Estimating Time-Dependent Gene Networks from Time Series Microarray Data by Dynamic Linear Models with Markov Switching

Ryo Yoshida*
Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku, Tokyo, 103-8569, Japan
yoshidar@ism.ac.jp

Seiya Imoto
Human Genome Center, Institute of Medical Science, University of Tokyo,
4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan
imoto@ims.u-tokyo.ac.jp

Tomoyuki Higuchi
Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku, Tokyo, 103-8569, Japan
higuchi@ism.ac.jp

Abstract

In gene network estimation from time series microarray data, dynamic models such as differential equations and dynamic Bayesian networks assume that the network structure is stable through all time points, while the real network might changes its structure depending on time, affection of some shocks and so on. If the true network structure underlying the data changes at certain points, the fitting of the usual dynamic linear models fails to estimate the structure of gene network and we cannot obtain efficient information from data. To solve this problem, we propose a dynamic linear model with Markov switching for estimating time-dependent gene network structure from time series gene expression data. Using our proposed method, the network structure between genes and its change points are automatically estimated. We demonstrate the effectiveness of the proposed method through the analysis of Saccharomyces cerevisiae cell cycle time series data.

1. Introduction

For estimating gene networks from time series gene expression data measured by microarrays , a lot of attention has been focused on statistical methods, including Boolean networks [1, 11], differential equations [3, 5], dynamic Bayesian networks [6, 7, 8], state space models [2, 4] and so on. While these methods have provided many successful applications, a serious drawback for using these method to estimate gene networks remains to be solved: a basic assumption of these methods is that the network structure does not change through all time points, while the real gene network has time-dependent structure. In this paper, we give a solution of this problem and establish a statistical methodology to estimate gene networks with time-dependent structure by using dynamic linear models with Markov switching.

Our model is based on the linear state space model, also known as the dynamic linear model (DLM). In the DLM, the high-dimensional observation vector is compressed into the lower dimensional hidden state variable vector. For the microarray analysis, the observation vector corresponds to the gene expression value vector and the state variables can be considered as a transcriptional module [9] that is a set of co-regulated genes. Unlike Boolean networks, differential equations and dynamic Bayesian networks, we consider the dependency between these state variables in the DLM. Since microarrays contain much number of genes, the learning of Boolean networks and other network models is often infeasible. On the other hand, in the DLM, the network of the state variables gives a practical solution to understand gene regulatory networks based on the possible

* Current affiliation: Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan, yoshidar@ims.u-tokyo.ac.jp

transcriptional modules. Furthermore, by considering the canonical form of the DLM, it implicitly represents a network between genes by the linear system with the first-order Markov property.

Although, the DLM is advocated for analyzing high-dimensional time series gene expression data, this model also assume that the network structure is stable through the all time points. If the network structure changes drastically at certain points, the fitting of the DLM to the data should fail and we cannot obtain efficient information from the estimated model. To solve this problem, we use the dynamic linear models with Markov switching [12] (DLM-MS) that is an extension of the DLM to capture the change points of the data. In this approach, the dynamics of the system at a certain point is generated by one of possible regimes evolving according to a Markov process. The parameters in the DLM-MS are estimated by the Bayes approach based on the Gibbs sampling. Thus, we obtain the posterior distribution of each parameter that can be used for determining the network structure between genes. The number of switching points of the network structure and the number of hidden state variables are also automatically determined by the estimated prediction error.

The rest of this article is organized as follows: In Section 2, we present the time-dependent dynamic linear models and elucidate how we estimate a networks between genes. Section 3 describes the dynamic linear models with Markov switching. Section 4 will discuss the Bayesian estimation problem of DLM-MS, mainly, in terms of the computational aspect. Section 5 provides some analytic tools, including the determination of the number of regime switching and the dimension of state vectors, and the estimation of the transcriptional modules. In Section 6, the potential usefulness of our approach will be demonstrated with the application to *Saccharomyces cerevisiae* cell cycle time series data produced by Spellman *et al.* [13], where a part of data is synthesized to have a switching structure. Finally, the concluding remarks are given in Section 7.

2. Dynamic Linear Model

Let $\boldsymbol{y}_t$ be a vector of d observed random variables which contains expression values of d genes at time point t. The DLM relates a collection of $\boldsymbol{y}_t$, $t = 1, \cdots, T$, to the hidden k-dimensional state vector $\boldsymbol{x}_t$ in the following way:

$$\boldsymbol{y}_t = \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{w}_t. \tag{1}$$

Here, the $\boldsymbol{A}_t$ is a $d \times k$ measurement matrix and the $\boldsymbol{w}_t$ is the Gaussian white noise as $\boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{R}_t)$. Usually the dimension of state vector is taken to be much smaller than that of data, $k < d$. In DLM, the time evolution of the state variables are modeled by a first-order Markov process as

$$\boldsymbol{x}_t = \boldsymbol{B}_t \boldsymbol{x}_{t-1} + \boldsymbol{v}_t, \tag{2}$$

where $\boldsymbol{B}_t$ is $k \times k$ state transition matrix and the additive system noise follows form the Gaussian distribution as $\boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t)$. Throughout this article, the noise covariance matrices are assumed to be diagonal, $\boldsymbol{R}_t = \mathrm{diag}\{r_{1t}, \cdots, r_{dt}\}$ and $\boldsymbol{Q}_t = \mathrm{diag}\{q_{1t}, \cdots, q_{kt}\}$, respectively. Notice that the model parameters $\{\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{R}_t, \boldsymbol{Q}_t\}$ depend on the time index. This implies that the underlying dynamics changes discontinuously at certain undetermined points in time.

The process of the DLM starts with an initial Gaussian state $\boldsymbol{x}_0$ that has mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. In DLM, the dynamics of $\boldsymbol{Y}_{(T)} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)$ and $\boldsymbol{X}_{(T)} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T)$ are governed by the joint probability distribution

$$p(\boldsymbol{X}_{(T)}, \boldsymbol{Y}_{(T)}) = p(\boldsymbol{x}_0) \prod_{t=1}^{T} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) p(\boldsymbol{y}_t|\boldsymbol{x}_t).$$

The all composition in this representation are the Gaussian density ϕ in which $p(\boldsymbol{x}_0) = \phi(\boldsymbol{x}_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \phi(\boldsymbol{x}_t; \boldsymbol{B}_t \boldsymbol{x}_{t-1}, \boldsymbol{Q}_t)$, and $p(\boldsymbol{y}_t|\boldsymbol{x}_t) = \phi(\boldsymbol{y}_t; \boldsymbol{A}_t \boldsymbol{x}_t, \boldsymbol{R}_t)$.

The DLM, in its canonical form, implicitly assumes an interesting casual relationship among the d variates (genes). To see this, consider the generalized singular value decomposition of $\boldsymbol{A}_t$, namely, $\boldsymbol{R}_t^{-1/2} \boldsymbol{A}_t = \boldsymbol{L}_t \boldsymbol{D}_t \boldsymbol{V}_t^{'}$ where $\boldsymbol{L}_t$ is a matrix of k orthogonal vectors of length d, the diagonal matrix $\boldsymbol{D}_t$ contains k singular values and $\boldsymbol{V}_t^{'}$ is a $k \times k$ orthogonal matrix. Multiplying the both terms in observed equation (1) by $\boldsymbol{A}_t^{+'} = \boldsymbol{V}_t \boldsymbol{D}_t^{-1} \boldsymbol{L}_t^{'}$ from the lefthand-side, one can obtain an expression as

$$\boldsymbol{A}_t^{+'} \boldsymbol{R}_t^{-1/2} (\boldsymbol{y}_t - \boldsymbol{w}_t) = \boldsymbol{x}_t.$$

The canonical variate $\boldsymbol{A}_t^{+'} \boldsymbol{R}_t^{-1/2} (\boldsymbol{y}_t - \boldsymbol{w}_t)$ is a linear mapping of d-dimensional data onto the subspace $\mathbf{R}^k$ after removing the effect of measurement noise. The matrix $\boldsymbol{A}_t^{+'}$ compresses the filtered data $\boldsymbol{R}_t^{-1/2} (\boldsymbol{y}_t - \boldsymbol{w}_t)$ into k modules in the state vector. If $(\boldsymbol{A}_t^{+'})_{ij}$ is positioned significantly far from zero, the j-th gene captures a large effect on the i-th module. In contrast, the influence of genes with the $(\boldsymbol{A}_t^{+'})_{ij}$ lying a region close to zero is removed.

Substituting the canonical variates $\boldsymbol{A}_t^{+'} \boldsymbol{R}_t^{-1/2} (\boldsymbol{y}_t - \boldsymbol{w}_t)$ into the system model (2) leads to a causal relationship between the k modules defined by

$$\boldsymbol{A}_t^{+'} \boldsymbol{R}_t^{-1/2} (\boldsymbol{y}_t - \boldsymbol{w}_t) = \boldsymbol{B}_t \boldsymbol{A}_{t-1}^{+'} \boldsymbol{R}_{t-1}^{-1/2} (\boldsymbol{y}_{t-1} - \boldsymbol{w}_{t-1}) + \boldsymbol{v}_t.$$

This canonical form of DLM characterizes the interaction between the previous modules to the current ones, that is, module-module interaction, where the state transition matrix $\boldsymbol{B}_t$ captures the intensity of interaction.

The DLM also retains the linear system for describing the gene regulatory network as

$$\boldsymbol{R}_t^{-1/2} \ (\boldsymbol{y}_t - \boldsymbol{w}_t) =$$

$$\boldsymbol{H}_t \boldsymbol{R}_{t-1}^{-1/2}(\boldsymbol{y}_{t-1} - \boldsymbol{w}_{t-1}) + \boldsymbol{R}_t^{-1/2}\boldsymbol{A}_t \boldsymbol{v}_t,$$

where the interaction matrices $\boldsymbol{H}_t$, $t = 1, \cdots, T$ are parameterized by

$$\boldsymbol{H}_t = \boldsymbol{R}_t^{-1/2}\boldsymbol{A}_t \boldsymbol{B}_t \boldsymbol{A}_{t-1}^{+'}.$$

The $\boldsymbol{H}_t$ governs the gene network from time point $t-1$ to t in the following way: once the k modules in the compressed data $\boldsymbol{A}_{t-1}^{+'}\boldsymbol{R}_{t-1}^{-1/2}(\boldsymbol{y}_{t-1} - \boldsymbol{w}_{t-1})$ are given, the modules at time t are constructed through the loading matrix $\boldsymbol{B}_t$, and then the updated k modules regulates the expression value of d genes with the measurement matrix $\boldsymbol{A}_t$.

To sum up, the time-dependent DLM describes the consecutive changes in module sets of genes, module-module interactions and gene-gene interactions with the underlying canonical form (see Figure 1). After learning $\boldsymbol{A}_t$, $\boldsymbol{B}_t$ and the projection matrix $\boldsymbol{A}_t^{+}$, we can identify the time-dependent network structure by testing whether or not these parameters lie in a region significantly far from zero. This problem amounts to the classical testing method or the bootstrap confidential intervals.

3. DLM with Markov Switching

The problem of modeling change in an evolving time series can be handled by incorporating the dynamics of some underlying model change discontinuously at certain undetermined points in time. In view of real biological system, the structural change might occur in smooth. To incorporate a reasonable switching structure, we employ the DLM-MS approach that assumes the $\boldsymbol{y}_t$ is generated by one of the G possible regimes evolving according to a Markov chain. In this context, the model parameters $\{\boldsymbol{A}_t, \boldsymbol{B}_t, \boldsymbol{R}_t, \boldsymbol{Q}_t\}$ are assumed to take one of the G possible configurations $\{\bar{\boldsymbol{A}}_g, \bar{\boldsymbol{B}}_g, \bar{\boldsymbol{R}}_g, \bar{\boldsymbol{Q}}_g\}$, $g = 1, \cdots, G$, at each time point. For notational convenience, we introduce the hidden vector of G class labels $(\boldsymbol{c}(t))_g = c_g(t)$ to indicate the configurations in the following way:

$$c_g(t) = \begin{cases} 1 & \boldsymbol{y}_t \in \text{regime } g \\ 0 & \text{otherwise.} \end{cases}$$

The DLM-MS, in its basic form, assumes that the discrete variable $\boldsymbol{c}(t)$ evolves according to the first-order Markov chain with the transition probability matrix $\boldsymbol{M}$ of order $G \times G$ where the (h, g) element defines a probability of event $\{\boldsymbol{y}_t \in \text{regime } g\} \cup \{\boldsymbol{y}_{t-1} \in \text{regime } h\}$, that is,

$$(\boldsymbol{M})_{hg} = \Pr(c_g(t) = 1 | c_h(t-1)).$$

Each row of $\boldsymbol{M}$, denoted by $\boldsymbol{m}_h$, is restricted to be $||\boldsymbol{m}_h||^2 = 1$. Smoothness of change in regimes are controlled by the entropy of $\boldsymbol{m_h}$ for $h = 1, \cdots, G$.

4. Bayesian Inference

For some gene expression data, each array contains some genes with fluorescence intensity measurements that were flagged by the experimenter and recorded as missing data points. In such a case, $\boldsymbol{y}_t$ is incomplete. To deal with the missing problem, we define the partition of d observed vector $\boldsymbol{y}_t = (\boldsymbol{y}_t^{o'}, \boldsymbol{y}_t^{m'})$ where $\boldsymbol{y}_t^o$ and $\boldsymbol{y}_t^m$ contain the observed and missing components, respectively. Consequently, the DLM-MS takes $\{\boldsymbol{C}_{(T)}, \boldsymbol{X}_{(T)}, \boldsymbol{Y}_{(T)}^m, \boldsymbol{Y}_{(T)}^o\}$ as a complete dataset having the joint distribution

$$\begin{aligned} p_\Theta(\boldsymbol{C}_{(T)}, \boldsymbol{X}_{(T)}, \boldsymbol{Y}_{(T)}) \; &= p(\boldsymbol{c}_0)p(\boldsymbol{x}_0)\prod_{t=1}^{T} p(\boldsymbol{c}_t|\boldsymbol{c}_{t-1}) \\ &\quad p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{c}_t)p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{c}_t). \end{aligned}$$

The parameters to be learned from the observed dataset are collected into a set $\boldsymbol{\Theta} = \{\bar{\boldsymbol{A}}_g, \bar{\boldsymbol{B}}_g, \bar{\boldsymbol{R}}_g, \bar{\boldsymbol{Q}}_g, \boldsymbol{M}\}_{g=1}^G$. The $p(\boldsymbol{x}_0)$ and $p(\boldsymbol{c}_0)$ denote the initial distributions to derive the dynamic system. Each composition in the above joint distribution is obvious, so the details are omitted here.

Our attention turns to the Bayesian learning of DLM-MS that requires the prior distribution of all model parameters $p(\boldsymbol{\Theta})$ and the initial distribution of the hidden states $p(\boldsymbol{x}_0)$ and $p(\boldsymbol{c}_0)$. In this study, we employ the natural conjugate priors. Let $\bar{\boldsymbol{a}}_{ig}$ and $\bar{\boldsymbol{b}}_{ig}$ be the i-th row of $\bar{\boldsymbol{A}}_g$ and $\bar{\boldsymbol{B}}_g$, respectively. A family of the conjugate priors of DLM-MS that we use are expressed as follows:

$$\begin{aligned} \bar{\boldsymbol{a}}_{ig} &\sim N_+(\boldsymbol{0}, \lambda_a \boldsymbol{I}), \; \forall i, g, \\ \bar{\boldsymbol{b}}_{ig} &\sim N(\boldsymbol{0}, \lambda_b \boldsymbol{I}), \; \forall i, g, \\ (\bar{\boldsymbol{R}}_g)_{ii} &\sim IG(\gamma_{r0}, \delta_{r0}), \; \forall i, g, \\ (\bar{\boldsymbol{Q}}_g)_{ii} &\sim IG(\gamma_{q0}, \delta_{q0}), \; \forall i, g, \\ \boldsymbol{m}_h &\sim Dir(u_1, \cdots, u_G), \; \forall h. \end{aligned}$$

where $IG(\gamma, \delta)$ stands for the inverse-gamma distribution with the shape γ and the scale parameter δ, and $Dir(u_1, \cdots, u_G)$ denotes the Dirichlet distribution with the prior sample size $u_1, \cdots, u_G$. Note that the prior distribution of $\bar{\boldsymbol{A}}_g$ is specified by the truncated Gaussian distribution $N_+(\boldsymbol{0}, \lambda_a \boldsymbol{I})$ whose support are restricted to the positive part $\boldsymbol{a}_{ig} \geq \boldsymbol{0}$. For DLM setting the underlying dynamical system is invariant under the transformations as $\boldsymbol{A}_t \to -\boldsymbol{A}_t$ and $\boldsymbol{x}_t \to -\boldsymbol{x}_t$. To avoid the lack of identifiability, we use the truncated prior distribution.

Once the prior distributions are given, the augmented parameters $\boldsymbol{\Theta}, \boldsymbol{X}_{(T)}, \boldsymbol{C}_{(T)}$, and $\boldsymbol{Y}_{(T)}^m$ are estimated through the posterior distribution

$$\begin{aligned} p(\boldsymbol{\Theta}, \; &\boldsymbol{X}_{(T)}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)}^m | \boldsymbol{Y}_{(T)}^o) \\ &\propto p_\Theta(\boldsymbol{C}_{(T)}, \boldsymbol{X}_{(T)}, \boldsymbol{Y}_{(T)}) \; p(\boldsymbol{\Theta}). \end{aligned}$$

Within Bayesian framework, all inferences are made based on the marginal posterior distribution, for instance,

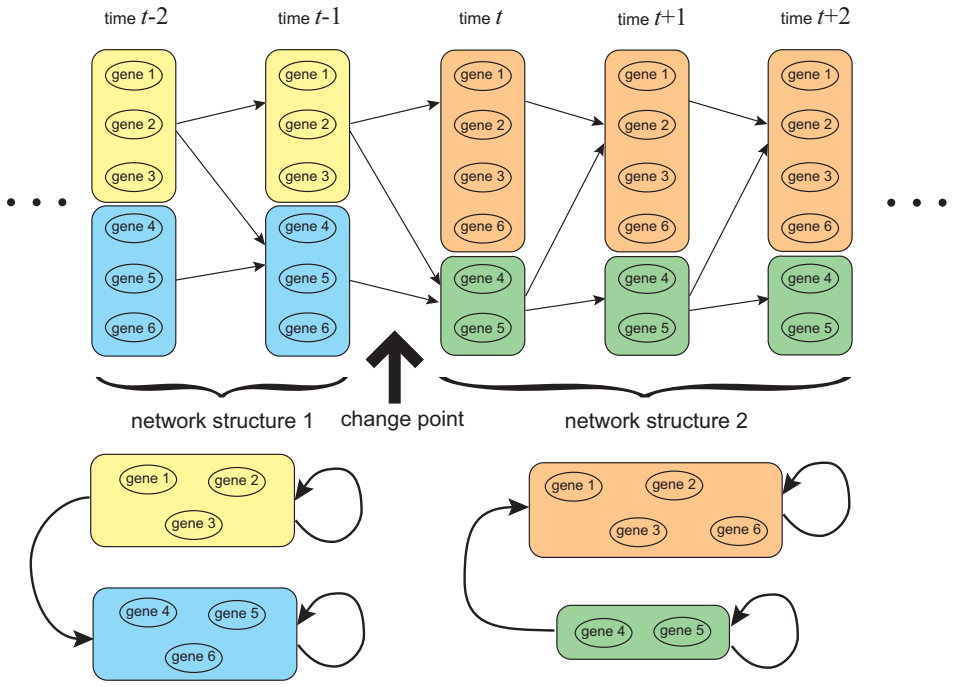

Figure 1. Schematic expression of time-dependent module-module networks represented by DLM-MS. Change in regime occurs at time point t. The interactions of the two transcriptional modules between previous and current time point in the first regime are different from those in the second regime.

$p(\boldsymbol{\Theta}|\boldsymbol{Y}^o_{(T)})$, and the goal is to characterize the marginal distributions by using some quantities, e.g. the posterior mean $\hat{\boldsymbol{\Theta}} = E(\boldsymbol{\Theta}|\boldsymbol{Y}^o_{(T)})$ and the maximum a posteriori estimator $\hat{\boldsymbol{\Theta}} = \mathrm{argmax}_{\Theta} p(\boldsymbol{\Theta}|\boldsymbol{Y}^o_{(T)})$ and so on. The direct evaluation of these quantities is, however, difficult under the DLM-MS setting. To overcome such intractability, we perform the Gibbs sampling algorithm that approximately computes the posterior quantities of interest by using simulated random draws from the posterior distributions. The Gibbs sampling is alternating conditional sampling which is defined in terms of subvector of $\boldsymbol{\Theta}, \boldsymbol{X}_{(T)}, \boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}^m_{(T)}$. Each iteration of the sampling scheme cycles through the subvector of drawing each subset conditional on the value of all the other. With arbitrary starting values $\boldsymbol{\Theta}^0$, $\boldsymbol{X}^0_{(T)}$, $\boldsymbol{C}^0_{(T)}$ and $\boldsymbol{Y}^{m0}_{(T)}$, it proceeds by successive iteration of the following eight steps:

1. Generate $\boldsymbol{X}_{(T)}$ conditional on $\boldsymbol{\Theta}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$.
2. Generate $\bar{\boldsymbol{A}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{A}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ for $g = 1, \cdots, G$.
3. Generate $\bar{\boldsymbol{R}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{R}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ for $g = 1, \cdots, G$.
4. Generate $\bar{\boldsymbol{B}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{B}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ for $g = 1, \cdots, G$.
5. Generate $\bar{\boldsymbol{Q}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{Q}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$for $g = 1, \cdots, G$.
6. Generate $\boldsymbol{m}_h$ conditional on $\boldsymbol{\Theta}_{-m_h}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ for $h = 1, \cdots, G$.
7. Generate $\boldsymbol{C}_{(T)}$ conditional on $\boldsymbol{\Theta}$, $\boldsymbol{X}_{(T)}$ and $\boldsymbol{Y}_{(T)}$.
8. Generate $\boldsymbol{Y}^m_{(T)}$ conditional on $\boldsymbol{\Theta}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}^o_{(T)}$.

Here, $\boldsymbol{\Theta}_{-Z}$ stands for all components of $\boldsymbol{\Theta}$, except for $\boldsymbol{Z}$, at their current values. The Markov structure of DLM-MS and the assumption of conjugate priors makes it easy to draw sample from the full-conditional distribution, for example, the functional form of $p(\boldsymbol{X}_{(T)}|\boldsymbol{\Theta}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)})$ is Gaussian where the mean and covariance matrix are successively computed by the well-known Kalman filter and smoother, and $p(\bar{\boldsymbol{a}}_{ig}|\boldsymbol{\Theta}_{-\bar{A}}, \boldsymbol{X}_{(T)}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)})$ is the positive part Gaussian distribution where its parameters are determined by the conventional rule based on the natural conjugate prior. The above steps are detailed in Appendix. The method proceeds by alternatively sampling

from these full-conditional distributions. If the iteration have proceeded long enough, the simulations is grossly representative of the target distribution. To diminish the effect of the starting point, we generally discard the first p simulated samples and focus attention on the rest $n-p$. The set $\{\boldsymbol{\Theta}_j, \boldsymbol{X}_{(T)j}, \boldsymbol{C}_{(T)j}, \boldsymbol{Y}^m_{(T)j}\}^n_{j=p+1}$ is used to summarize the posterior distribution and to compute quantiles, and the other summaries of interest as needed.

5. Implementations

A basic issue arising in the DLM-MS approach is the determination of the number of regimes in the switching system, G, and the number of modules k. We address this problem by selecting a particular combination $\{G^*, k^*\}$ to attain the best predictive ability.

To this end, we firstly construct B set of the bootstrap samples in which B vectors of the quasi-missing observations, $\{\boldsymbol{y}^m_b\}^B_{b=1}$, is generated by resampling of all elements in $\boldsymbol{Y}^o_{(T)}$ with probability α. One intuitive approach is to select a combination $\{G^*, k^*\}$ to minimize the prediction error

$$\mathrm{Err}(G,k) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{L_b}||\boldsymbol{y}^m_b - \hat{\boldsymbol{y}}^m_b(G,k)||^2, \qquad (3)$$

where the L_b is the number of the quasi-missing observations contained in the b-th bootstrap set and the $\hat{\boldsymbol{y}}^m_b(G,k)$ stands for the corresponding posterior mean computed by the Monte Carlo samples.

In DLM-MS approach, the existing regimes are deduced from the estimated posterior distribution of the class labels. The Bayes rule explores the G regimes by assigning each time point t to a particular regime as follows:

$$\hat{c}_g(t) = \begin{cases} 1 & \text{if } g = \underset{h\in\{1,\cdots,G\}}{\mathrm{argmax}} \sum_{j=p+1}^{n} c_h(t)_j, \\ 0 & \text{otherwise,} \end{cases}$$

where $\{c_h(t)_j\}^n_{j=p+1}$ is the simulated draws generated by the Gibbs sampling.

Once the model parameters are estimated, the DLM-MS approach offers a set of consecutive k modules $\boldsymbol{A}^{+'}_t \boldsymbol{R}^{-1/2}_t(\boldsymbol{y}_t - \boldsymbol{w}_t)$ along with the time line $t = 1, \cdots, T$, and also their estimated networks. Interpretation of the k coordinates corresponding to the estimated modules is important for real data analysis. This task can be addressed by investigating the direction of projection matrix $\boldsymbol{A}^{+'}_t = \bar{\boldsymbol{A}}^{+'}_g$ that projects $\boldsymbol{y}_t$ onto $\mathbf{R}^k$. In practice, it will be helpful to list the top L genes to attain the highest positive score of $(\bar{\boldsymbol{A}}^+_g)_{ij}$ at Ω^+_{ig} and the highest negative score at Ω^-_{ig} for $j = 1, \cdots, k$ and $g = 1, \cdots, G$. These $2kG$ sets can be useful either to visualize the calibrated networks and also to elucidate a causal link from the estimated networks to some biological resources.

6. Computational Experiments

We demonstrate our proposed method through the analysis of *Saccharomyces cerevisiae* cell cycle time series data collected by Spellman *et al.* [13]. Although the cell cycle dataset contains two short time series data and four medium time series data, we use cdc15 time series data (24 time points). Originally, 800 genes were identified as the cell cycle-related genes by Spellman *et al*. [13]. From these 800 genes, 43 genes are also complied in the cell cycle pathway in KEGG. Therefore, we use these 43 genes and estimate the network of these genes in this analysis. While the 24 time course data were collected at unequal time intervals, the sampling time points are assumed to be equally spaced within our DLM framework. Without loss of generality, the DLM approach can incorporate the problem of unequally-spaced time points into the estimation procedure.

First, to select an optimal number of regimes G and the number of modules, i.e. the dimension of state variable k, we use diffuse prior distributions for all candidate models as follows: λ_a, $\lambda_b = 20$, γ_{r0}, $\gamma_{q0} = 10$, δ_{r0}, $\delta_{q0} = 10$, $u_g = 1$ for $g = 1, \cdots, G$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\Sigma}_0 = \mathrm{diag}\{10, \cdots, 10\}$ and $\Pr(c_g(0) = 1|\mathcal{I}_0) = 1/G$ for $g = 1, \cdots, G$. In the Gibbs sampling algorithm, the number of discarded draws is fixed at $p = 250000$, and total $n - p = 50000$ samples are used to compute the posterior quantities.

After fitting a variety of models ranging $G = 1, \cdots, 3$ and $k = 1, \cdots, 6$, the model of $G = 1$ and $k = 5$ was judged to be optimal by using the 10-fold cross validation criterion (3). Our proposed method provided no evidence for the presence of the regime switching. However, this result is not desirable for demonstrating the performance of our proposed method and its applicability.

We therefore decided to construct a quasi-cell cycle microarray data which are synthesized to capture a switching structure. Data fabrication that we enforced are summarized in both Figure 2 and below:

For $t = 11, \cdots, 17$, the expression values of 43 genes are interpolated in the following way:

1. *Module 1*

$$\begin{aligned} \text{if } i = 1, 2 \quad & y_{i_t} = -0.4y_{1_{t-1}} + 0.6y_{2_{t-1}} + \nu_t, \\ \text{if } 3 \le i \le 15 \quad & y_{i_t} = -0.5y_{1_{t-1}} + 0.6y_{2_{t-1}} \\ & \quad -0.3y_{31_{t-1}} - 0.8y_{32_{t-1}} + \nu_t, \end{aligned}$$

2. *Module 2*

$$\begin{aligned} \text{if } i = 16 \quad & y_{i_t} = -0.4y_{16_{t-1}} + \nu_t \\ \text{if } 17 \le i \le 30 \quad & y_{i_t} = 0.7y_{16_{t-1}} - 0.6y_{2_{t-1}} + \nu_t \end{aligned}$$

3. *Module 3*

$$\begin{aligned}\text{if } i = 31, 32 \quad y_{i_t} &= 0.6y_{15_{t-1}} + 0.7y_{31_{t-1}} \\ &\quad -0.4y_{32_{t-1}} + \nu_t \\ \text{if } 33 \leq i \leq 43 \quad y_{i_t} &= 0.6y_{15_{t-1}} + 0.5y_{i_{t-1}} \\ &\quad +0.2y_{31_{t-1}} + 0.4y_{32_{t-1}} + \nu_t\end{aligned}$$

The order of genes, i.e. $i = 1, \cdots, 43$, follows that of Figure 2.

In this synthesized regime, three quasi-modules are regulated by each other, as module 1 causes module 1 itself and module 2, module 2 causes module 2 and 3, and module 3 regulates module 1 and 2, respectively. Primitive genes that drive dynamics in this regime are comprised of y_{1t}, y_{2t}, y_{15t}, y_{16t}, y_{31t} and y_{32t}. Figure 3 shows a schematic expression of data synthesis and the resulting expression patterns. The member of each module is also listed at there.

Among a range of candidate models, cross validation criterion attains the best score at the number of switching points $G = 2$ and the number of hidden modules $k = 3$ that is consistent with the existing data structure. Figure 4 summarizes the time evolution of the estimated gene-gene interaction matrix $\boldsymbol{H}_t = \boldsymbol{R}^{-1/2}\boldsymbol{A}_{t+}\boldsymbol{B}_t\boldsymbol{A}_{t-1}^{'}$ for $t = 2, \cdots, T$ where the coefficients are computed by averaging the Monte Carlo samples. Change in the regime from $t = 11, \cdots, 17$ is clearly detected while the estimated interaction in the other regime are stable through the evolving times. Visualizing time-dependent interaction matrices must be very helpful for understanding the switching structure and finding the time points of variation.

Figure 3 displays sets of module transcriptional genes listed at Ω_{ig}^{+} and Ω_{ig}^{-} for $i = 1, \cdots, 3$ and $g = 1, \cdots, 2$ in which shown here is a part of the selected genes in each set, and the estimated module-module interactions are also presented. The existing two change points were correctly estimated. In the synthesized regime, the calibrated 6 modules are likely to reflect the quasi-three modules as all members listed at a set belong to one quasi-module. The estimated interactions is also consistent to the true data structure.

7. Discussion

We focused on a time-dependent DLM to deal with structural change of biological system in gene expression. As was elucidated in this paper, the DLM, in its canonical form, implicitly represents gene-gene interaction via module-module interaction. The time-dependent DLM assumes that these interactions change over time. This assumption is natural in terms of real gene expression process, but the occurrence of structural change must be smooth. To incorporate smoothness, we proposed use of the DLM-MS that represents change in regime evolving according to the first-order Markov process. We established some analytic tools associated with DLM-MS; the Bayesian parameter estimation based on the Gibbs sampling algorithm; the cross validation approach for the determination of the number of switching time points and the number of module transcriptionals; visualization technique for the evolving gene-gene interactions and the module-module interaction. We demonstrated its potential usefulness with the application to *Saccharomyces cerevisiae* cell cycle time course data where a part of data is synthesized as to have a switching structure in the gene network.

The gene regulation system stated by the our proposed model is an autonomous process that does not depend on any external variables e.g. proyteins, metabolites and so on. This fundamental assumption might be quite questionable. Within our framework, a dependence of the gene regulation on the external system can be included by incorporating the utilizable driving inputs into the observational equation or the system model.

The Bayesian parameter estimation gives a scope to the overfitting problem occurred due to small sample size and a way of incorporating the biological knowledge to the parameter estimation procedure. However, in this study, ambiguity in the determination of hyperparameters of the prior distributions is remained. In practice, we have to further explore the robustness of some estimates for any priors in the class. In a case where no prior knowledges are available, the hierarchical Bayes method must be useful to avoid such ambiguity or to model relatively complicated situations. Alternatively, a family of noninformative priors, e.g. Jeffery's prior or uniform prior, is also incorporated into our method without loss of generality.

Although some tasks remain to be solved, we believe that our proposed method will provide the successful applications for gene network estimation problem.

Appendix: Gibbs Sampling for DLM-MS

The following steps explain details of the Gibbs sampling algorithm, given arbitrary starting values $\boldsymbol{\Theta}^0$, $\boldsymbol{X}_{(T)}^0$ and $\boldsymbol{C}_{(T)}^0$. Hereafter, we will use the following notations:

$$\boldsymbol{x}_{t|s} = E(\boldsymbol{x}_t|\mathcal{I}_s)$$

and

$$\boldsymbol{F}_{t|s} = E\Big[(\boldsymbol{x} - \boldsymbol{x}_{t|s})(\boldsymbol{x} - \boldsymbol{x}_{t|s})^{'}\Big|\mathcal{I}_s\Big],$$

where the set $\mathcal{I}_s$ contains all information up to time point s.

1. Generate $\boldsymbol{X}_{(T)}$ conditional on $\boldsymbol{\Theta}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ according to

$$p(\boldsymbol{X}_{(T)}|\boldsymbol{\Theta}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)}) \propto p(\boldsymbol{x}_T|\boldsymbol{\Theta}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)}) \prod_{t=1}^{T-1} p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \boldsymbol{\Theta}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)}).$$

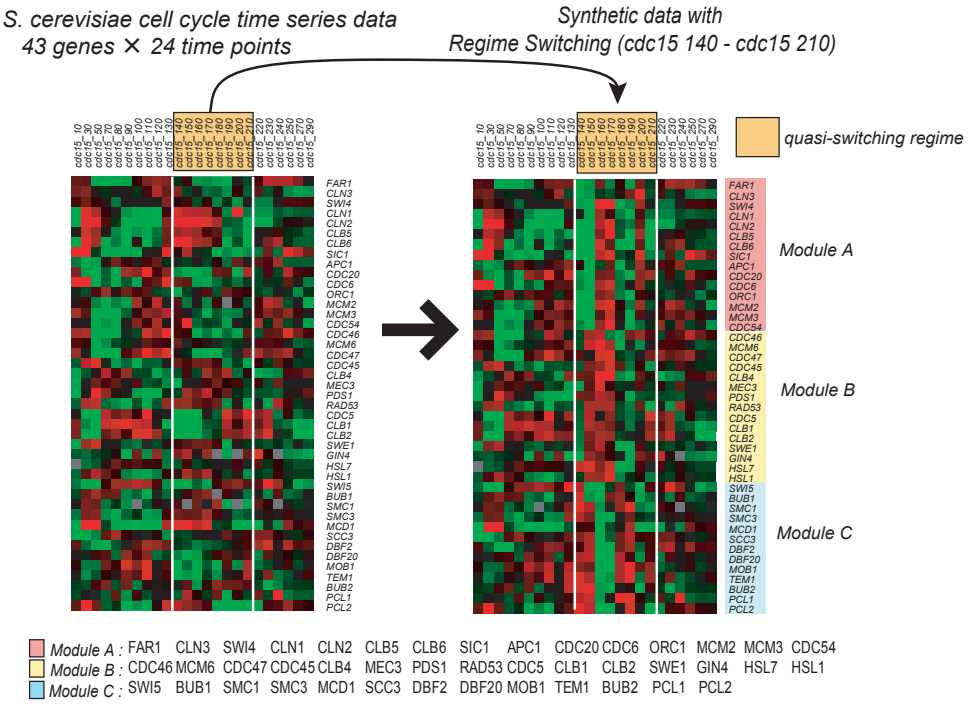

Figure 2. Fabrication of synthetic data: Original dataset contains the 43 gene expression values *Saccharomyces cerevisiae* **cell cycle measured at 24 time points. In the regime from $t = 11$ to 17, the quasi-expression values are interpolated by following a time series model. The synthetic gene expression organizes the three transcriptional modules.**

Note that $p(\boldsymbol{x}_T|\boldsymbol{\Theta}, \boldsymbol{C}_{(T)}, \boldsymbol{Y}_{(T)})$ takes in the form of Gaussian which is equivalent to the filtering distribution corresponding to the conventional linear state space model. Hence the computation of the mean and the covariance matrix is accomplished via the Kalman filter.

(a) With initial conditions $\boldsymbol{x}_{0|0} = \boldsymbol{\mu}_0$ and $\boldsymbol{F}_{0|0} = \boldsymbol{\Sigma}_0$, run the Kalman filter algorithm for $t = 1, \cdots, T$:

i. (Prediction)

$$\begin{aligned} \boldsymbol{x}_{t|t-1} &= \boldsymbol{B}_t \boldsymbol{x}_{t-1|t-1}, \\ \boldsymbol{F}_{t|t-1} &= \boldsymbol{B}_t \boldsymbol{F}_{t-1|t-1} \boldsymbol{B}_t^{'} + \boldsymbol{Q}_t, \end{aligned}$$

ii. (Filtering)

$$\begin{aligned} \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{A}_t \boldsymbol{x}_{t|t-1}), \\ \boldsymbol{F}_{t|t} &= (\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{A}_t) \boldsymbol{F}_{t|t-1}, \\ \boldsymbol{K}_t &= \boldsymbol{R}_t^{-1} \boldsymbol{A}_t (\boldsymbol{A}_t^{'} \boldsymbol{R}_t^{-1} \boldsymbol{A}_t + \boldsymbol{F}_{t|t-1}^{-1})^{-1}. \end{aligned}$$

(b) The last iteration of the Kalman filter provide us with $\boldsymbol{x}_{T|T}$ and $\boldsymbol{F}_{T|T}$, and $\boldsymbol{x}_T$ can be generated from

$$\boldsymbol{x}_T \sim N(\boldsymbol{x}_{T|T}, \boldsymbol{F}_{T|T}).$$

(c) For $t = T-1, T-2, \cdots, 1$, generate repeatedly $\boldsymbol{x}_t$ according to

$$\boldsymbol{x}_t \sim N(\boldsymbol{m}_t, \boldsymbol{P}_t)$$

where the mean and the covariance matrix are computed by the updating equation

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{x}_{t|t} + \boldsymbol{G}_{t+1}(\boldsymbol{x}_{t+1} - \boldsymbol{B}_{t+1} \boldsymbol{x}_{t|t}), \\ \boldsymbol{P}_t &= (\boldsymbol{I} - \boldsymbol{G}_{t+1} \boldsymbol{B}_{t+1}) \boldsymbol{F}_{t|t}, \end{aligned}$$

with

$$\boldsymbol{G}_{t+1} = \boldsymbol{F}_{t|t} \boldsymbol{B}_{t+1}^{'} (\boldsymbol{B}_{t+1} \boldsymbol{F}_{t|t} \boldsymbol{B}_{t+1}^{'} + \boldsymbol{Q}_{t+1})^{-1}.$$

2. For $g = 1, \cdots, G$, generate $\bar{\boldsymbol{A}}_g^{'} = (\bar{\boldsymbol{a}}_{1g} \cdots \bar{\boldsymbol{a}}_{dg})$ conditional on $\boldsymbol{\Theta}_{-\bar{A}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$. in the following way:

$$\bar{\boldsymbol{a}}_{ig} \sim N(\boldsymbol{\eta}_{ig}, \boldsymbol{\Psi}_{ig}),$$

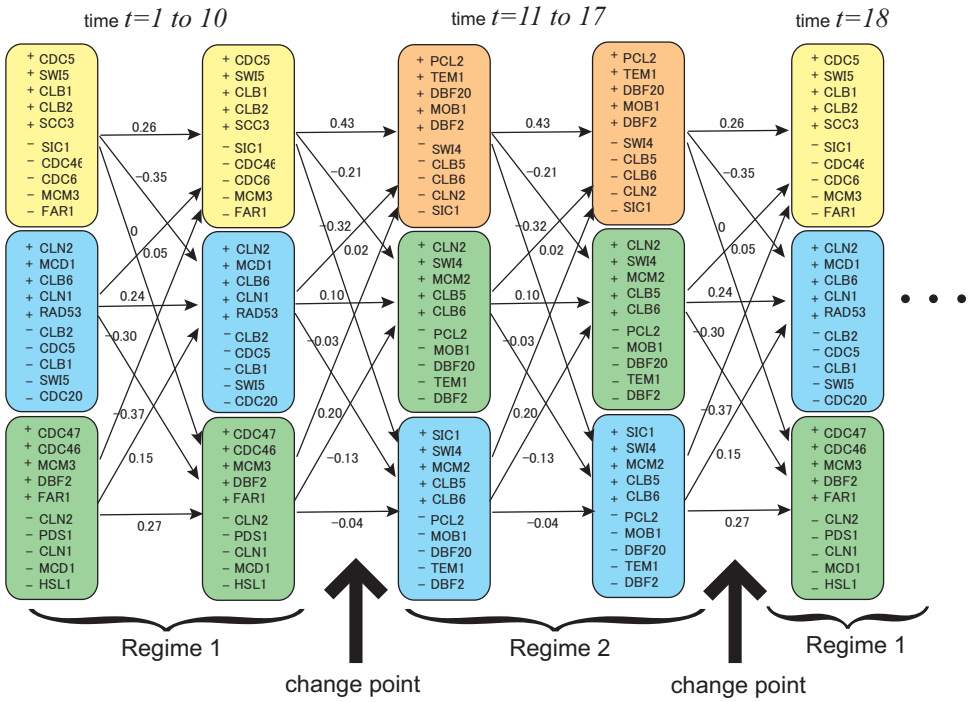

Figure 3. The calibrated module genes listed at Ω_{ig}^{+} and Ω_{ig}^{-} for $i = 1, \cdots, 3$ and $g = 1, \cdots, 2$ and the module-module interactions. Switching time points are estimated as $t = 11$ and $t = 18$. Gene names prefixed with $+$ and $-$ were listed at Ω_{ig}^{+} and Ω_{ig}^{-}, respectively. Each score represents the intensity of interaction between two modules.

where the mean and the covariance matrix are computed by

$$\begin{aligned} \boldsymbol{\eta}_{ig} &= \left(\frac{\lambda_a}{\bar{r}_{ig}}\boldsymbol{I} + \boldsymbol{X}_g^{'}\boldsymbol{X}_g\right)^{-1}(\boldsymbol{X}_g^{'}\boldsymbol{y}_{ig}) \\ \boldsymbol{\Psi}_{ig} &= (\lambda_a \boldsymbol{I} + \bar{r}_{ig}^{-1}\boldsymbol{X}_g^{'}\boldsymbol{X}_g)^{-1}. \end{aligned}$$

for $i = 1, \cdots, d$. Here the vector $\boldsymbol{y}_{ig}$ contains the i-th gene expression value $(\boldsymbol{y}_t)_i$ belonging to the g-th regime, that is, having the current class label $c_g(t) = 1$. Each row of the design matrix $\boldsymbol{X}_g$ is the current state vector $\boldsymbol{x}_t$ having $c_g(t) = 1$, where the number of rows is equal to $\{\text{num. of time points} \in \text{regime } g\} = \sum_{t=1}^{T} c_g(t)$.

3. For $g = 1, \cdots, G$, generate $\bar{\boldsymbol{R}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{R}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ according to

$$r_{ig} \sim IG(\gamma_{r1}, \delta_{r1})$$

where

$$\gamma_{r1} = \gamma_{r0} + \sum_{t=1}^{T} c_g(t),$$

$$\delta_{r1} = \delta_{r0} + ||\boldsymbol{y}_{ig} - \boldsymbol{X}_g \bar{\boldsymbol{a}}_{ig}||^2,$$

for $i = 1, \cdots, d$.

4. For $g = 1, \cdots, G$, generate $\bar{\boldsymbol{B}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{B}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ in the following way:

$$\bar{\boldsymbol{b}}_{ig} \sim N(\boldsymbol{\xi}_{ig}, \boldsymbol{\Phi}_{ig}),$$

where the mean and the covariance matrix are computed by

$$\begin{aligned} \boldsymbol{\xi}_{ig} &= \left(\frac{\lambda_b}{\bar{q}_{ig}}\boldsymbol{I} + \boldsymbol{S}_g^{'}\boldsymbol{S}_g\right)^{-1}(\boldsymbol{S}_g^{'}\boldsymbol{x}_{ig}) \\ \boldsymbol{\Phi}_{ig} &= (\lambda_b \boldsymbol{I} + \bar{q}_{ig}^{-1}\boldsymbol{S}_g^{'}\boldsymbol{S}_g)^{-1}. \end{aligned}$$

Here the response vector $\boldsymbol{x}_{ig}$ contains the i-th element of $\boldsymbol{x}_t$ having the current label $c_g(t) = 1$, and thus the length $\sum_{t=1}^{T} c_g(t)$. Each row of the design matrix $\boldsymbol{S}_g$ consists of the the corresponding inputs vector $\boldsymbol{x}_{t-1}$ in the system model.

5. Generate $\bar{\boldsymbol{Q}}_g$ conditional on $\boldsymbol{\Theta}_{-\bar{Q}_g}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ according to

$$\bar{q}_{ik} \sim IG(\gamma_{q1}, \delta_{q1})$$

where

$$\gamma_{q1} = \gamma_{q0} + \sum_{t=1}^{T} c_g(t),$$
$$\delta_{q1} = \delta_{q0} + ||\boldsymbol{x}_{ig} - \boldsymbol{S}_g \bar{\boldsymbol{b}}_{ig}||^2,$$

for $i = 1, \cdots, k$.

6. Generate $\boldsymbol{m}_h$ conditional on $\boldsymbol{\Theta}_{-m}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ as

$$\boldsymbol{m}_h \sim Dir(n_{h1} + u_1, n_{h2} + u_2, \cdots, n_{hG} + u_G),$$

where n_{hg} stands for the number of samples having $c_g(t) = 1$ and $c_h(t-1) = 1$.

7. Generate $\boldsymbol{C}_{(T)}$ conditional on $\boldsymbol{\Theta}$, $\boldsymbol{X}_{(T)}$ and $\boldsymbol{Y}_{(T)}$ according to

$$p(\boldsymbol{C}_{(T)}|\boldsymbol{\Theta}, \boldsymbol{X}_{(T)}, \boldsymbol{Y}_{(T)}) \propto p(\boldsymbol{c}_{t+1}|\boldsymbol{c}_t, \boldsymbol{\Theta}) \, p(\boldsymbol{c}_t|\boldsymbol{\Theta}, \boldsymbol{X}_{(t)}, \boldsymbol{Y}_{(t)}).$$

To this end, the following steps can be employed:

Starting from an initial distribution $p(\boldsymbol{c}_0|\mathcal{I}_0) = 1/G$, run the filtering algorithm to calculate $p(\boldsymbol{c}_t|\mathcal{I}_t)$, $t = 1, \cdots, T$ where $\mathcal{I}_t = \{\boldsymbol{\Theta}, \boldsymbol{X}_{(t)}, \boldsymbol{Y}_{(t)}\}$ as,

(a) Given $p(\boldsymbol{c}_{t-1}|\mathcal{I}_{t-1})$ at the beginning of time t, the $\Pr(c_g(t) = 1, c_h(t-1) = 1|\mathcal{I}_{t-1})$ are calculated by

$$\Pr(\, c_g(t) = 1, c_h(t-1) = 1|\mathcal{I}_{t-1}) = \Pr(c_k(t){=}1|c_h(t{-}1){=}1)\Pr(c_h(t{-}1)|\mathcal{I}_{t-1}).$$

(b) Once $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ is observed at the end of time t, we can update the filtered probability as

$$\Pr(c_g(t){=}\, 1 \,|\mathcal{I}_t) = \sum_{h=1}^{G} \Pr(c_g(t){=}1, c_h(t{-}1){=}1|\mathcal{I}_t),$$

where

$$\Pr\left(c_g(t) = 1, c_h(t-1) = 1|\mathcal{I}_t\right) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t, c_g(t) = 1) p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, c_g(t) = 1) \Pr(c_g(t){=}1, c_h(t-1){=}1|\mathcal{I}_{t-1}).$$

8. Generate $\boldsymbol{Y}^m_{(T)}$ conditional on $\boldsymbol{\Theta}$, $\boldsymbol{X}_{(T)}$, $\boldsymbol{C}_{(T)}$ and $\boldsymbol{Y}^o_{(T)}$ as

$$\boldsymbol{y}^m_t \sim N(\boldsymbol{A}^{(m)}_t \boldsymbol{x}_t, \boldsymbol{R}^{(m)}_t).$$

where $\boldsymbol{A}^{(m)}_t$ and $\boldsymbol{R}^{(m)}_t$ are, respectively, the partitioned measurement matrix and the covariance matrix of $\boldsymbol{A}_t$ and $\boldsymbol{R}_t$ corresponding to the missing parts.

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. *Pac. Symp. Biocomput.*, **4**, 17–28, 1999.

[2] 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*, **21**(3), 349–356, 2005.

[3] T. Chen, H. He and G. Church. Modeling gene expression with differential equations. *Pacific Symposium on Biocomputing*, **4**, 29–40, 1999.

[4] C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E. Sotheran, A. Gaiba, D.L. Wild, and F. Falciani. Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics*, **20**(9), 1361–1372, 2004.

[5] M.J.L. 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. *Pac. Symp. Biocomput.*, **8**, 17–28, 2003.

[6] N. Friedman, K. Murphy and S. Russell. Learning the structure of dynamic probabilistic networks. *Proc. Conference on Uncertainty in Artificial Intelligence*, 139–147, 1998.

[7] S. Kim, S. Imoto and S. Miyano . Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief. Bioinform.*, **4**(3), 228–235, 2003.

[8] S. Kim, S. Imoto and S. Miyano. Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems*, **75**(1-3), 57–65, 2004.

[9] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller and N. Friedman. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**(2), 166–176, 2003.

[10] N. Shephard. Partial non-Gaussian state space. *Biometrika*, **81**, 115–131, 1994.

[11] I. Shmulevich, E.R. Dougherty, S. Kim, and W. Zhang. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, **18**(2), 261–274, 2002.

[12] R.H. Shumway and D.S. Stoffer. Dynamic linear models with switching. *J. American Statistical Association*, **86**, 763–769, 1991.

[13] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein and B. Futcher. Comprehensive identification of cell cycleregulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297, 1998.

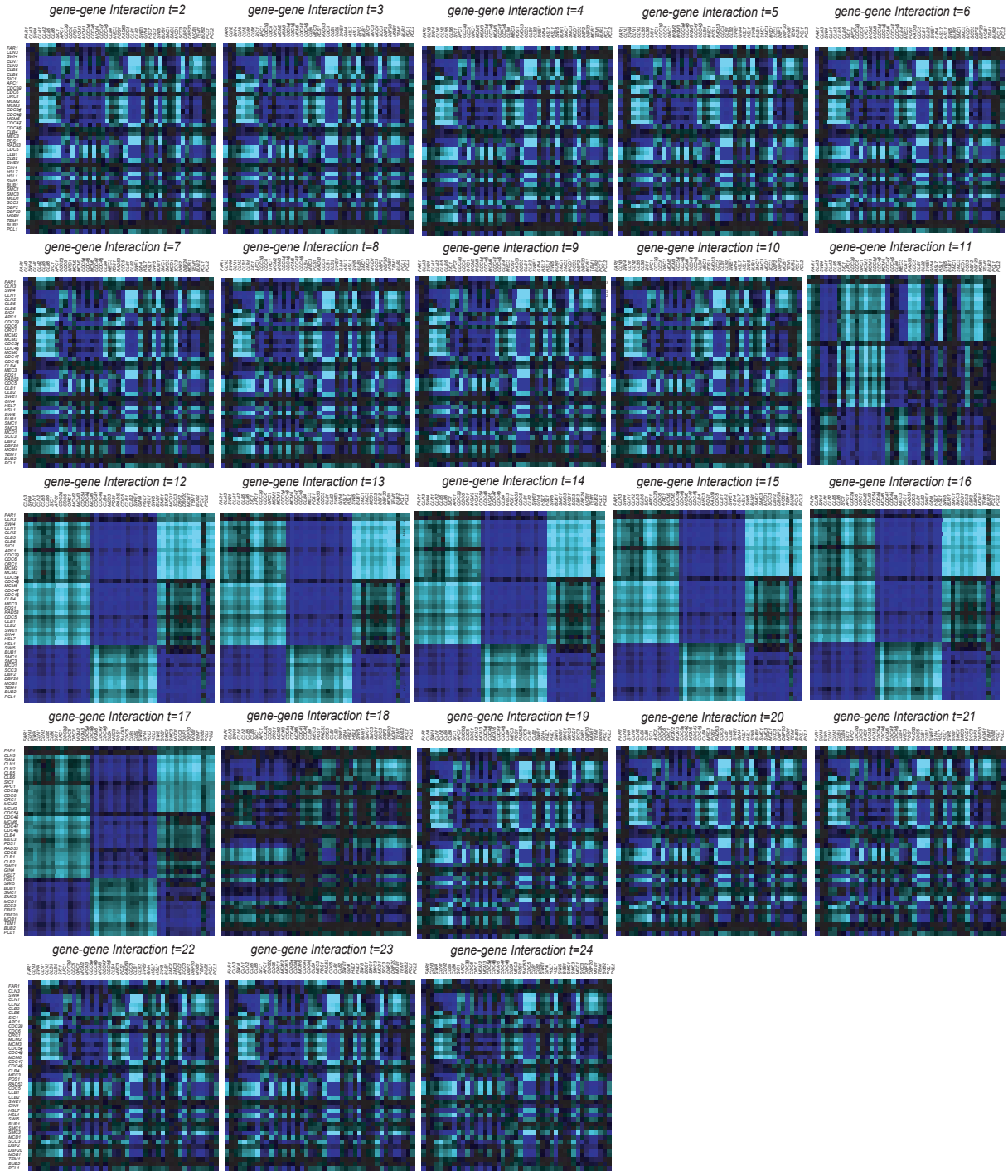

Figure 4. The estimated time-dependent interaction matrices, $\boldsymbol{H}_t$, $t = 2, \cdots, 24$: Each colored pixel expresses an intensity of gene-gene interaction from $t-1$ (column) to t (row).