Stochastic Dynamic Modeling of Short Gene Expression Time-Series Data

Zidong Wang*, *Senior Member, IEEE*, Fuwen Yang, *Senior Member, IEEE*, Daniel W. C. Ho, *Senior Member, IEEE*, Stephen Swift, Allan Tucker, and Xiaohui Liu

***Abstract*—In this paper, the expectation maximization (EM) algorithm is applied for modeling the gene regulatory network from gene time-series data. The gene regulatory network is viewed as a stochastic dynamic model, which consists of the noisy gene measurement from microarray and the gene regulation first-order autoregressive (AR) stochastic dynamic process. By using the EM algorithm, both the model parameters and the actual values of the gene expression levels can be identified simultaneously. Moreover, the algorithm can deal with the sparse parameter identification and the noisy data in an efficient way. It is also shown that the EM algorithm can handle the microarrary gene expression data with large number of variables but a small number of observations. The gene expression stochastic dynamic models for four real-world gene expression data sets are constructed to demonstrate the advantages of the introduced algorithm. Several indices are proposed to evaluate the models of inferred gene regulatory networks, and the relevant biological properties are discussed.**

***Index Terms*—Clustering, DNA microarray technology, expectation maximization (EM) algorithm, gene expression, modeling, time-series data.**

I. Introduction

DNA MICROARRAY technology has provided an efficient way of measuring the expression levels of thousands of genes in a single experiment on a single "chip." It enables the monitoring of expression levels of thousands of genes simultaneously. This allows, for the first time, a global view on the expression levels of all genes when the cell undergoes specific conditions or processes. The potential of such technologies for functional genomics is tremendous. Measuring gene expression levels in different conditions may prove useful in medical diagnosis, treatment, and drug design. Microarray technology has been heralded as the new biological revolution after the advent of the human genome project, since it become possible to extract the important information from gene expression time-series data.

In order to infer useful biological information and determine the relationships between individual genes, many current research efforts have focused on clustering. Cluster analysis of the gene expression data appeared first in [13] and has quickly attracted considerable research attention. A number of clustering algorithms have been examined on gene expression data, such as hierarchical clustering [13], self-organizing map [34], k-means [35], and Gaussian-model-based clustering [28], [41], to name just a few [19]. However, a fundamental shortcoming of such clustering schemes is that they are based on the assumption that there exists the correlation similarity between genes. Recently, there has been an increasing research interest to reconstruct models for gene regulatory networks from time-series data [10], [31], such as Boolean network model [1], [18], [22], [32], linear differential equation model [7], [9], [11], [17], Bayesian model [16], [20], [23], [27], state-space model [4], [29], [40], and stochastic model [8], [37].

Obviously, selecting a good model to fit gene regulatory networks is essential to a meaningful analysis of the expression data. It turns out that the model for gene regulatory networks should posses the following three properties. First, the model should be easy to evolve the biological information such as the linear dynamical model. Second, the model should reflect the "stochastic" characteristics, since it is well known that the gene expression is an inherently stochastic phenomenon [21], [25], [27], [36]. Third, the observations (measurement outputs) of the model should be regarded as noisy due to our inability to perfectly and accurately (noise-free) measure gene expression levels. Fourth, in biology and medicine, the available time series (e.g., gene expression time series) typically consists of a large number of variables but with a small number of observations. Therefore, the modeling method should be capable of tackling short time series with acceptable accuracy.

There have been attempts to reconstruct models for gene regulatory networks by taking into account the aforementioned three properties. Dynamic Bayesian networks have been proposed to model gene expression time-series data [20], [23], [27]. The merits of dynamic Bayesian networks include the ability to model stochasticity and handle noisy/hidden variables. However, dynamic Bayesian networks need more complex algorithms such as the genetic algorithm [20], [33] to infer gene regulatory networks. Another model is the state-space model [4], [29], [40], whose main feature is that the gene expression

Manuscript received November 30, 2005; revised March 9, 2007. This work was supported in part by the Biotechnology and Biological Sciences Research Council (BBSRC) of the U.K. under Grants BB/C506264/1 and 100/EGM17735, in part by the Engineering and Physical Sciences Research Council (EPSRC) of the U.K. under Grants GR/S27658/01 and EP/C524586/1, in part by an International Joint Project sponsored by the Royal Society of the U.K., in part by the Nuffield Foundation of the U.K. under Grant NAL/00630/G, and in part by the Alexander von Humboldt Foundation of Germany. *Asterisk indicates corresponding author.*

*Z. Wang is with the Department of Information Systems and Computing, Brunel University, Uxbridge, UB8 3PH, U.K. (e-mail: zidong.wang@brunel.ac.uk).

F. Yang, S. Swift, A. Tucker, and X. Liu are with the Department of Information Systems and Computing, Brunel University, Uxbridge, UB8 3PH, U.K. (e-mail: Fuwen.Yang@brunel.ac.uk; stephen.swift@brunel.ac.uk; allan.tucker@brunel.ac.uk; Xiaohui.liu@brunel.ac.uk).

D. W. C. Ho is with the Department of Mathematics, City University of Hong Kong, Hong Kong (e-mail: madaniel@cityu.edu.hk).

Digital Object Identifier 10.1109/TNB.2008.2000149

value depends not only on the current internal state variables but also on the external inputs. It is very interesting that the external input is viewed as the previous time step observation, and the gene regulation matrix is obtained from the relationship between the current measurement, the previous measurement, and internal state variables [4], [29]. For the use of state-space models, the measurements need to be accurate, and a suitable dimension for the internal state variables needs to be determined beforehand, which raises considerable difficulties in experimentation and computation.

In this paper, we view the gene regulatory network as a dynamic stochastic model, which is composed 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 first-order autoregressive (AR) stochastic dynamic process. Note that it is very important to regard the models as stochastic, since the gene expression is of inherent stochasticity. Stochastic models can help conduct more realistic simulations of biological systems, and also set up a practical criterion for measuring the robustness of the mathematical models against stochastic noises. After specifying the model structure, we apply the expectation maximization (EM) algorithm for identifying both the model parameters and the actual value of gene expression levels. Note that the EM algorithm is a learning algorithm that can handle sparse parameter identification and noisy data very well. It is also shown that the EM algorithm can cope with the microarrary gene expression data with large number of variables but a small number of observations. Four real-world gene expression data sets are employed to demonstrate the effectiveness of our algorithm, and some indices are used to evaluate the models of inferred gene regulatory networks from the viewpoint of bioinformatics.

The remainder of this paper is organized as follows. In Section II, a stochastic dynamic model is described for genetic regulatory network, which takes into account the noisy measurement as well as the inherently stochastic phenomenon of the genetic regulatory process. The EM algorithm is introduced in Section III for handling the sparse parameter identification problem and the noisy data analysis. In Section IV, our developed algorithm is applied to four real-world gene expression data sets, and the biological significance is discussed in terms of certain criteria. Further discussion is made in Section V to explain the advantages and shortcomings of our method. Some concluding remarks and future research topics are provided in Section VI.

II. Stochastic Dynamic Model For Gene Expression Data

Measuring gene expression levels by DNA microarray technologies has made a great progress in understanding the interaction among genes and extracting functional information. However, the gene expression data measured are often contaminated by measurement noises in a discrete-time fashion, because gene expression time series represent discrete "snapshots" of gene expression at various time points. Therefore, gene expression levels measured can be modeled as

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

where $y_i(k)$ is the measurement data of the ith gene expression levels from microarray 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 genes, and m is the measurement time points. Without the loss of generality, we assume that $v_i(k)$ is a zero mean Gaussian white noise sequence with covariance $V_i > 0$.

Next, we model the gene regulatory network containing n genes by the following stochastic discrete-time dynamic system:

$$x_i(k+1) = -\lambda_i x_i(k) + \sum_{j=1}^{n} a_{i,j} x_j(k) + w_i(k),$$
$$i = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, m \tag{2}$$

where λ_i is the self-degradation rate of the ith gene expression product and $a_{i,j}$ represents the regulatory relationship and degree among genes. A positive value for $a_{i,j}$ means the jth gene stimulating the expression of the ith gene and, similarly, a negative value for $a_{i,j}$ stands for the jth gene repressing the expression of the ith gene, while a value of zero indicates that jth gene does not influence the transcription of ith gene. This way, each gene in the organism can have multiple inputs, both positive and negative, of differing strength. $w_i(k)$ is the system noise. We also assume that $w_i(k)$ is a zero mean Gaussian white noise sequence with covariance $W_i > 0$, and $w_i(k)$ and $v_i(k)$ are mutually independent.

Remark 1: The measurement noise $v_i(k)$ and the system noise $w_i(k)$ in the model (1) and (2) are assumed to be zero mean Gaussian white noises. Such an assumption, however, does not lose the generality. The noises can also be modeled as the colored noises, which does not cause difficulties in our algorithm proposed later. More details about whitening noises can be found in [2]. For simplicity, we only consider the case in which $v_i(k)$ and $w_i(k)$ are zero mean Gaussian white noises.

Now, denote

$$x(k) = [\, x_1(k) \quad x_2(k) \quad \cdots \quad x_n(k) \,]^T, \qquad k = 1, 2, \ldots, m \tag{3}$$

and

$$a_i = [\, a_{i,1}\ a_{i,2} \cdots -\lambda_i + a_{i,i} \cdots a_{i,n} \,], \quad i = 1, 2, \ldots, n. \tag{4}$$

We can rewrite (2) as

$$x_i(k+1) = a_i x(k) + w_i(k), \qquad i = 1, 2, \ldots, n. \tag{5}$$

In this paper, our aim is to establish the model (1) and (5) from the measurement data

$$Y := \{y_1(1), y_1(2), \ldots, y_1(m), y_2(1), y_2(2), \ldots, y_2(m), \ldots, y_n(1), y_n(2), \ldots, y_n(m)\}.$$

This is a system identification problem. Notice that for gene expression time-series data, we typically have a *large* number of variables but a *small* number of observations. Unfortunately, traditional identification methods such as the least square method cannot be suitably used for the system identification problem with large number of variables but small number of observations, since they basically require a large amount of observations. Therefore, this problem becomes an "underdetermined" one from the viewpoint of system identification. To handle the data shortage problem, we introduce the EM algorithm to identify the model (1) and (5). Before introducing our algorithm, we define the vector

$$\theta = [a_1 a_2 \cdots a_n W_1 W_2 \cdots W_n V_1 V_2 \cdots V_n] \tag{6}$$

which consists of all parameters to be estimated in (1) and (5).

In the next section, we will develop a computationally efficient iterative method based on EM algorithm for identifying the parameter θ.

III. EM Algorithm for Parameter Identification

In this section, we first introduce the main idea of the EM algorithm. Then, to solve the specified gene network modeling problem, we will derive the iterative computation procedure for the proposed model (1) and (5) by using the Kalman filtering and Kalman smoothing approaches.

The EM algorithm, for time-series analysis, was first presented by Shumway and Stoffer [30]. It is a general iterative method to compute maximum likehood (ML) estimates of a set of parameters. It has been shown in various signal processing applications that the use of EM algorithm leads to computationally efficient estimation algorithms [38], [42].

The EM algorithm can be divided into two steps: E-step and M-step. The E-step is to estimate the logarithm likehood of the complete data using the observed data and the current parameter estimate, and the M-step is to maximize the estimated logarithm likehood function to obtain the new parameter estimate. Here, given the observations Y and the current parameter estimate, we define the natural logarithm of the conditional expectation of probability density functions for the completed data as the following logarithm likehood function [30]

$$J(\theta, \theta^{(l)}) = \mathbb{E}_{\theta^{(l)}}\left[L(X, Y, \theta)|Y\right] \tag{7}$$

where

$$L(X,Y,\theta) = -\sum_{i=1}^{n}\left\{\frac{m}{2}\ln|W_i| + \frac{1}{2}\sum_{k=1}^{m}[x_i(k) - a_i x(k-1)]^T \times W_i^{-1}[x_i(k) - a_i x(k-1)] + \frac{(m+1)}{2}\ln|V_i| + \frac{1}{2}\sum_{k=0}^{m}[y_i(k) - x_i(k)]^T \times V_i^{-1}[y_i(k) - x_i(k)]\right\} + C \tag{8}$$

with the constant C being independent of θ.

The new parameter estimate can be obtained by

$$\theta^{(l+1)} = \arg\max_{\theta} J(\theta, \theta^{(l)}). \tag{9}$$

Next, the new parameter estimate $\theta^{(l+1)}$ can be found by maximizing $J(\theta, \theta^{(l)})$. We maximize $J(\theta, \theta^{(l)})$ with respect to a_i, W_i^{-1}, and V_i^{-1}, respectively, and obtain

$$\frac{\partial J}{\partial a_i} = \mathbb{E}_{\theta^{(l)}}\left\{\sum_{k=1}^{m} W_i^{-1}[x_i(k) - a_i x(k-1)]x(k-1)^T|Y\right\} = 0 \tag{10}$$

$$\frac{\partial J}{\partial W_i^{-1}} = \mathbb{E}_{\theta^{(l)}}\left\{\frac{m}{2}W_i - \frac{1}{2}\sum_{k=1}^{m}[x_i(k) - a_i x(k-1)][x_i(k) - a_i x(k-1)]^T|Y\right\} = 0 \tag{11}$$

$$\frac{\partial J}{\partial V_i^{-1}} = \mathbb{E}_{\theta^{(l)}}\left\{\frac{m+1}{2}V_i - \frac{1}{2}\sum_{k=0}^{m}[y_i(k) - x_i(k)][y_i(k) - x_i(k)]^T|Y\right\} = 0. \tag{12}$$

From (10)–(12), we have

$$a_i^{(l+1)} = \left\{\sum_{k=1}^{m}\mathbb{E}_{\theta^{(l)}}[x_i(k)x(k-1)^T|Y]\right\} \times \left\{\sum_{k=1}^{m}\mathbb{E}_{\theta^{(l)}}[x(k-1)x(k-1)^T|Y]\right\}^{-1} \tag{13}$$

$$W_i^{(l+1)} = \frac{1}{m}\left\{\sum_{k=1}^{m}\mathbb{E}_{\theta^{(l)}}[x_i(k)x_i^T(k)|Y] - a_i^{(l+1)} \times \sum_{k=1}^{m}\mathbb{E}_{\theta^{(l)}}[x(k-1)x_i^T(k)|Y] - \sum_{k=1}^{m}\mathbb{E}_{\theta^{(l)}}[x_i(k)x^T(k-1)|Y](a_i^{(l+1)})^T + a_i^{(l+1)}\sum_{k=1}^{m}\mathbb{E}_{\theta^{(l)}}[x(k-1)x^T(k-1)|Y](a_i^{(l+1)})^T\right\} \tag{14}$$

$$V_i^{(l+1)}=\frac{1}{m+1}\left\{\sum_{k=0}^{m}[y_i(k)y_i^T(k)]-\sum_{k=0}^{m}\mathbb{E}_{\theta^{(l)}}[x_i(k)|Y]y_i^T(k)\right.$$
$$\left.-\sum_{k=0}^{m}y_i(k)\mathbb{E}_{\theta^{(l)}}[x_i^T(k)|Y]+\sum_{k=0}^{m}\mathbb{E}_{\theta^{(l)}}[x_i(k)x_i^T(k)|Y]\right\}. \tag{15}$$

The EM algorithm is an iterative numerical method for computing the maximum likehood estimate. Letting θ^0 be the initial parameter estimate, the EM algorithm generates a sequence of parameter estimates as follows.

1) *E-Step:* Set $\theta=\theta^{(l)}$ and compute $J(\theta,\theta^{(l)})$ in (7).
2) *M-Step:* Compute $a_i^{(l+1)}$, $W_i^{(l+1)}$, and $V_i^{(l+1)}$ in (13)–(15) from $i=1$ to n.

Obviously, in order to compute (7) and (13)–(15), we should first get the conditional expectations for $\mathbb{E}_{\theta^{(l)}}[x_i(k)|Y]$, $\mathbb{E}_{\theta^{(l)}}[x_i(k)x_i^T(k)|Y]$, and $\mathbb{E}_{\theta^{(l)}}[x_i(k)x(k-1)^T|Y]$. In the following, we will provide the Kalman filtering and Kalman smoothing algorithms to compute them.

Before giving the algorithm, we denote

$$\hat{x}^{(l)}(k|m):=\mathbb{E}_{\theta^{(l)}}[x(k)|Y] \tag{16}$$
$$\Sigma^{(l)}(k|m):=\mathbb{E}_{\theta^{(l)}}\{[x(k)-\hat{x}^{(l)}(k|m)][x(k)-\hat{x}^{(l)}(k|m)]^T|Y\} \tag{17}$$
$$\Pi^{(l)}(k,k-1|m):=\mathbb{E}_{\theta^{(l)}}\{[x(k)-\hat{x}^{(l)}(k|m)][x(k-1)-\hat{x}^{(l)}(k-1|m)]^T|Y\}. \tag{18}$$

Since

$$\mathbb{E}_{\theta^{(l)}}\{[x(k)-\hat{x}^{(l)}(k|m)][x(k)-\hat{x}^{(l)}(k|m)]^T|Y\}$$
$$=\mathbb{E}_{\theta^{(l)}}[x(k)x^T(k)|Y]-\hat{x}^{(l)}(k|m)[\hat{x}^{(l)}(k|m)]^T$$

and

$$\mathbb{E}_{\theta^{(l)}}\{[x(k)-\hat{x}^{(l)}(k|m)][x(k-1)-\hat{x}^{(l)}(k-1|m)]^T|Y\}$$
$$=\mathbb{E}_{\theta^{(l)}}[x(k)x^T(k-1)|Y]-\hat{x}^{(l)}(k|m)[\hat{x}^{(l)}(k-1|m)]^T$$

thus, we can get $\mathbb{E}_{\theta^{(l)}}[x(k)x^T(k)|Y]$ and $\mathbb{E}_{\theta^{(l)}}[x(k)x^T(k-1)|Y]$ from $\hat{x}^{(l)}(k|m)$, $\Sigma^{(l)}(k|m)$, and $\Pi^{(l)}(k,k-1|m)$.

The computation of the conditional expectations in (16)–(18) can be carried out using the Kalman filtering and smoothing methods. To do that, we may represent (1) and (5) in the following state-space form:

$$x(k+1)=Ax(k)+w(k) \tag{19}$$
$$y(k)=x(k)+v(k) \tag{20}$$

where

$$A=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix},\quad w(k)=\begin{bmatrix}w_1(k)\\w_2(k)\\\vdots\\w_n(k)\end{bmatrix},\quad y(k)=\begin{bmatrix}y_1(k)\\y_2(k)\\\vdots\\y_n(k)\end{bmatrix},$$

$$v(k)=\begin{bmatrix}v_1(k)\\v_2(k)\\\vdots\\v_n(k)\end{bmatrix}. \tag{21}$$

Denote the current parameter estimate $\theta=\theta^{(l)}$, then A, Q, and R is replaced by $A^{(l)}$, $Q^{(l)}$, and $R^{(l)}$, where

$$Q=\mathrm{diag}\{W_1,W_2,\ldots,W_n\},\quad R=\mathrm{diag}\{V_1,V_2,\ldots,V_n\}.$$

Therefore, at the current iteration cycle, $\hat{x}^{(l)}(k|m)$, $\Sigma^{(l)}(k|m)$, and $\Pi^{(l)}(k,k-1|m)$ can be obtained from the following algorithm [15], [42]:

1) *Forward (Filtering) Recursions:* For $k=1,2,\ldots,m$ Propagation equations

$$\hat{x}^{(l)}(k+1|k)=A^{(l)}\hat{x}^{(l)}(k|k) \tag{22}$$
$$\Sigma^{(l)}(k|k-1)=A^{(l)}\Sigma^{(l)}(k-1|k-1)(A^{(l)})^T+Q^{(l)}. \tag{23}$$

Updating equations

$$K^{(l)}(k)=\Sigma^{(l)}(k|k-1)[\Sigma^{(l)}(k|k-1)+R^{(l)}]^{-1} \tag{24}$$
$$\hat{x}^{(l)}(k|k)=\hat{x}^{(l)}(k|k-1)+K^{(l)}(k)\times[y(k)-C\hat{x}^{(l)}(k|k-1)] \tag{25}$$
$$\Sigma^{(l)}(k|k)=\Sigma^{(l)}(k|k-1)-\Sigma^{(l)}(k|k-1)\times[\Sigma^{(l)}(k|k-1)+R^{(l)}]^{-1}\Sigma^{(l)}(k|k-1). \tag{26}$$

2) *Backward (Smoothing) Recursions:* For $k=m,m-1,\ldots,1$

$$\Lambda^{(l)}(k-1)=\Sigma^{(l)}(k-1|k-1)(A^{(l)})^T[\Sigma^{(l)}(k|k-1)]^{-1} \tag{27}$$
$$\hat{x}^{(l)}(k-1|N)=\hat{x}^{(l)}(k-1|k-1)+\Lambda^{(l)}(k-1)\times[\hat{x}^{(l)}(k|N)-\hat{x}^{(l)}(k|k-1)] \tag{28}$$
$$\Sigma^{(l)}(k-1|N)=\Sigma^{(l)}(k-1|k-1)+\Lambda^{(l)}(k-1)\times[\Sigma^{(l)}(k|N)-\Sigma^{(l)}(k|k-1)]\Lambda^{(l)}(k-1)^T \tag{29}$$

and

$$\Pi^{(l)}(k,k-1|m)=\Pi^{(l)}(k,k-1|k)+[\Sigma^{(l)}(k|m)-\Sigma^{(l)}(k|k)][\Sigma^{(l)}(k|k)]^{-1}\Pi^{(l)}(k,k-1|k) \tag{30}$$
$$\Pi^{(l)}(k,k-1|k)=[I-K_k^{(l)}]A^{(l)}\Sigma^{(l)}(k-1|k-1). \tag{31}$$

Remark 2: The EM algorithm is only guaranteed to converge to a local maximum of the likehood function. Therefore, in order to ensure convergence to the global maximum, a good initialization procedure may be required. To initialize the Kalman smoothing equations, we need to specify $\hat{x}^{(l)}(0|0)$ and $\Sigma^{(l)}(0|0)$. We may use the first observed data to specify $\hat{x}^{(0)}(0|0)$ and $\Sigma^{(0)}(0|0)$. These initial estimates can then be iteratively improved by using the final estimates from the previous iteration cycle, i.e., $\hat{x}^{(l+1)}(0|0)=\hat{x}^{(l)}(0|m)$ and $\Sigma^{(l+1)}(0|0)=\Sigma^{(l)}(0|m)$.

Remark 3: Since biologically the resulting gene regulatory is expected to be sparse, we set some of the matrix entries equal

to zero, and infer the network using only the nonzero entries. If we know some parameters $a_{i,j}$ *a priori*, we do not need to include the known $a_{i,j}$ in θ for computation and only specify them in the matrix A. Moreover, if one group of genes are not related with other groups of genes, we can divide them into several groups. Several small gene regulatory networks are only computed, which reduces the computational complexity. Note that other conventional system identification algorithms such as least square method cannot be used to deal with the sparse data in such an effective way.

IV. Simulation Results

In order to evaluate the performances of the proposed algorithm, we adopt four real-world gene expression data sets, that is, the yeast gene expression time series [41], the virus gene expression time series [20], the human malaria gene expression time series [5], and the worm gene expression time series [3], [26]. Our modeling process is carried out after data preprocessing. Normalization is applied to the gene expression profile by taking log ratios first, and then, mean centering. After the normalization, the aforementioned EM algorithm is employed to these data sets in order to model the gene expression network dynamics.

In order to be concise, we will elaborate the identification process for yeast and virus gene expression time series in Sections IV-A and IV-B, but will briefly describe the simulation results for dynamic modeling of human malaria and worm gene expression time series in Section IV-C.

A. Modeling of Yeast Gene Expression Time Series

The first data set is from the yeast gene expression experiment, which consists of expression levels of 237 genes at 17 equally spaced time points, selected by Yeung *et al*. [41]. This data set is available from the website http://faculty.washington.edu/kayee/model/.

By using the proposed EM algorithm, for the first data set, the gene regulation matrix for group 3 of yeast gene expression experiment is obtained by $A = [\, A_1 \quad A_2 \quad A_3 \,]$, where

$$A_1 = \begin{bmatrix} -0.4824 & -0.3440 & -0.4248 & 0.5235 & 0.3822 & -0.4154 \\ -0.0579 & -0.0962 & 0.1152 & 0.3520 & -0.0779 & -0.1062 \\ -0.1092 & -0.1901 & 0.2481 & 0.2095 & -0.0055 & 0.1430 \\ 0.0161 & 0.3041 & -0.1126 & -0.0678 & -0.1018 & -0.2795 \\ -0.1161 & 1.0210 & 0.5017 & 0.0996 & -0.0030 & 0.8956 \\ 0.2056 & -0.0771 & -0.0145 & 0.2066 & -0.0373 & -0.1362 \\ 0.2625 & 0.5455 & -0.2064 & -0.1530 & 0.1482 & -0.5774 \\ 0.1012 & 0.0835 & 0.1955 & 0.0490 & 0.1986 & 0.5117 \\ 0.1452 & 0.6964 & -0.0621 & 0.2799 & 0.0441 & 0.3137 \\ -0.0162 & 0.2625 & -0.3951 & -0.3521 & -0.2136 & -0.2032 \\ 0.2221 & 0.2684 & 0.0338 & -0.0121 & 0.2217 & -0.5210 \\ 0.2114 & -1.2157 & 0.3920 & -0.0611 & 0.3203 & 0.3511 \\ 0.0352 & 0.2100 & -0.2358 & -0.1400 & -0.2304 & 0.0839 \\ 0.2967 & 0.3958 & -0.1568 & 0.4209 & 0.0581 & 0.0183 \\ -0.2405 & 0.3095 & 0.0738 & 0.2097 & 0.0047 & -0.0408 \\ 0.1657 & -0.1242 & -0.1044 & 0.1269 & -0.1694 & 0.2038 \\ 0.3554 & -0.2441 & 0.0842 & -0.2436 & 0.3571 & 0.4504 \\ -0.0808 & 0.2869 & 0.1084 & 0.2325 & -0.0934 & -0.0234 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.6976 & 0.3902 & 0.5201 & -0.5319 & 0.1039 & -0.2079 \\ 0.0411 & -0.1198 & 0.6151 & 0.2917 & -0.2760 & 0.4203 \\ -0.0605 & -0.1433 & 0.1162 & 0.4696 & 0.1618 & 0.0102 \\ -0.1204 & -0.1456 & 0.1277 & 0.3725 & -0.2130 & -0.2519 \\ -0.0261 & -0.2367 & -0.7025 & -0.0071 & 0.0810 & -0.4779 \\ 0.1804 & -0.1333 & 0.3043 & 0.2478 & 0.2310 & -0.4495 \\ -0.0740 & -0.5611 & 0.0397 & -0.1946 & 0.2223 & -0.5854 \\ -0.0467 & -0.3921 & 0.0190 & 0.1391 & 0.5357 & 0.2563 \\ 0.2650 & -0.6973 & 0.2149 & 0.1320 & 0.1059 & -0.1450 \\ -0.2038 & -0.3563 & 0.1770 & -0.4520 & -0.4880 & 0.1747 \\ -0.1232 & 0.1624 & 1.1356 & -0.3625 & 0.1256 & 0.5414 \\ -0.0671 & 0.8208 & -0.3664 & -0.0691 & -0.3463 & 0.0199 \\ -0.0543 & 0.0862 & -0.0724 & 0.0897 & 0.0350 & 0.0254 \\ 0.1491 & 0.0122 & -0.3639 & -0.2656 & 0.3253 & 0.1790 \\ -0.0051 & -0.2914 & 0.6284 & 0.5105 & 0.0863 & -0.0651 \\ 0.2998 & 0.0877 & 0.3486 & 0.3826 & 0.2239 & -0.1961 \\ 0.1233 & 0.2794 & -0.0905 & 0.6867 & -0.3379 & 0.0683 \\ -0.0159 & -0.0276 & -0.0028 & 0.2788 & -0.1808 & 0.0664 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} -0.4381 & -0.7611 & 0.1691 & -0.1328 & -0.4932 & 0.1544 \\ 0.0156 & 0.0058 & 0.1407 & -0.1606 & -0.2516 & 0.7453 \\ 0.0221 & -0.2250 & 0.4366 & -0.0785 & -0.2756 & 0.0278 \\ 0.3515 & -0.1879 & 0.1373 & 0.1641 & 0.2014 & -0.1100 \\ 0.3996 & 0.7104 & -0.1843 & 0.1939 & -0.1773 & -1.3029 \\ -0.0246 & -0.2462 & 0.2970 & -0.0906 & -0.2234 & -0.4861 \\ 0.0124 & 0.2977 & -0.0892 & -0.0805 & 0.3771 & 0.5676 \\ 0.1859 & -0.2419 & 0.0550 & 0.0537 & -0.2569 & 0.4901 \\ 0.1126 & 0.1182 & 0.0445 & 0.0969 & -0.2355 & -1.0189 \\ 0.5669 & -0.4872 & -0.5928 & 0.2526 & -0.0132 & 0.7997 \\ 0.6818 & -0.2117 & -0.1965 & -0.0988 & -0.1643 & -0.0749 \\ -0.1555 & 0.2701 & -0.4681 & -0.3162 & 0.3189 & 0.7409 \\ 0.5366 & -0.4739 & -0.0497 & 0.4096 & 0.0768 & 0.2256 \\ 0.4707 & 0.1121 & 0.1047 & 0.1813 & 0.2154 & -0.7320 \\ 0.1140 & 0.0285 & 0.0575 & -0.0551 & -0.0846 & -0.5759 \\ -0.1705 & -0.3424 & -0.0993 & 0.0050 & -0.1188 & 0.1565 \\ 0.0780 & -0.1529 & -0.3316 & -0.0158 & -0.1467 & 0.1716 \\ -0.0564 & -0.1127 & 0.3264 & 0.0387 & 0.1280 & 0.0218 \end{bmatrix}.$$

The covariance of the yeast gene network model W and the covariance of the yeast gene expression measurement noise V are, respectively, calculated as shown at the bottom of the next page.

As we can see from the aforementioned, all the parameters of the proposed stochastic dynamic model can be easily obtained by using our algorithm. Furthermore, the predicted values of gene expression levels are also obtained. We can observe that, for the gene expression levels, there exist differences between the actual values and the predicted (simulated) values, and the prediction errors of every yeast gene are shown in Figs. 1–4.

B. Modeling of Virus Gene Expression Time Series

The second data set is for the virus gene expression microarray data from [20], which consists of 106 genes expressed at eight equally spaced time points.

Again, by using the proposed EM algorithm to the second data set, the gene regulation matrix A for group one of virus gene expression experiment is obtained as shown at the bottom of the next page.

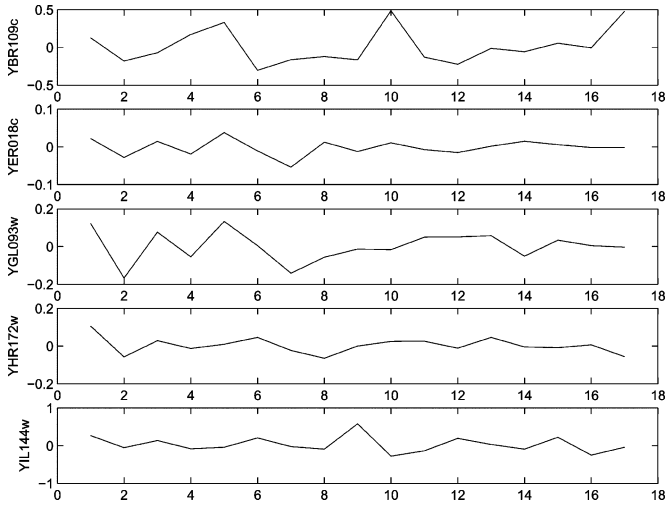

Fig. 1. Measurement errors of yeast genes (part 1).

The covariance of viral gene network model is

$$W = [0.0900\ \ 0.0560\ \ 0.0803\ \ 0.2862\ \ 0.0688$$
$$0.1253\ \ 0.1194\ \ 0.0753\ \ 0.0263\ \ 0.1453]$$

and the covariance of viral gene expression measurement noise is

$$V = [1.0196\ \ 0.0350\ \ 0.4137\ \ 9.5014\ \ 0.5865$$
$$4.2930\ \ 4.7446\ \ 0.9643\ \ 0.0000\ \ 11.4422].$$

All the parameters of stochastic dynamic model as well as the noise intensity are simultaneously calculated, and the prediction errors of every virus gene are illustrated in Figs. 5 and 6.

C. Modeling of Human Malaria and Worm Gene Expression Time Series

The third data set is from the human malaria gene expression time series [5]. As stated in [5], *Plasmodium falciparum*is responsible for the vast majority of the 300–500 million episodes of malaria worldwide and accounts for 0.7–2.7 million annual deaths. A comprehensive understanding of *Plasmodium* molecular biology will be essential for the development of new chemotherapeutic and vaccine strategies. Therefore, it is of great importance to model the human malaria expression data, which are made throughout the invasion process, with no observable abrupt change in the expression program upon successful reinvasion. The human malaria expression data set consists of 530 genes expressed at 48 equally spaced time points. We select a group of 15 genes and apply the proposed EM algorithm. All the model parameters can be obtained, which are not given here for the purpose of saving space. To illustrate the usefulness of the proposed modeling method, we display the prediction errors of every human malaria gene in Figs. 7–9.

The fourth data set is from the worm gene expression time series [3], [26], which consists of 98 genes expressed at 123 equally spaced time points. Again, we select a group of 15 genes, apply the proposed EM algorithm, and display the prediction errors of the selected worm gene in Figs. 10–12.

V. Discussions

A. Model Quality Evaluation

Since it is generally difficult to understand the real gene regulatory networks completely by biological experiments at present, some researchers [39], [40] proposed several indices to evaluate the models for gene regulatory networks from the viewpoint of bioinformatics, such as the computational cost, the prediction power (error), the stability, the robustness, and the periodicity. Obviously, different evaluation standards should

$$W = [0.0555\ \ 0.0075\ \ 0.0109\ \ 0.0083\ \ 0.0388\ \ 0.0126\ \ 0.0365\ \ 0.0076\ \ 0.0051$$
$$0.0114\ \ 0.0073\ \ 0.0253\ \ 0.0090\ \ 0.0090\ \ 0.0132\ \ 0.0203\ \ 0.0231\ \ 0.0020]$$
$$V = [0.1327\ \ 0.0006\ \ 0.0311\ \ 0.0105\ \ 0.0787\ \ 0.0011\ \ 0.1682\ \ 0.0075\ \ 0.0170$$
$$0.0021\ \ 0.0017\ \ 0.0157\ \ 0.0012\ \ 0.0014\ \ 0.0706\ \ 0.0805\ \ 0.0661\ \ 0.0001].$$

$$A = \begin{bmatrix} -0.0380 & 0.2864 & 0.0224 & 0.1894 & -0.1095 & 0.0563 & 0.1044 & 0.2009 & 0.4790 & 0.2404 \\ 0.0358 & -0.0149 & 0.4452 & 0.2760 & -0.3539 & 0.0146 & 0.0201 & -0.0330 & -1.1245 & 0.4650 \\ -0.0578 & 0.2458 & 0.2424 & 0.2885 & -0.2094 & 0.0932 & 0.1267 & 0.1010 & -0.0704 & 0.0864 \\ 0.5355 & -0.0647 & 0.7688 & 0.1994 & -0.0052 & 0.0681 & 0.2249 & -0.0024 & 0.6778 & 0.3909 \\ 0.0727 & 0.4132 & 0.1664 & 0.0991 & -0.1088 & -0.0145 & -0.0234 & -0.0024 & -0.1631 & 0.3268 \\ 0.1106 & 0.1621 & 0.2381 & 0.1420 & 0.0652 & -0.0501 & 0.0183 & -0.0303 & 0.2039 & 0.2081 \\ -0.0108 & 0.2278 & 0.2461 & 0.1063 & -0.1258 & 0.0094 & -0.0296 & -0.0394 & 0.1782 & 0.0468 \\ 0.1280 & 0.1246 & 0.2874 & 0.0801 & 0.0095 & -0.0049 & 0.0025 & -0.1375 & 0.4621 & 0.1838 \\ -0.0758 & 0.2457 & 0.1075 & -0.1461 & -0.2045 & -0.0659 & 0.0181 & -0.1992 & -0.1900 & -0.2829 \\ 0.3163 & -0.3176 & 0.0884 & -0.0962 & -0.2591 & -0.0060 & 0.2219 & 0.1804 & 0.8216 & -0.0740 \end{bmatrix}.$$

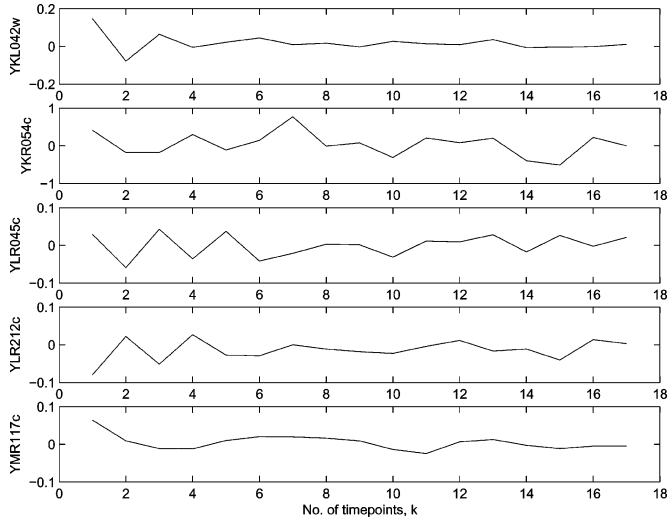

Fig. 2. Measurement errors of yeast genes (part 2).

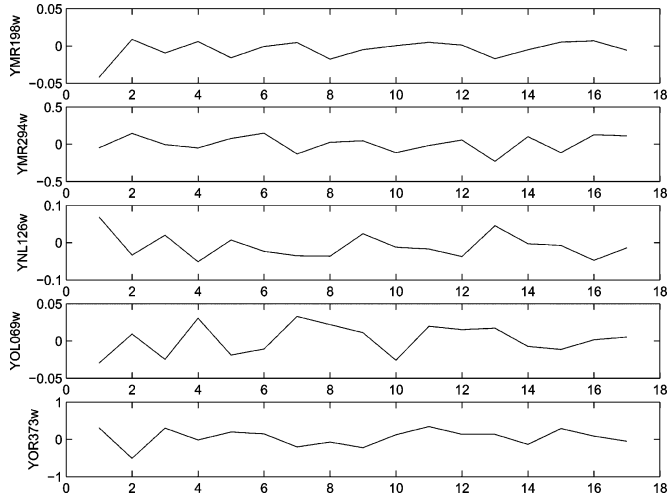

Fig. 3. Measurement errors of yeast genes (part 3).

be applied to different kinds of models. Since our models are stochastic and the measurements are noisy, our evaluation indices will mainly focus on the computational cost, the estimation covariance, the stability, and the robustness.

For the computational cost, our EM algorithm is an iterative learning algorithm that does not involve searching. The computational complexity is only related with the number of genes, the time points, and the number of iteration. From our simulations on the four data sets, the computational time is in seconds on a PC computer; hence, the computational cost is light.

From our experiments on the four data sets, the estimation covariances W are small, which means that our models fit the data very well. The covariances V represent the quality of measuring gene expression levels using microarray. For example, examining the covariances V for the established yeast and virus time-series models, we can see that the measurement of yeast gene expression levels is accurate, whereas the measurement of virus gene expression levels is not quite accurate, because the covariances V is smaller for the yeast measurement and the covariances V is bigger for the virus measurement. Furthermore, in order to evaluate the model quality in a quantitative way, let us introduce the following criterion for the modeling errors (error ratio in percentage) between the actual and predicted data [24]:

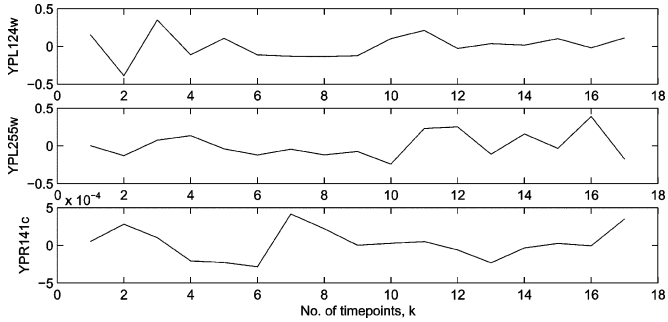

Fig. 4. Measurement errors of yeast genes (part 4).

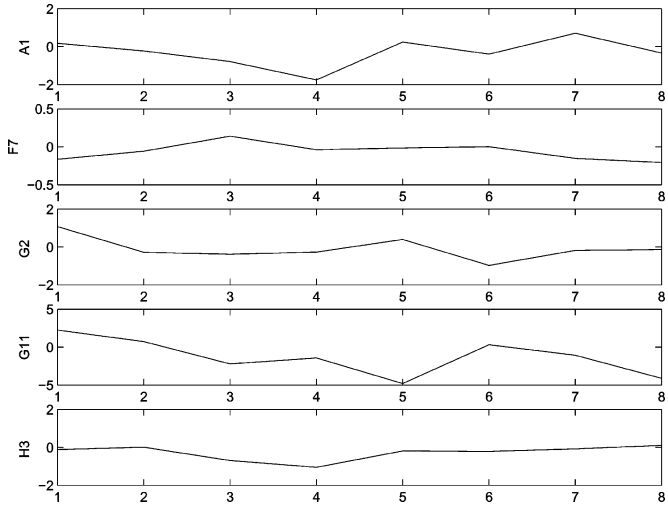

Fig. 5. Measurement errors of virus genes (part 1).

$$\text{Error ratio} = 100 \times \frac{1}{l}\sum_{c=1}^{l}\left[\sqrt{\frac{\sum_{k=1}^{s}(y_{ck}-\hat{y}_{ck})^2}{\sum_{k=1}^{s} y_{ck}^2}}\right]\% \quad (32)$$

where l is the number of genes (dimension) involved in the modeling, s is the number of observations (length), and y_{ck} is the actual gene expression value for cth gene at the kth time point. The results are given in the following table:

It can be seen from Table I that, the model quality is generally satisfactory. The publicly available yeast gene expression time-series data is of a good quality that leads to the best model. It is not surprising that the model for the virus gene expression time series is relatively the worst simply because of the poor quality of the data set (only eight observations are made for each gene). In fact, the lengths for all the four time series considered here are very short, and, as will be discussed later, traditional modeling approaches fail to cope with the short time-series modeling due to the assumption on the length of the time series.

In order to check the stability and robustness of our models, we need to compute the eigenvalues of the regulation matrices of the four models. For the virus gene expression time series,

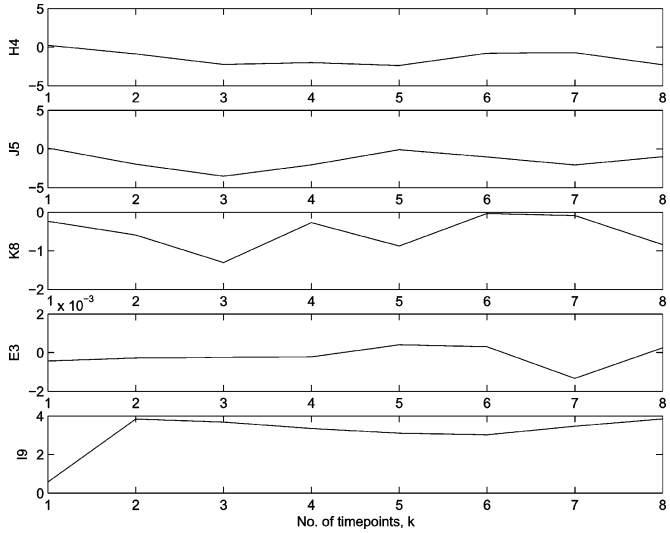

Fig. 6. Measurement errors of virus genes (part 2).

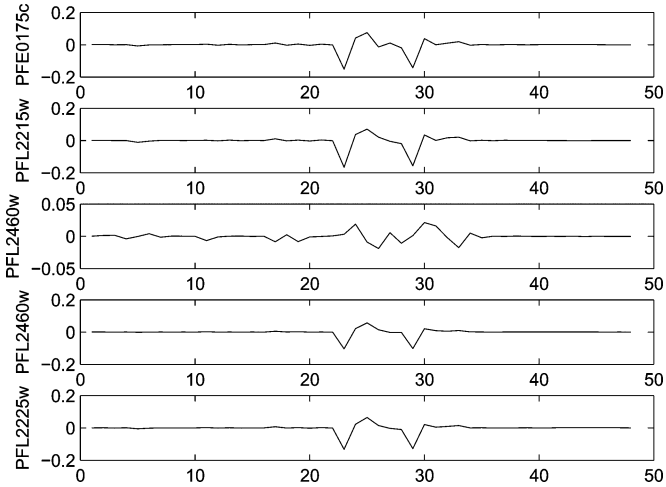

Fig. 7. Measurement errors of human malaria genes (part 1).

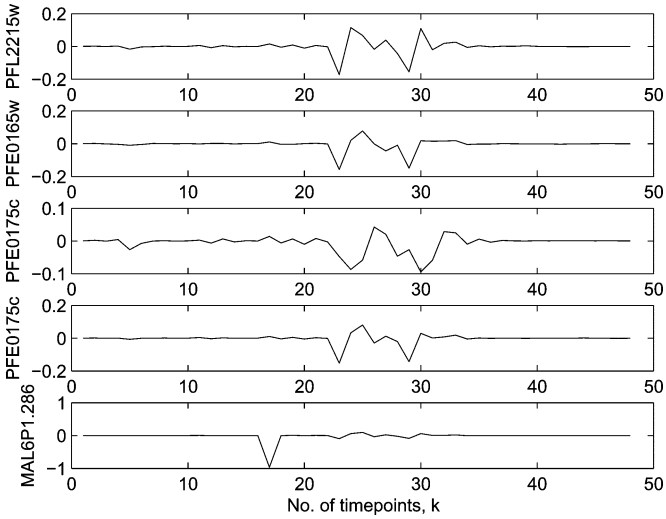

Fig. 8. Measurement errors of human malaria genes (part 2).

the malaria gene expression time series and the worm gene expression time series, the sets of the eigenvalues of the regulation matrix are, respectively,

$$\begin{aligned}\mathbb{E}_{\text{virus}} &= \{-0.1277 \pm 0.8784i, 0.0893 \pm 0.4355i, 0.8692,\\ &\quad -0.4490, -0.3501, -0.0729, -0.0604, -0.0611\}\\ \mathbb{E}_{\text{malaria}} &= \{-0.4902 \pm 0.7282i, -0.8811, 0.3662\\ &\quad \pm 0.6502i, 0.6410, -0.3408, 0.1044, -0.1660,\\ &\quad -0.1055, -0.0847, -0.0486 \pm 0.0046i,\\ &\quad -0.0606 \pm 0.0038i\}\\ \mathbb{E}_{\text{worm}} &= \{0.9993, 0.5957, -0.1149 \pm 0.2028i, 0.0307,\\ &\quad -0.0953 \pm 0.0522i, -0.0461, -0.0562,\\ &\quad -0.0875 \pm 0.0014i, -0.0720, -0.0807,\\ &\quad -0.0767 \pm 0.0017i\}.\end{aligned}$$

Obviously, for the models of virus, malaria, and worm gene expression time series, all eigenvalues lie well inside the unit circle. Therefore, the models are stable and robust. For the yeast data set, 18 eigenvalues of the regulation matrix are given by $0.9067 \pm 0.3304i, 0.7305 \pm 0.6659i, 0.3706 \pm 0.8439i, 0.0406 \pm 0.9303i, -0.2429 \pm 0.8529i, -0.5551 \pm 0.6032i, -0.7877 \pm 0.3466i, -0.0063 \pm 0.0102i, -1.0260, -0.3956$. All of these except one lie inside the unit circle. Although one of the eigenvalues is outside the unit circle, it is very close to 1. If the gene regulatory network is periodic, this eigenvalue would not cause the instability. Hence, the model is almost stable and robust.

B. Comparisons With Existing Modeling Methods

In biology and medicine, it is quite common for the mu ltivariate time series (MTS) data to be rather short, either because of the expense involved in obtaining data, e.g., in high-throughput bioinformatics areas such as microarrays, or due to the practicalities such as patients' treatment period or mortality.

Traditionally, statistical methods have been proposed when modeling the MTS, e.g., the Vector Auto-Regressive (VAR) process, VAR Moving Average, Markov Chain Monte Carlo methods, and other nonlinear and Bayesian systems [6], [12]. In the computing community, many MTS forecasting methods have been proposed using recurrent or time-delay neural networks, evolutionary computation, inductive logic programming, and support vector machines, see [6], [14], and the references therein.

However, one area where there has been little work is the analysis of a particular type of time series in which the data set consists of a large number of variables but with a small number of observations. Almost all traditional methods for modeling MTS place constraints on the minimum number of time-series observations in the dataset; many methods require distribution assumptions to be made regarding the observed time series, e.g., the maximum likelihood method for parameter estimation [6].

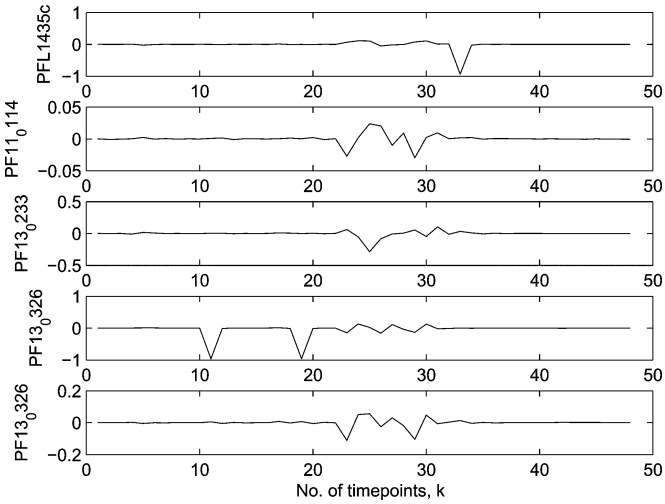

Fig. 9. Measurement errors of human malaria genes (part 3).

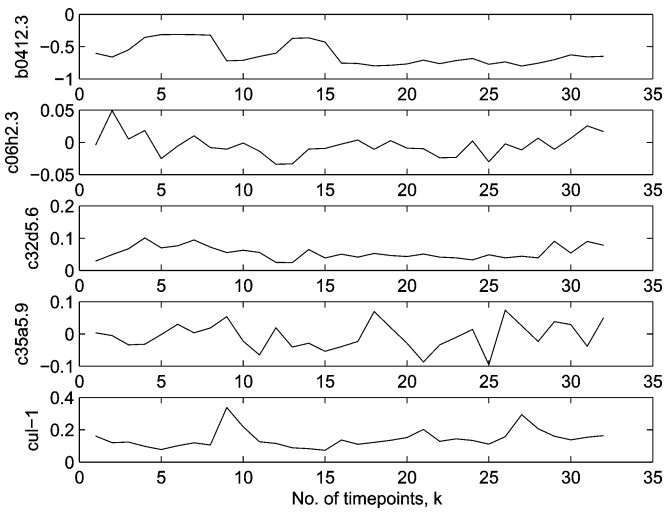

Fig. 10. Measurement errors of worm genes (part 1).

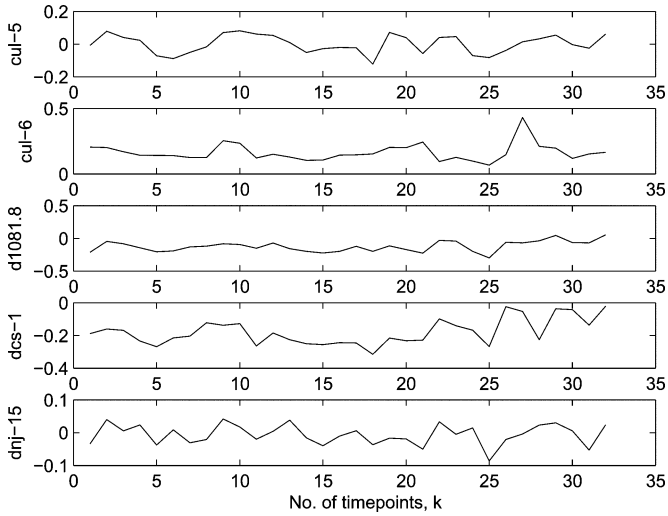

Fig. 11. Measurement errors of worm genes (part 2).

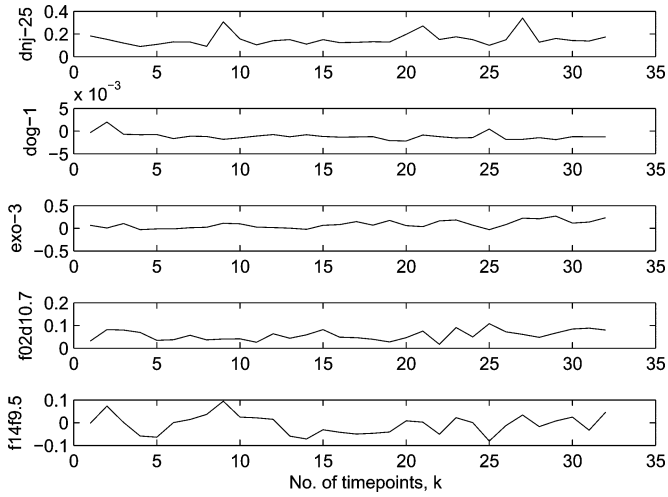

Fig. 12. Measurement errors of worm genes (part 3).

For example, a traditional way of modeling MTS data is the VAR process for a model of order P. The standard statistical methods for fitting a VAR process to a set of data often consist of two steps: order selection and parameter estimation. Order selection is commonly performed through the use of information-theory-based matrices such as Akaike's Information Criterion. Many of these matrices will impose a restriction on the minimum length of an MTS, N, based on the number of degrees of freedom of the model being estimated: $N > KP + 1$, where K is the number of variables being modeled and P is the order of the VAR process. For example, for an MTS consisting of 100 variables, to find the most appropriate order of a VAR process with a maximum order of three under consideration, N must be at least 302. This restriction is clearly unacceptable for modeling many short, high-dimensional time series, which are common in biology and medicine.

For the four gene expression data sets considered in this paper, the number of genes (dimension) and the number of observations (length) are given in Table II. It can be seen clearly from Table II that the gene expression time series is pretty short but with high dimensions, for which the traditional MTS modeling methods are simply impossible to be applied in a satisfactory way.

Recently, the modeling problem for short, high-dimensional time series has begun to receive some research interests. For example, dynamic Bayesian networks have been proposed to model gene expression time-series data [20], [27] with the ability of handling noisy/hidden variables. However, dynamic Bayesian networks need more complex algorithms such as the genetic algorithm [20], [33] to infer gene regulatory networks, and the noise intensity cannot be obtained directly. In previous sections, we have provided an algorithm to infer a gene regulatory network in the form of stochastic dynamic models. Using our algorithm, the gene regulation matrix can be extracted from *a very short number* of gene expression time-series data. This matrix can be employed to figure out how genes act "in concert" to achieve specific phenotypic characteristics [39].

TABLE I
QUANTITATIVE MODEL EVALUATION

Gene expression time series	Yeast	Virus	Malaria	Worm
Modeling error ratio	10.60%	42.19%	23.26%	28.91%

TABLE II
DIMENSION VS LENGTH

Gene expression time series	Yeast	Virus	Malaria	Worm
Dimension	237	106	530	98
Length	17	8	48	123

Compared to the existing MTS modeling methods, our algorithm has the following advantages.

1) Our algorithm can effectively tackle short, high-dimensional time series that typically occurs in biology and medical sciences.
2) The defined gene regulation matrix can reflect the relationship and interaction between genes, where a_{ij} stands for the effect of jth gene on ith gene. The model is of direct biological significance.
3) Our scheme can identify the entire network structure. For the existing differential equation method, there exists an "underdetermined" problem if the number of genes is equal to or larger than the number of experiments due to the shortage of gene expression data. In this paper, we use iterative "learning" procedure so that the global dynamic model can be obtained from a small number of gene expression data.
4) Our method can deal with noises in gene expression data measurement and sparse connectivity. Since the measurements of mRNA concentration using microarrays are typically noisy, it is very advantageous that our algorithm is robust to noises while identifying gene regulatory networks.
5) Our algorithm can tackle the spare gene regulatory networks only by setting some of the matrix entries as zero. The algorithm is especially efficient for larger gene regulatory networks that can be divided into several individual gene regulatory networks.
6) Our algorithm is very efficient for computation, and the computational cost is light even if there is a lack of gene expression data.

Nevertheless, our algorithm cannot deal with the data sets with missing values, which often occurs in many gene measurement data. Furthermore, a better data preprocessing method should be explored to maintain the biological meaning while simplifying the computation, since different preprocessing (normalization) methods would cause different gene regulation matrices. On the modeling issue, the models would be different for different iteration times and initial conditions. Proper biological knowledge should be incorporated to our algorithm, and some constraints need to be added to make the solution set smaller. Moreover, for simplicity, we have adopted a stochastic linear model with constant coefficients to infer gene regulatory network. How to consider the influence of time-varying coefficients, nonlinearities, and time delays will be the topic of our further investigation.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have developed an EM algorithm for modeling the gene regulatory networks from gene expression time-series data. The gene regulation matrix has been obtained by an iterative learning procedure based on the stochastic linear dynamic model. Gene expression stochastic dynamic models for four real-world gene expression data sets have been constructed to show how our algorithm works. Our algorithm can tackle the spare gene regulatory networks only by setting some of the matrix entries as zero. Furthermore, our algorithm is especially efficient for larger gene regulatory networks that can be divided into several individual gene regulatory networks. Therefore, our algorithm can be ideally applied to modeling the gene regulatory networks where the real connectivity of the network is specified *a priori*. We expect that our algorithm will be useful for reconstructing gene regulatory networks on a genome-wide scale, and hope that our results will benefit for biologists and biological computation scientists. In the near future, we will continue to investigate some real-world gene expression data sets and apply our algorithm to reconstruct gene regulatory networks with missing data, sparse connectivity, periodicity, and time delays. We are also getting connection with the biologists to explain our results from the biological point of view.

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," in *Proc. Pacific Symp. Biocomput.*, 1999, vol. 4, pp. 17–28.
[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[3] 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.
[4] 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*, vol. 21, no. 3, pp. 349–356, Oct. 2004.
[5] Z. Bozdech, M. Llinas, B. L. Pulliam, E. D. Wong, and J. Zhu, "The transcriptome of the intraerythrocytic developmental cycle of plasmodium falciparum," *PLoS Biol.*, vol. 1, no. 1, pp. 85–100, 2003.
[6] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. London, U.K.: Chapman and Hall, 2004.
[7] T. Chen, H. L. He, and G. M. Church, "Modeling gene expression with differential equations," in *Proc. Pacific Symp. Biocomput.*, 1999, vol. 4, pp. 29–40.
[8] D. L. Cook, A. N. Gerber, and S. J. Tapscott, "Modeling stochastic gene expression: Implications for haploinsufficiency," in *Proc. Nat. Acad. Sci.*, USA, 1998, vol. 95, pp. 15641–15646.
[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," in *Proc. Pacific Symp. Biocomput.*, 2003, pp. 17–28.
[10] H. de Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *J. Comput. Biol.*, 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," in *Proc. Pacific Symp. Biocomput.*, 1999, pp. 41–52.

[12] P. Diggle, *Time Series: A Biostatistical Introduction*, Oxford Statistical Science Series 5, 1990.
[13] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," in *Proc. Nat. Acad. Sci.*, USA, 1998, vol. 95, pp. 14863–14868.
[14] F. Gers, N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learning Res.*, vol. 3, pp. 115–143, 2002.
[15] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Univ. Toronto, Toronto, Canada, Tech. Rep., CRG-TR-96-2, 1996.
[16] Z. Ghahramani, "Learning dynamic Bayesian networks," in *Adaptive Processing of Sequences and Data Structures*, C. L. Giles and M. Gori, Eds. Lecture Notes in Artificial Intelligence. Berlin, Germany: Springer-Verlag, pp. 168–197.
[17] N. S. Holter, A. Maritan, M. Cieplak, N. V. Fedoroff, and J. R. Banavar, "Dynamic modeling of gene expression data," in *Proc. Nat. Acad. Sci.*, USA, 2001, vol. 98, pp. 1693–1698.
[18] S. Huang, "Gene expression profiling, genetic networks, and cellular states: An integrating concept for tumorigenesis and drug discovery," *J. Molecular Med.*, vol. 77, pp. 469–480, 1999.
[19] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
[20] P. Kellam, X. Liu, N. Martin, C. Orengo, S. Swift, and A. Tucker, "A framework for modelling virus gene expression data," *Intell. Data Anal.*, vol. 6, pp. 265–279, 2002.
[21] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations," *Biophys. J.*, vol. 81, no. 6, pp. 3116–3136, 2001.
[22] S. Liang, S. Fuhrman, and R. Somogyi, "REVEAL: A general reverse engineering algorithm for inference of genetic network architectures," in *Proc. Pacific Symp. Biocomput.*, 1998, vol. 3, pp. 18–29.
[23] T. Liu, W. Sung, and A. Mittal, "Model gene network by semi-fixed Bayesian network," *Expert Syst. Appl.*, vol. 30, no. 1, pp. 42–49, 2006.
[24] L. Ljung, *System Identification—Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
[25] H. M. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," in *Proc. Nat. Acad. Sci.*, USA, 1997, vol. 94, pp. 814–819.
[26] M. F. Maduro and J. H. Rothman, "Making worm guts: The gene regulatory network of the Caenorhabditis elegans endoderm," *Dev. Biol.*, vol. 246, pp. 68–85, 2002.
[27] K. Murphy and S. Mian, "Modelling gene expression data using dynamic Bayesian networks", Univ. California, Berkeley, CA, Tech. Rep., 1999.
[28] M. F. Ramoni, P. Sebastiani, and I. S. Kohane, "Cluster analysis of gene expression dynamics," in *Proc. Nat. Acad. Sci.*, USA, 2002, vol. 99, pp. 9121–9126.
[29] 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.
[30] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, vol. 3, pp. 253–264, 1982.
[31] P. Smolen, D. A. Baxter, and J. H. Byrne, "Mathematical modeling of gene networks review," *Neuron*, vol. 26, no. 3, pp. 567–580, 2000.
[32] 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.
[33] S. Swift and X. Liu, "Predicting glaucomatous visual field deterioration through short multivariate time series modelling," *Artif. Intell. Med.*, vol. 24, pp. 5–24, 2002.
[34] 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," in *Proc. Nat. Acad. Sci.*, USA, 1999, vol. 96, pp. 2907–2912.
[35] 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.
[36] T. Thattai and A. van Oudenaarden, "Stochastic gene expression in fluctuating environments," in *Proc. Genetics Soc. Am.*, 2004, pp. 523–530.
[37] T. Tian and K. Burrage, "Stochastic neural network models for gene regulatory networks," in *Proc. 2003 IEEE Congr. Evol. Comput.*, pp. 162–169.
[38] E. Weinstein, A. V. Oppenheim, M. Feder, and J. R. Buck, "Iterative and sequential algorithms for multisensor signal enhancement," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 846–859, Apr. 1994.
[39] L. F. Wessels, E. P. van Someren, and M. J. Reinders, "A comparison of genetic network models," in *Proc. Pacific Symp. Biocomput.*, 2001, pp. 508–519.
[40] F. Wu, W. Zhang, and A. J. Kusalik, "Modeling gene expression from microarray expression data with state-space equations," in *Proc. Pacific Symp. Biocomput.*, Hawaii Island, HI, 2004, pp. 581–592.
[41] 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.
[42] I. Ziskind and D. Hertz, "Maximum-likelihood localization of narrowband autoregressive sources via the EM algorithm," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2719–2724, Aug. 1993.

Zidong Wang (M'03–SM'04) was born in Jiangsu, China, in 1966. He received the B.Sc. degree in mathematics from Suzhou University, Suzhou, China, in 1986, the M.Sc. degree in applied mathematics in 1990, and the Ph.D. degree in electrical and computer engineering in 1994, from Nanjing University of Science and Technology, Nanjing, China.

He was with Nanjing University of Science and Technology earlier as a Lecturer and later as an Associate Professor. During 1997–1998, he was an Alexander von Humboldt Research Fellow at the Control Engineering Laboratory, Ruhr-University, Bochum, Germany. During 1999–2001, he was a Lecturer at the Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany. During 2001–2002, he was a University Senior Research Fellow at the School of Mathematical and Information Sciences, Coventry University, Coventry, U.K. Since 2002, he has been with the Department of Information Systems and Computing, Brunel University, Uxbridge, U.K., earlier as a Lecturer and Reader, and currently, as a Professor. He is an Adjunct TePin Professor at Donghua University, Shanghai, China, a Visiting Professor at Fuzhou University, Fuzhou, China, and a Guest Professor at Nanjing Normal University, Nanjing. He is the author or coauthor of more than 80 papers published in international journals. His current research interests include dynamical systems, signal processing, bioinformatics, control theory, and applications.

Dr. Wang is a Fellow of the Royal Statistical Society, a member of program committee for many international conferences, and a Reviewer for many international journals. He is an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C, the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, the *Circuits, Systems and Signal Processing*, an Action Editor for the *Neural Networks*, an Editorial Board Member for the *International Journal of Systems Science*, the *Neurocomputing*, the *International Journal of Computer Mathematics*, the *International Journal of General Systems*, and an Associate Editor on the Conference Editorial Board for the IEEE Control Systems Society. He was awarded the Humboldt Research Fellowship in 1996 from Alexander von Humboldt Foundation, the Japan Society for the Promotion of Science (JSPS) Research Fellowship in 1998 from the JSPT, and the William Mong Visiting Research Fellowship in 2002 from the University of Hong Kong. He was the recipient of the Outstanding Science and Technology Development Awards, from the National Education Committee of China, once in 2005 and twice in 1997, from the Military Industry General Company of China in 1997, from Jiangsu Province Government of China in 1997, and from the National Education Ministry of China in 1999. He was nominated an Appreciated Reviewer for the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 2006 and 2007, an Outstanding Reviewer for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL in 2004, and the *Automatica* in 2000. He was the Standing Member of the Technical Committee on Control of International Association of Science and Technology for Development in 2000. From 2001 to 2007, he was a member of technical program committee for 35 international conferences.

Fuwen Yang (SM'04) received the B.Sc. degree in electrical engineering from Anhui University of Science and Technology, Huainan, China, in 1983, and the M.Sc. and Ph.D. degrees in control engineering from Huazhong University of Science and Technology, Wuhan, China, in 1986 and 1990, respectively.

From 1986 to 1987, he was with Fujian Institute of Electronic Technique as a Research Engineer. Since 1990, he has been with the Department of Electrical Engineering, Fuzhou University, China, where he is currently a Professor of Control Engineering. He has held research positions at King's College London, Brunel University, Uxbridge, U.K., the University of Manchester, Manchester, U.K., and the University of Hong Kong, Hong Kong. He is the author or coauthor of more than 90 journal and conference papers. His current research interests include H_∞ control and filtering, iterative learning control, nonfragile control, signal processing, industrial real-time control, and power electronics.

Dr. Yang was the recipient of the Teaching Excellence Award for Young Teachers in 1995 from Fok Ying Tung Education Foundation, China, three Science and Technology Development Awards in 1996, 1999, and 2002, from Fujian Province of China, and two Youth Science and Technology Awards both in 1998 from Fujian Province of China and from Yun Sheng Foundation. He was listed in the Ten Outstanding Youth from Fujian Province of China in 1999, and was honored the May 4 Youth Medal from Fujian Province of China in 2000. Since 2001, he has been acting as a State Consultant of the People's Republic of China.

Daniel W. C. Ho (SM'04) received the B.Sc., M.Sc., and Ph.D. degrees in mathematics from the University of Salford, Salford, U.K., in 1980, 1982, and 1986, respectively.

From 1985 to 1988, he was with the Industrial Control Unit, University of Strathclyde, Glasgow, U.K., as a Research Fellow. Since 1989, he has been with the Department of Mathematics, City University of Hong Kong, Hong Kong, where he is currently a Professor. His current research interests include H_∞ control theory, robust pole assignment problem, adaptive neural wavelet identification, and nonlinear control theory.

Dr. Ho is an Associate Editor of the *Asian Journal of Control*.

Stephen Swift received the B.Sc. degree in mathematics and computing from the University of Kent, Canterbury, U.K., in 1991, the M.Sc. degree in artificial intelligence from Cranfield University, Cranfield, U.K., and the Ph.D. degree in intelligent data analysis from Birkbeck College, University of London, London, U.K. in 1993 and 2002, respectively.

He is currently with the School of Information Systems, Computing, and Mathematics, Brunel University, Uxbridge, U.K., as a Research Lecturer. He was a Web Designer, a Programmer, a Technical Architect, and a Postdoctoral Research Fellow. His current research interests include multivariate time-series analysis, heuristic search, data clustering, and evolutionary computation.

Allan Tucker received the B.Sc. degree in cognitive science from the University of Sheffield, Sheffield, U.K., in 1996, and the Ph.D. degree in computer science from Birkbeck College, University of London, London, U.K., in 2001.

He is currently with the Department of Information Systems and Computing, Brunel University, Uxbridge, U.K., a Research Lecturer. His current research interests include machine learning, Bayesian networks, bioinformatics, and medical informatics.

Xiaohui Liu received the B.Eng. degree in computing from Hohai University, Nanjing, China, in 1982 and the Ph.D. degree in computer science from Heriot-Watt University, Edinburg, U.K., in 1988.

He is currently with the Department of Information Systems and Computing, Brunel University, Uxbridge, U.K., as a Professor, where he leads the Intelligent Data Analysis (IDA) Group, engaging in research on artificial intelligence, dynamic systems, image and signal processing, and statistics.

Prof. Liu is on editorial boards of four computing journals, founded the biennial international conference series on IDA in 1995.