}(k|k) = \hat{x}(k|k-1) + K_k[y(k) - g(\hat{x}(k|k-1))]$.

Error covariance measurement update: $P(k|k) = (I - K_k\hat{C}(k))P(k|k-1)$.

Remark 2. The convergence of EKF has been a research topic of recurring interest for some decades, and a variety of criteria have been proposed to guarantee the convergence. Recently, it has been shown in [26] that if the system is C^2 and uniformly observable with bounded second partial derivatives, then the EKF converges locally for a broad class of nonlinear systems. In this case, if the initial estimation error of the filter is not too large, then the error goes to zero exponentially as time goes to infinity. More detailed discussions on the convergence can be found in [22], [26] and the references therein.

3.3 The EKF for Parameter Identification

EKF is a very practical method in identification of nonlinear systems. Augmenting the unknown parameters to the state vector makes it possible to use EKF for parameter identification, too. In general, the nonlinear system dynamics can be described by

$$x(k+1) = f(x(k), \theta) + \xi(k), \tag{13}$$

with a measurement given by

$$z(k) = g(x(k), \theta) + v(k), \tag{14}$$

where $x(k)$ and $z(k)$ are the state vector and the measurement vector, respectively, $\xi(k)$ and $v(k)$ are the zero-mean white noise processes, and the parameters to be estimated are denoted as θ. In many applications as well as in our system, it is natural to assume that the parameters are constant. In order to estimate a vector of parameters θ from the nonlinear state-space model, the state vector $x(k)$ is augmented to include the parameters as states

$$X(k) = \begin{bmatrix} x(k) \\ \theta(k) \end{bmatrix},$$

and the parameters are modeled as constants with uncertain initial conditions: $\theta \sim \mathcal{N}(\theta_0, P_\theta)$. The resulting dynamic equation is

$$X(k+1) = F(X(k)) + w(k), \tag{15}$$

$$z(k) = G(X(k)) + v(k), \tag{16}$$

where $w(k) = [\xi^T(k), 0]^T$, $F(X(k)) = [f^T(x(k), \theta(k)), \theta^T(k)]^T$, and $G(X(k)) = g(x(k), \theta(k))$. Then, the standard EKF algorithm can be applied.

4 Applications to Real-World Gene Expression Time Series Data

In this section, we focus our attention on the model (1) and show how to estimate its parameters via the EKF approach.

For presentation convenience, we denote

$$A(k) = (a_{ij}(k))_{n\times n}, \tag{17}$$

$$B(k) = (b_{ij}(k))_{n\times n}, \tag{18}$$

$$\begin{aligned} A_e(k) = [&a_{11}(k), a_{21}(k), \ldots, a_{n1}(k), a_{12}(k), a_{22}(k), \ldots, \\ &a_{n2}(k), \ldots, a_{1n}(k), a_{2n}(k), \ldots, a_{nn}(k)]^T, \end{aligned} \tag{19}$$

$$\begin{aligned} B_e(k) = [&b_{11}(k), b_{21}(k), \ldots, b_{n1}(k), b_{12}(k), b_{22}(k), \ldots, \\ &b_{n2}(k), \ldots, b_{1n}(k), b_{2n}(k), \ldots, b_{nn}(k)]^T, \end{aligned} \tag{20}$$

$$\mu(k) = [\mu_1(k), \mu_2(k), \ldots, \mu_n(k)]^T, \tag{21}$$

$$I_0(k) = [I_{01}(k), I_{02}(k), \ldots, I_{0n}(k)]^T, \tag{22}$$

$$\theta(k) = [A_e^T(k) \quad B_e^T(k) \quad \mu^T(k) \quad I_0^T(k)]^T, \tag{23}$$

$$X(k) = [x^T(k) \quad \theta^T(k)]^T. \tag{24}$$

Let I_n be the $n \times n$ identity matrix, and the matrix $H = [I_m \ 0]$ with appropriate dimension.. In order to facilitate the application of the EKF in the parameter estimation problem, we rewrite (1) as follows:

$$X(k+1) = F(X(k)) + w(k), \tag{25}$$

$$y(k) = HX(k) + v(k), \tag{26}$$

where

$$\begin{aligned} F(X(k)) &= \begin{bmatrix} A(k)x(k) + B(k)f(x(k), \mu(k)) + I_0(k) \\ \theta(k) \end{bmatrix}, \\ w(k) &= \begin{bmatrix} \xi(k) \\ 0 \end{bmatrix}, \end{aligned} \tag{27}$$

with $f(x, \mu) = [f_1(x_1, \mu_1), f_2(x_2, \mu_2), \ldots, f_n(x_n, \mu_n)]^T$ and

$$f_j(x_j, \mu_j) = \frac{1}{1 + e^{-\mu_j x_j}}.$$

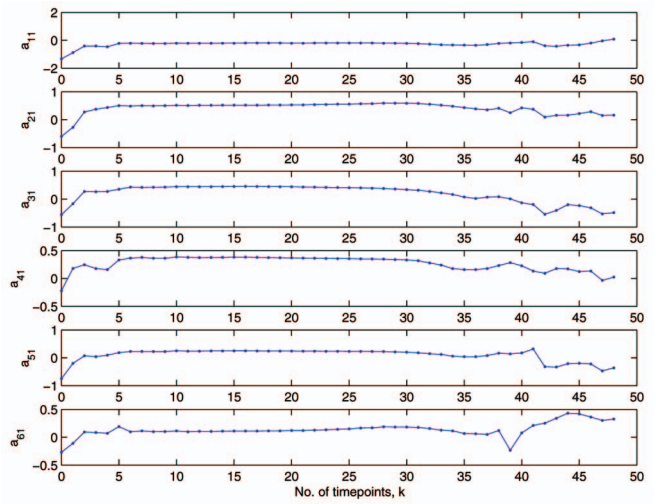


Fig. 1. The estimated time series of parameters $a_{11}, a_{21}, a_{31}, a_{41}, a_{51}, a_{61}$ (Malaria).

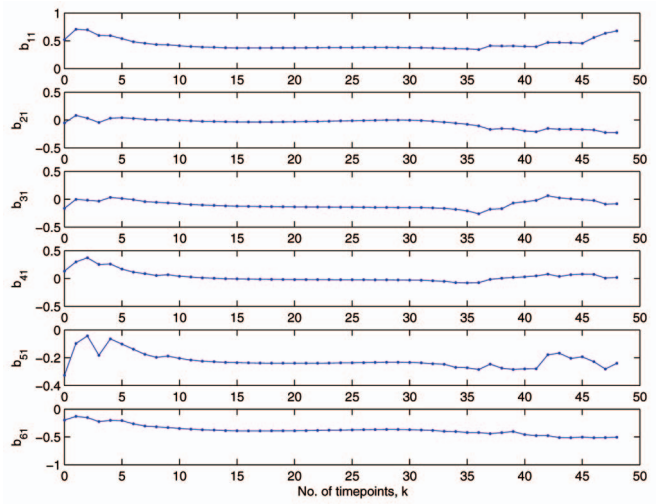


Fig. 2. The estimated time series of parameters $b_{11}, b_{21}, b_{31}, b_{41}, b_{51}, b_{61}$ (Malaria).

Since $\xi(k) \sim \mathcal{N}(0, R_0)$, it is obvious that $w(k) \sim \mathcal{N}(0, R)$ with $R = H^T R_0 H$, and it is also not difficult to see that

$$\frac{\partial F(X(k))}{\partial X(k)} = \begin{bmatrix} \Xi_1(X(k)) & \Xi_2(X(k)) & \Xi_3(X(k)) & \Xi_4(X(k)) & I \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{bmatrix}, \quad (28)$$

where

$$\Xi_1(X(k)) = A(k) + B(k)\frac{\partial f(x(k), \mu(k))}{\partial x(k)},$$
$$\Xi_2(X(k)) = [x_1(k)I, x_2(k)I, \ldots, x_n(k)I],$$
$$\Xi_3(X(k)) = [f(x_1(k), \mu_1(k))I, f(x_2(k), \mu_2(k))I, \ldots, f(x_n(k), \mu_n(k))I],$$
$$\Xi_4(X(k)) = B(k)\mathrm{diag}\left\{\frac{x_1(k)e^{-\mu_1(k)x_1(k)}}{(1+e^{-\mu_1(k)x_1(k)})^2}, \frac{x_2(k)e^{-\mu_2(k)x_2(k)}}{(1+e^{-\mu_2(k)x_2(k)})^2}, \ldots, \frac{x_n(k)e^{-\mu_n(k)x_n(k)}}{(1+e^{-\mu_n(k)x_n(k)})^2}\right\},$$

with

$$\frac{\partial f(x(k), \mu(k))}{\partial x(k)} = \mathrm{diag}\left\{\frac{\mu_1(k)e^{-\mu_1(k)x_1(k)}}{(1+e^{-\mu_1(k)x_1(k)})^2}, \frac{\mu_2(k)e^{-\mu_2(k)x_2(k)}}{(1+e^{-\mu_2(k)x_2(k)})^2}, \ldots, \frac{\mu_n(k)e^{-\mu_n(k)x_n(k)}}{(1+e^{-\mu_n(k)x_n(k)})^2}\right\}.$$

Remark 3. Based on the EKF approach, we can identify $2n^2 + 2n$ parameters. Specifically, we can estimate n^2 parameters a_{ij}s, n^2 parameters b_{ij}s, n parameters μ_is, and n parameters I_{0i}s. It is also worth pointing out that we can identify the n state variables as well.

4.1 Identifying the Malaria Model Parameters from Time Series

The first data set is from the Malaria gene expression time series [4]. It consists of 530 genes expressed 48 equally spaced time points. We select the first six genes expression time series given by

$$Z = [Z_1 \;\; Z_2 \;\; Z_3 \;\; Z_4 \;\; Z_5 \;\; Z_6],$$

where

$$Z_1 = \begin{bmatrix} 4.3140 & 3.2789 & 1.6684 & 1.7445 & 1.0716 & 0.9868 & 0.9900 & 0.7780 \\ 2.2710 & 1.8179 & 0.7923 & 1.2726 & 0.7282 & 0.5669 & 0.5280 & 0.4488 \\ 2.7890 & 2.3653 & 1.4219 & 1.3902 & 1.0680 & 0.8739 & 0.6490 & 0.7413 \\ 3.7880 & 2.5943 & 1.2601 & 1.8115 & 0.9243 & 0.8472 & 0.8310 & 0.6240 \\ 4.1620 & 2.9244 & 0.9809 & 2.1758 & 0.9998 & 0.8891 & 0.7450 & 0.5897 \\ 2.2080 & 2.0724 & 0.9977 & 1.3763 & 0.7307 & 0.4528 & 0.4890 & 0.5092 \end{bmatrix},$$

$$Z_2 = \begin{bmatrix} 0.8355 & 0.5796 & 0.4910 & 0.3782 & 0.3446 & 0.1460 & 0.1465 & 0.2114 \\ 0.5778 & 0.3129 & 0.2540 & 0.2401 & 0.2036 & 0.1260 & 0.1608 & 0.1577 \\ 0.5219 & 0.5056 & 0.3680 & 0.3691 & 0.3232 & 0.1730 & 0.1002 & 0.1133 \\ 0.9553 & 0.4316 & 0.4230 & 0.2943 & 0.2634 & 0.1360 & 0.1280 & 0.1028 \\ 0.9722 & 0.3823 & 0.3700 & 0.3343 & 0.3019 & 0.1280 & 0.1482 & 0.1168 \\ 0.4854 & 0.3545 & 0.2580 & 0.2504 & 0.2264 & 0.1170 & 0.1313 & 0.1554 \end{bmatrix},$$

$$Z_3 = \begin{bmatrix} 0.2061 & 0.1720 & 0.1678 & 0.1700 & 0.2155 & 0.2226 & 0.2101 & 0.1976 \\ 0.1710 & 0.2110 & 0.2138 & 0.2620 & 0.3233 & 0.2806 & 0.3582 & 0.4357 \\ 0.1239 & 0.1010 & 0.0642 & 0.0630 & 0.0632 & 0.0655 & 0.0467 & 0.0280 \\ 0.0811 & 0.0970 & 0.0518 & 0.0490 & 0.0427 & 0.0524 & 0.0496 & 0.0469 \\ 0.1129 & 0.1180 & 0.0839 & 0.0810 & 0.0948 & 0.0917 & 0.0995 & 0.1074 \\ 0.1666 & 0.2140 & 0.2089 & 0.2790 & 0.2675 & 0.3096 & 0.3894 & 0.4691 \end{bmatrix},$$

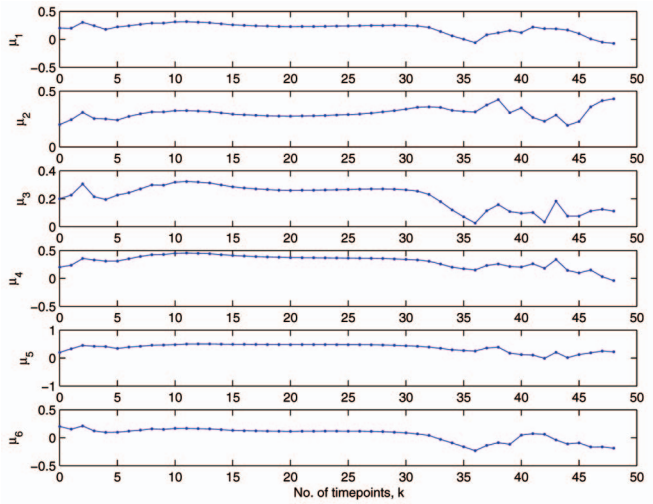


Fig. 3. The estimated time series of parameters $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ (Malaria).

$$Z_4 = \begin{bmatrix} 0.2375 & 0.2131 & 0.2530 & 0.1947 & 0.2148 & 0.2349 & 0.2650 & 0.6056 \\ 0.3711 & 0.4639 & 0.6410 & 0.6707 & 0.8082 & 0.9458 & 1.1440 & 1.3391 \\ 0.0608 & 0.0410 & 0.0440 & 0.0391 & 0.0850 & 0.1309 & 0.2050 & 0.3874 \\ 0.0544 & 0.0475 & 0.0750 & 0.0707 & 0.1066 & 0.1425 & 0.2100 & 0.5808 \\ 0.1016 & 0.1090 & 0.1280 & 0.1381 & 0.1739 & 0.2098 & 0.3030 & 0.5905 \\ 0.4062 & 0.5582 & 0.5920 & 0.7738 & 0.8656 & 0.9574 & 1.2510 & 1.2578 \end{bmatrix},$$

$$Z_5 = \begin{bmatrix} 1.0130 & 1.4945 & 1.9910 & 2.5285 & 1.7578 & 1.8211 & 2.5851 & 3.8840 \\ 1.9144 & 2.0826 & 2.3190 & 2.5555 & 2.9656 & 2.3457 & 3.3361 & 3.2779 \\ 0.9661 & 1.3078 & 1.8535 & 2.4930 & 1.7872 & 2.0033 & 3.4185 & 4.6765 \\ 1.0017 & 1.7174 & 1.9343 & 2.2905 & 2.0121 & 1.9548 & 4.0059 & 4.5845 \\ 0.8967 & 1.6631 & 1.7467 & 2.3982 & 1.8186 & 1.5144 & 3.6226 & 2.8834 \\ 1.9266 & 2.0004 & 2.4258 & 2.4844 & 2.8291 & 2.3201 & 7.6102 & 2.9527 \end{bmatrix},$$

$$Z_6 = \begin{bmatrix} 3.8805 & 6.0726 & 5.4836 & 4.6334 & 3.2207 & 1.0636 & 1.5610 & 1.1717 \\ 3.1208 & 4.1553 & 2.2738 & 2.0388 & 1.8348 & 1.5575 & 1.9512 & 1.4513 \\ 4.7711 & 6.6787 & 4.1907 & 4.6189 & 2.5593 & 2.3816 & 2.9104 & 2.3003 \\ 5.1805 & 6.1378 & 4.4675 & 4.1250 & 3.2643 & 1.9541 & 2.6247 & 1.9389 \\ 3.6588 & 6.9146 & 5.1801 & 4.6347 & 3.9337 & 2.8011 & 3.4341 & 2.1344 \\ 2.7262 & 4.1970 & 2.3114 & 2.3628 & 2.0484 & 1.6607 & 2.0003 & 1.3854 \end{bmatrix}.$$

Here, the ith row of matrix Z stands for the expression time series of the ith gene. We take $P(0|0) = 0.5I$, and let $\xi(k) \sim \mathcal{N}(0, 0.3^2 I)$ and $v(k) \sim \mathcal{N}(0, 0.4^2 I)$. Then, based on the EKF algorithm, we can identify all the parameters, which consist of $2n^2 + 2n$ system parameters and n system variables. Both the identified parameters and variables are expressed in the form of time series, and the time series for error covariances is also obtained simultaneously.

For the purpose of space saving, we only select partial parameters. They are $a_{11}, a_{21}, \ldots, a_{61}; b_{11}, b_{21}, \ldots, b_{61}; \mu_1, \mu_2, \ldots, \mu_6; I_{01}, I_{02}, \ldots, I_{06}$. The time series of the corresponding estimated parameters are depicted in Figs. 1, 2, 3, and 4.

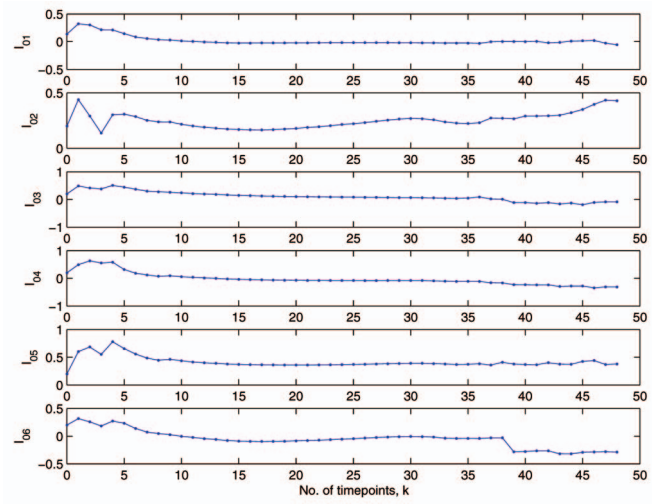


Fig. 4. The estimated time series of parameters $I_{01}, I_{02}, I_{03}, I_{04}, I_{05}, I_{06}$ (Malaria).

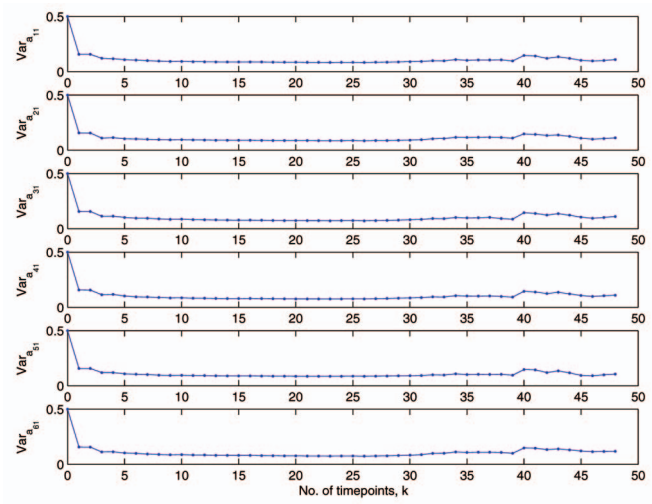


Fig. 5. The variances of estimated time series of parameters $a_{11}, a_{21}, a_{31}, a_{41}, a_{51}, a_{61}$ (Malaria).

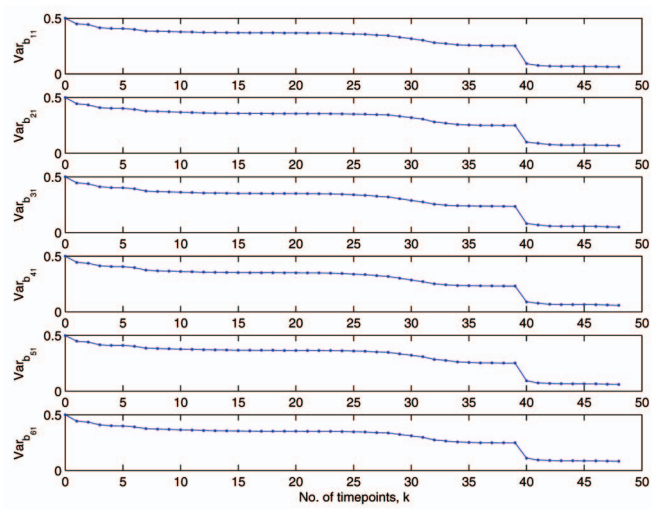


Fig. 6. The variances of estimated time series of parameters $b_{11}, b_{21}, b_{31}, b_{41}, b_{51}, b_{61}$ (Malaria).

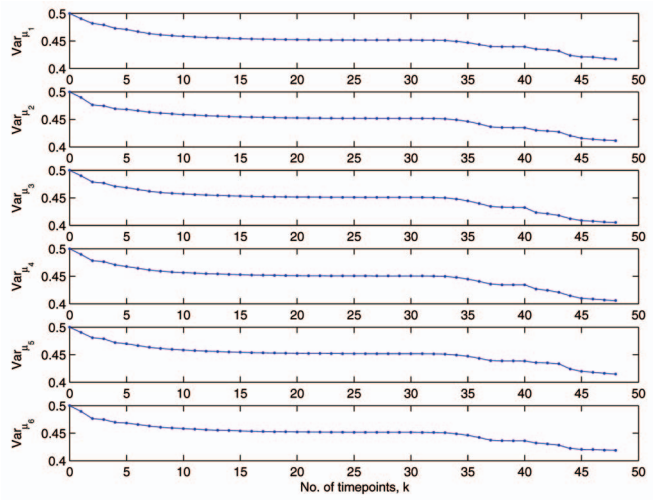


Fig. 7. The variances of estimated time series of parameters $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ (Malaria).

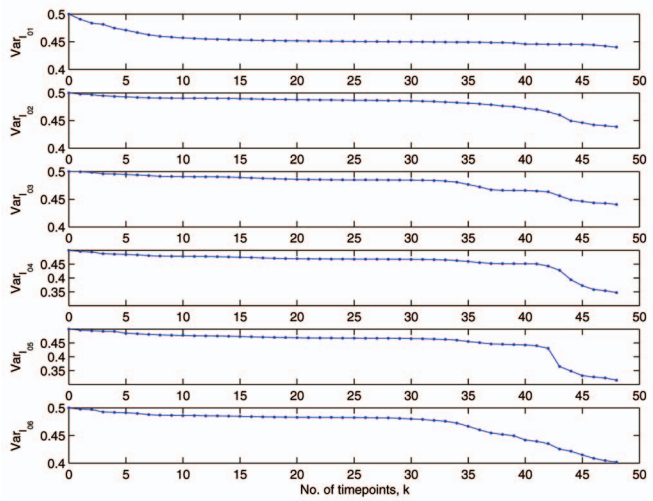


Fig. 8. The variances of estimated time series of parameters $I_{01}, I_{02}, I_{03}, I_{04}, I_{05}, I_{06}$ (Malaria).

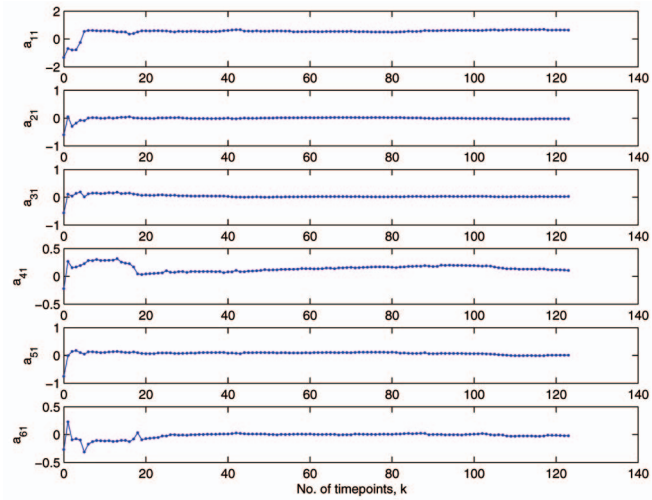


Fig. 9. The estimated time series of parameters $a_{11}, a_{21}, a_{31}, a_{41}, a_{51}, a_{61}$ (Worm).

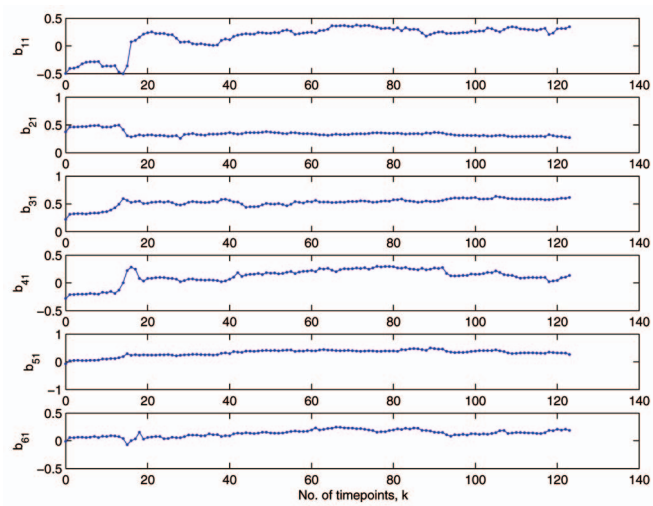


Fig. 10. The estimated time series of parameters $b_{11}, b_{21}, b_{31}, b_{41}, b_{51}, b_{61}$ (Worm).

In identifying the parameters by EKF approach, we can obtain the error covariance matrix $P(k|k)$ simultaneously. Note that the time series for error covariance matrix gives a quantitative criterion to evaluate the identification errors. Here, we display the individual parameter error variance in Figs. 5, 6, 7, and 8. In these figures, for a parameter χ, we denote its variance by Var_χ. We can observe in Figs. 5, 6, 7, and 8 that, when time points increase, the estimation error variances decrease. This shows that the EKF algorithm works well when modeling nonlinear dynamic gene regulatory networks via short gene expression time series.

4.2 Identifying the Worm Model Parameters from Time Series

The second data set is from the Worm gene expression time series [2], [30]. It consists of 98 genes expressed 123 equally spaced time points. We select the first 12 genes expression time series for our purpose. Here, we take $P(0|0) = 0.5I$, and let $\xi(k) \sim \mathcal{N}(0, 0.3^2 I)$ and $v(k) \sim \mathcal{N}(0, 0.4^2 I)$. Then, based on EKF algorithm, we can identify all the parameters. For the same purpose of saving room, we select only partial parameters. They are $a_{11}, a_{21}, \ldots, a_{61}; b_{11}, b_{21}, \ldots, b_{61}; \mu_1, \mu_2, \ldots, \mu_6; I_{01}, I_{02}, \ldots, I_{06}$. The time series of the corresponding estimated parameters are demonstrated in Figs. 9, 10, 11, and 12.

Remark 4. We can also specify the individual parameter error variance as done previously. However, in order to keep this paper concise, we omit the related figures hereafter.

4.3 Identifying the Yeast Model Parameters from Time Series

The third data set is from the Worm gene expression time series [46]. It consists of 237 genes expressed 17 equally spaced time points. We select the first 12 genes expression time series for our purpose. Here, we take $P(0|0) = 0.5I$, and let $\xi(k) \sim \mathcal{N}(0, 0.3^2 I)$ and $v(k) \sim \mathcal{N}(0, 0.4^2 I)$. Then, based on EKF algorithm, we can identify all the parameters. For the same purpose of saving room, we select only partial parameters. They are $a_{11}, a_{21}, \ldots, a_{61}; b_{11}, b_{21}, \ldots, b_{61}; \mu_1, \mu_2, \ldots, \mu_6; I_{01}, I_{02}, \ldots, I_{06}$. The time series of the corresponding

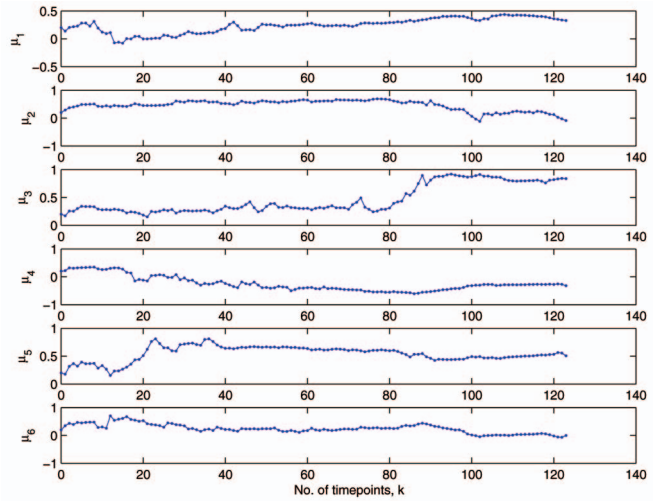


Fig. 11. The estimated time series of parameters $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ (Worm).

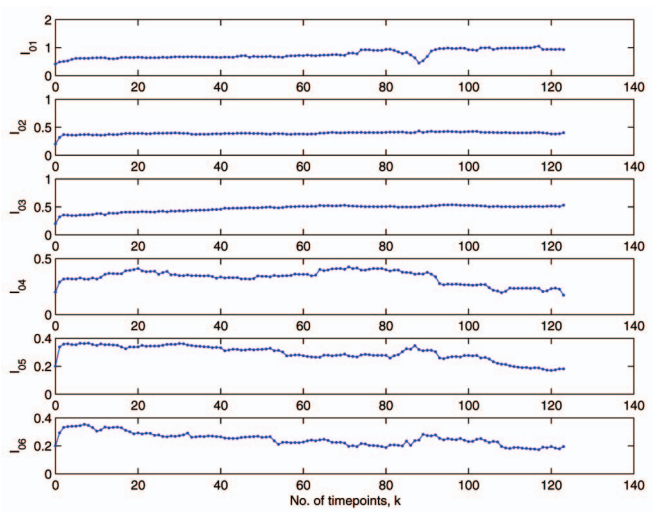


Fig. 12. The estimated time series of parameters $I_{01}, I_{02}, I_{03}, I_{04}, I_{05}, I_{06}$ (Worm).

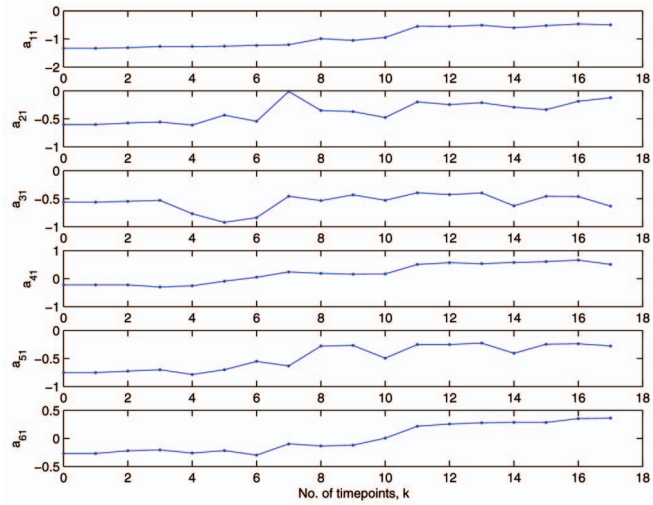


Fig. 13. The estimated time series of parameters $a_{11}, a_{21}, a_{31}, a_{41}, a_{51}, a_{61}$ (Yeast).

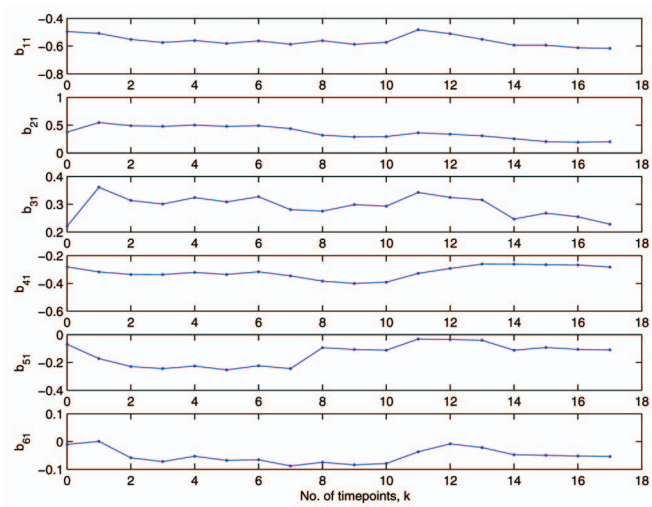


Fig. 14. The estimated time series of parameters $b_{11}, b_{21}, b_{31}, b_{41}, b_{51}, b_{61}$ (Yeast).

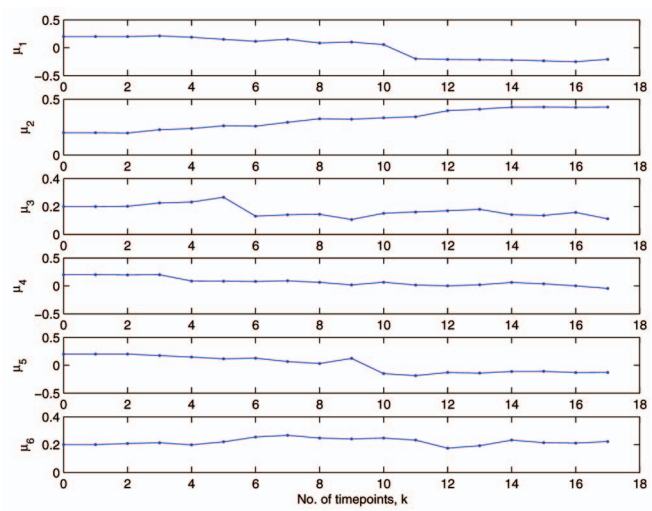


Fig. 15. The estimated time series of parameters $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ (Yeast).

estimated parameters are demonstrated in Figs. 13, 14, 15, and 16.

4.4 Identifying the Virus Model Parameters from Time Series

The fourth data set is for the virus gene expression microarray data from [24], which consists of 106 genes expressed at eight equally spaced time points. Similar to the previous cases, we can select partial parameters and display the time series of the corresponding estimated parameters. To keep this paper concise, we have omitted the figures.

5 Conclusions

In this paper, the EKF algorithm has been applied to model the gene regulatory network from gene time series data. We assume that the dynamics of the gene regulatory network under investigation is governed by a class of nonlinear stochastic differential equation. With the help of the EKF approach, we have identified all the parameters of the

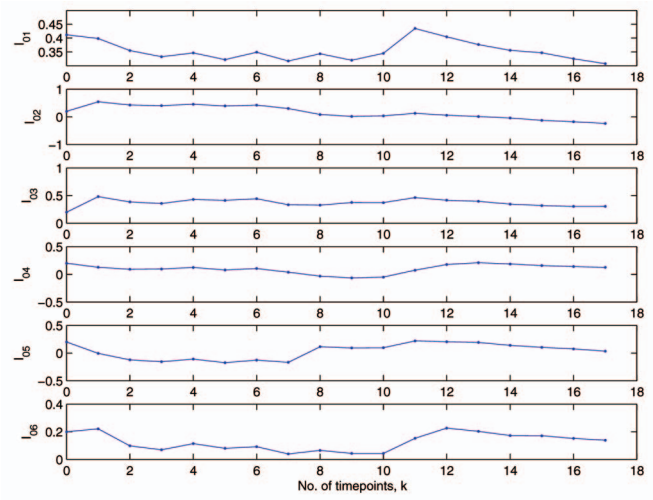


Fig. 16. The estimated time series of parameters $I_{01}, I_{02}, I_{03}, I_{04}, I_{05}, I_{06}$ (Yeast).

system. The numerical examples have shown that the EKF approach works well for modeling the given gene regulatory network.

ACKNOWLEDGMENTS

This work was supported in part by the Biotechnology and Biological Sciences Research Council (BBSRC) of the UK under Grants BB/C506264/1 and 100/EGM17735, the Engineering and Physical Sciences Research Council (EPSRC) of the UK under Grants GR/S27658/01 and EP/C524586/1, an International Joint Project sponsored by the Royal Society of the UK, the National Natural Science Foundation of China under Grants 60774073 and 60804028, the Teaching and Research Fund for Excellent Young Teachers at Southeast University of China, the Specialized Research Fund for the Doctoral Program of Higher Education for New Teachers in China under Grant 200802861044, and the Alexander von Humboldt Foundation of Germany.

REFERENCES

[1] T. Akutsu, S. Miyano, and S. Kuhara, "Identification of Genetic Networks from a Small Number of Gene Expression Patterns under The Boolean Network Model," *Proc. Pacific Symp. Biocomputing,* vol. 4, pp. 17-28, 1999.

[2] L.R. Baugh, A.A. Hill, J.M. Claggett, K. Hill-Harfe, J.C. Wen, D.K. Slonim, E.L. Brown, and C.P. Hunter, "The Homeodomain Protein PAL-1 Specifies a Lineage-Specific Regulatory Network in the *C. Elegans Embryo*," *Development,* vol. 132, pp. 1843-1854, 2005.

[3] M.J. Beal, F. Falciani, Z. Ghahramani, C. Rangel, and D.L. Wild, "A Bayesian Approach to Reconstructing Genetic Regulatory Networks with Hidden Factors," *Bioinformatics,* Oct. 2004.

[4] Z. Bozdech, M. Llinas, B.L. Pulliam, E.D. Wong, and J. Zhu, "The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium Falciparum," *PLoS Biology,* vol. 1, no. 1, pp. 85-100, 2003.

[5] T. Chen, H.L. He, and G.M. Church, "Modeling Gene Expression with Differential Equations," *Proc. Pacific Symp. Biocomputing,* vol. 4, pp. 29-40, 1999.

[6] L. Chen and K. Aihara, "Chaos and Asymptotical Stability in Discrete-Time Neural Networks," *Physica D: Nonlinear Phenomena,* vol. 104, pp. 286-325, 1997.

[7] D.L. Cook, A.N. Gerber, and S.J. Tapscott, "Modeling Stochastic Gene Expression: Implications for Haploinsufficiency," *Proc. Nat'l Academy of Science USA,* vol. 95, pp. 15641-15646, 1998.

[8] A. Corigliano and S. Mariani, "Parameter Identification in Explicit Structural Dynamics: Performance of the Extended Kalman Filter," *Computer Methods in Applied Mechanics and Eng.,* vol. 193, pp. 3807-3835, 2004.

[9] M.J. de Hoon, S. Imoto, K. Kobayashi, N. Ogasawara, and S. Miyano, "Inferring Gene Regulatory Networks from Time-Ordered Gene Expression Data of Bacillus Subtilis Using Differential Equations," *Proc. Pacific Symp. Biocomputing,* pp. 17-28, 2003.

[10] H. de Jong, "Modeling and Simulation of Genetic Regulatory Systems: A Literature Review," *J. Computational Biology,* vol. 9, no. 1, pp. 67-103, 2002.

[11] P. D'haeseleer, X. Wen, S. Fuhrman, and R. Somogyi, "Linear Modeling of mRNA Expression Levels during CNS Development and Injury," *Proc. Pacific Symp. Biocomputing,* pp. 41-52, 1999.

[12] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein, "Cluster Analysis and Display of Genome-Wide Expression Patterns," *Proc. Nat'l Academy of Science USA,* vol. 95, pp. 14863-14868, 1998.

[13] H. Gao, X. Meng, and T. Chen, "A Parameter-Dependent Approach to Robust H_∞ Filtering for Time-Delay Systems," *IEEE Trans. Automatic Control,* vol. 53, no. 10, pp. 2420-2425, 2008.

[14] H. Gao, X. Meng, and T. Chen, "A New Design of Robust H_2 Filters for Uncertain Systems," *Systems and Control Letters,* vol. 57, no. 7, pp. 585-593, 2008.

[15] Z. Gao, H. Wang, and T. Chai, "A Robust Fault Detection Filtering for Stochastic Distribution Systems via Descriptor Estimator and Parametric Gain Design," *IET Control Theory Applications,* vol. 1, no. 5, pp. 1286-1293.

[16] Z. Gao and S.X. Ding, "State and Disturbance Estimator for Time-Delay Systems with Application to Fault Estimation and Signal Compensation," *IEEE Trans. Signal Processing,* vol. 55, no. 12, pp. 5541-5551, 2007.

[17] Z. Gao and D.W.C. Ho, "State/noise Estimator for Descriptor Systems with Application to Sensor Fault Diagnosis," *IEEE Trans. Signal Processing,* vol. 54, no. 4, pp. 1316-1326, 2006.

[18] Z. Gao, X. Dai, T. Breikin, and H. Wang, "High-Gain Observer-Based Parameter Identification with Application in a Gas Turbine," *Proc. IFAC Control Congress,* pp. 1408-1413, July 2008.

[19] Z. Ghahramani, "Learning Dynamic Bayesian Networks," *Adaptive Processing of Sequences and Data Structures,* C.L. Giles and M. Gori, eds., pp. 168-197, Springer-Verlag, 1998.

[20] N.S. Holter, A. Maritan, M. Cieplak, N.V. Fedoroff, and J.R. Banavar, "Dynamic Modeling of Gene Expression Data," *Proc. Nat'l Academy of Science USA,* vol. 98, pp. 1693-1698, 2001.

[21] S. Huang, "Gene Expression Profiling, Genetic Networks, and Cellular States: An Integrating Concept for Tumorigenesis and Drug Discovery," *J. Molecular Medicine,* vol. 77, pp. 469-480, 1999.

[22] S. Huang and D. Dissanayake, "Convergence and Consistency Analysis for Extended Kalman Filter Based SLAM," *IEEE Trans. Robotics,* vol. 23, no. 5, pp. 1036-1049, 2007.

[23] A.K. Jain, M.N. Murty, and P.J. Flynn, "Data Clustering: A Review," *ACM Computing Surveys,* vol. 31, no. 3, pp. 264-323, 1999.

[24] P. Kellam, X. Liu, N. Martin, C. Orengo, S. Swift, and A. Tucker, "A Framework for Modeling Virus Gene Expression Data," *Intelligent Data Analysis,* vol. 6, pp. 265-279, 2002.

[25] T.B. Kepler and T.C. Elston, "Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations," *Biophysics J.,* vol. 81, no. 6, pp. 3116-3136, 2001.

[26] A.J. Krener, "The Convergence of the Extended Kalman Filter," *Directions in Mathematical Systems Theory and Optimization,* pp. 173-182, 2003.

[27] S. Liang, S. Fuhrman, and R. Somogyi, "REVEAL: A General Reverse Engineering Algorithm for Inference of Genetic Network Architectures," *Proc. Pacific Symp. Biocomputing,* vol. 3, pp. 18-29, 1998.

[28] T. Liu, W. Sung, and A. Mittal, "Model Gene Network by Semi-Fixed Bayesian Network," *Expert Systems with Applications,* vol. 30, no. 1, pp. 42-49, 2006.

[29] H.M. McAdams and A. Arkin, "Stochastic Mechanisms in Gene Expression," *Proc. Nat'l Academy of Science USA,* vol. 94, pp. 814-819, 1997.

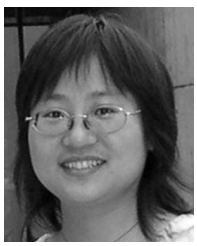
[30] M.F. Maduro and J.H. Rothman, "Making Worm Guts: The Gene Regulatory Network of the Caenorhabditis Elegans Endoderm," *Developmental Biology,* vol. 246, pp. 68-85, 2002.

[31] K. Murphy and S. Mian, "Modeling Gene Expression Data Using Dynamic Bayesian Networks," technical report, Univ. of California, 1999.
[32] M. Quach, N. Brunel, and F. d'Alché-Buc, "Estimating Parameters and Hidden Variables in Non-Linear State-Space Models Based on ODEs for Biological Networks Inference," *Bioinformatics,* vol. 23, no. 23, pp. 3209-3216, 2007.
[33] M.F. Ramoni, P. Sebastiani, and I.S. Kohane, "Cluster Analysis of Gene Expression Dynamics," *Proc. Nat'l Academy of Science USA,* vol. 99, pp. 9121-9126, 2002.
[34] C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E.A. Sotheran, A. Gaiba, D.L. Wild, and F. Falciani, "Modeling T-Cell Activation Using Gene Expression Profiling and State Space Models," *Bioinformatics,* vol. 20, no. 9, pp. 1361-1372, 2004.
[35] P. Smolen, D.A. Baxter, and J.H. Byrne, "Mathematical Modeling of Gene Networks Review," *Neuron,* vol. 26, no. 3, pp. 567-580, 2000.
[36] R. Somogyi and C.A. Sniegoski, "Modeling the Complexity of Genetic Networks: Understanding Multigenic and Pleiotropic Regulation," *Complexity,* vol. 1, no. 6, pp. 45-63, 1996.
[37] S. Swift and X. Liu, "Predicting Glaucomatous Visual Field Deterioration through Short Multivariate Time Series Modeling," *Artificial Intelligence in Medicine,* vol. 24, pp. 5-24, 2002.
[38] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub, "Interpreting Patterns of Gene Expression with Self-Organizing Maps: Methods and Application to Hematopoietic Differentiation," *Proc. Nat'l Academy of Science USA,* vol. 96, pp. 2907-2912, 1999.
[39] S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, and G.M. Church, "Systematic Determination of Genetic Network Architecture," *Nature Genetics,* vol. 22, no. 3, pp. 281-285, 1999.
[40] T. Thattai and A. van Oudenaarden, "Stochastic Gene Expression in Fluctuating Environments," *Proc. Genetics Soc. Am.,* pp. 523-530, 2004.
[41] T. Tian and K. Burrage, "Stochastic Neural Network Models for Gene Regulatory Networks," *Proc. 2003 IEEE Congress Evolutionary Computation,* pp. 162-169, 2003.
[42] Z. Wang, F. Yang, D.W.C. Ho, S. Swift, A. Tucker, and X. Liu, "Stochastic Dynamic Modeling of Short Gene Expression Time Series Data," *IEEE Trans. NanoBioscience,* vol. 7, no. 1, pp. 44-55, 2008.
[43] Z. Wang, H. Gao, J. Cao, and X. Liu, "On Delayed Genetic Regulatory Networks with Polytopic Uncertainties: Robust Stability Analysis," *IEEE Trans. NanoBioscience,* vol. 7, no. 2, pp. 154-163, 2008.
[44] F. Wu, W. Zhang, and A.J. Kusalik, "Modeling Gene Expression from Microarray Expression Data with State-Space Equations," *Proc. Pacific Symp. Biocomputing,* pp. 581-592, 2004.
[45] F. Lewis, L. Xie, and D. Popa, *Optimal and Robust Estimation: With an Introduction to Stochastic Control Theory,* CRC Press, 2007.
[46] K.Y. Yeung, C. Fraley, A. Murua, A.E. Raftery, and W.L. Ruzzo, "Model-Based Clustering and Data Transformations for Gene Expression Data," *Bioinformatics,* vol. 17, no. 10, pp. 977-987, 2001.

Zidong Wang (SM'03) received the BSc degree in mathematics from Suzhou University, Suzhou, China, in 1986, and the MSc degree in applied mathematics and the Ph.D. degree in electrical and computer engineering from Nanjing University of Science and Technology, Nanjing, China, in 1990 and 1994, respectively. He is now a professor of dynamical systems and computing at Brunel University, UK. His research interests include dynamical systems, signal processing, bioinformatics, and control theory and applications. He has published more than 120 papers in refereed international journals. He is currently serving as an associate editor for 12 international journals including the *IEEE Transactions on Automatic Control*, the *IEEE Transactions on Neural Networks*, the *IEEE Transactions on Signal Processing*, the *IEEE Transactions on Systems, Man, and Cybernetics—Part C*, and the *IEEE Transactions on Control Systems Technology*.

Xiaohui Liu received the BEng degree in computing from Hohai University, Nanjing, China, in 1982, and the PhD degree in computer science from Heriot-Watt University, Edinburg, UK, in 1988. He is currently a professor of computing at Brunel University. He leads the Intelligent Data Analysis (IDA) Group, performing interdisciplinary research involving artificial intelligence, dynamic systems, image and signal processing, and statistics, particularly for applications in biology, engineering, and medicine. He serves on editorial boards of four computing journals, founded the biennial international conference series on IDA in 1995, and has given numerous invited talks in bioinformatics, data mining, and statistics conferences.

Yurong Liu received the BSc degree in mathematics from Suzhou University, Suzhou, China, in 1986, the MSc degree in applied mathematics from Nanjing University of Science and Technology, Nanjing, China, in 1989, and the PhD degree in applied mathematics from Suzhou University, Suzhou, China, in 2000. Currently, he is a professor in the Department of Mathematics, Yangzhou University, China. His current interests include neural networks, nonlinear dynamics, time-delay systems, and chaotic dynamics.



Jinling Liang received the BSc and MSc degrees in mathematics from Northwest University, Xi'an, China, in 1997 and 1999, respectively, and the PhD degree in applied mathematics from Southeast University, Nanjing, China, in 2006. She is now an associate professor in the Department of Mathematics, Southeast University, Nanjing, China. From January 2004 to March 2004, she was a research assistant at the University of Hong Kong. From March to April 2004, she was a research assistant at the City University of Hong Kong. From April 2007 to March 2008, she was a postdoctoral research fellow in the Department of Information Systems and Computing, Brunel University, UK, sponsored by the Royal Society Sino-British Fellowship Trust Award of the UK. Her research interests include neural networks, complex networks, nonlinear systems, and bioinformatics. She has published around 20 papers in refereed international journals. She is a member of program committee for some international conferences and serves as a very active reviewer for many international journals.

Veronica Vinciotti received the first degree in mathematics from the University of Perugia, Italy, in 1999, and the PhD degree in statistics from Imperial College London, UK, in 2002. After a four-year research position at Brunel University, London, UK, working on the statistical analysis of gene expression data, she has been a lecturer in statistics at Brunel University since 2007. Her research interests include statistical classification methods, statistical analysis of bioinformatics and medical data, mathematical and statistical modeling of biological regulatory networks, and credit scoring. She has published a number of papers on the statistical analysis of gene expression data and on the statistical modeling of regulatory networks.