State-space approach with the maximum likelihood principle to identify the system generating time-course gene expression data of yeast

Rui Yamaguchi*

The Institute of Medical Science,
The University of Tokyo,
4-6-1 Shirokanedai, Minato-ku,
Tokyo, 108-8639, Japan
E-mail: ruiy@ims.u-tokyo.ac.jp
*Corresponding author

Tomoyuki Higuchi

The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku,
Tokyo, 106-8569, Japan
E-mail: higuchi@ism.ac.jp

Abstract: We use linear Gaussian state-space models to analyse time-course gene expression data of yeast. They are modelled to be generated from hidden state variables in a system. To identify the system, we estimate parameters of the model by EM algorithm and determine the dimension of the state variable by BIC.

Keywords: state-space models; microarray; time-course data; EM algorithm; BIC; bioinformatics; data mining; cell cycle; gene networks.

Reference to this paper should be made as follows: Yamaguchi, R. and Higuchi, T. (2006) 'State-space approach with the maximum likelihood principle to identify the system generating time-course gene expression data of yeast', *Int. J. Data Mining and Bioinformatics*, Vol. 1, No. 1, pp.77–87.

Biographical notes: Rui Yamaguchi received his PhD in Science at the Kyushu University, in 2003. He is currently an Assistant Professor of the Institute of Medical Science, the University of Tokyo. His main research interest is the study and applications of state-space models and Bayesian networks. He is especially interested in gene networks.

Tomoyuki Higuchi received his PhD in Science at the University of Tokyo, in 1986. He is currently a Vice Director-General at the Institute of Statistical Mathematics (ISM), Research Organisation of Information and Systems. He is also a Professor of ISM and of the Graduate University for Advanced Studies. His primary research interests are in Bayesian modelling of space-time data and data mining.

1 Introduction

DNA microarray technology has allowed us to measure gene expression levels on genomic scale. Although many details inside a cell are not precisely known, gene expression data on a genomic scale provides useful insights into a living cell.

Data thus collected enhance fundamental understanding of life on the molecular level, and may prove useful also in medical diagnosis, treatment and drug design (Drăghici, 2003).

Towards time-course gene expression data, a wide variety of models, such as Boolean networks (Akutsu et al., 1999), differential equations (Chen et al., 1999) and linear and non-linear auto-regression models (Schmitt and Stephanopoulos, 2003) have been introduced to model cellular systems. Many of these published models can be considered to be special cases of a general class of graphical models known as dynamic Bayesian networks (Jensen, 2001; Murphy, 2002). Dynamic Bayesian networks are suitable for modelling gene expression data, because they can handle noisy or missing data, or handle hidden variables such as protein levels that may have an effect on mRNA expression levels. Although microarray technologies have made it possible to measure time course of the expression level of many genes simultaneously, we cannot hope to measure all possible factors contributing to genetic regulatory interactions. As well as this, gene expression data are known to include missing and outliers. Therefore, the ability of Bayesian networks to handle such hidden variables would appear to be one of the main advantages as a modelling tool (Kim et al., 2003). Although modelling gene networks by dynamic Bayesian networks is useful, the dimension of the observation variables (i.e., the number of genes) tends to be limited to relatively small numbers owing to its computational complexity. Therefore models which retain goodness of dynamic Bayesian networks and have lower computational complexity is demanded. Linear Gaussian state-space models (Harvey, 1989; Shumway and Stoffer, 1982), is a candidate for such models. Linear Gaussian state-space models are in a class of dynamic Bayesian networks, which assume that the observed time-course measurements (e.g., expression levels of genes) are generated by some hidden state variables that evolve according to Markovian dynamics. If the dimension of the state variable is relatively small, we can estimate the parameters in the model even if the number of genes is large, and thus may extract information about the system to generate the time course of gene expressions.

In this study, we use the linear Gaussian state-space models to model time-course of gene expression levels owing to the above reason. If we apply linear Gaussian state models to such data, there would be two problems:

- how to estimate the parameters of the model
- how to determine the appropriate dimension of the hidden state variable.

In this study, we estimate the parameters of the model based on maximum likelihood estimation method by using Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The dimension of the hidden state variable is determined by using Bayesian information criterion (BIC; Schwarz, 1978). The purpose of this study is to apply the linear Gaussian state-space model to a published cDNA microarray time-course data for yeast cell-cycle regulated genes obtained by Spellman et al. (1998) and estimate the parameters and determine the optimal dimension of the state variable. Then we compare the results with those reported in other papers (Wu et al., 2004; Yukinawa et al., 2005), in

which they used almost the same data and the same model but used the different method to estimate the parameters and the different criterion to determine the dimension of the state variable. In the discussion section, we especially focus on the comparison of the selected dimensions.

The organisation of this paper is as follows. In Section 2, we explain models and the parameter estimation methods. The method to determine the optimal dimension of the internal variable is also explained. In Section 3, we analyse a publicly available time-course gene expression data by the above methods and show the results. Then we discuss the results in Section 4. Section 5 is the concluding remarks.

2 System identification methods

2.1 Linear Gaussian state-space models

Let y_n be an l-dimensional vector containing observed expression levels of l genes at the n-th time step, where $n = 1, \ldots, N$. In order to model such time-course data, we use linear Gaussian state-space models, which are often simply called state-space models and a candidate to overcome difficulties mentioned above. State-space models have been used in a wide variety of applications with great success (Kitagawa and Gersch, 1996; West and Harrison, 1997). There are also studies using state-space models for gene expression data, which showed successful results (e.g., Rangel et al., 2004).

In state-space models, a sequence of the observation vectors $\{y_1, \ldots, y_N\}$ is modelled by assuming that at each time step y_n was generated from k-dimensional hidden state variable denoted by x_n. A basic model of state-space models is shown as follows:

$$x_n = Fx_{n-1} + v_n, \quad \text{(System model)}, \tag{1}$$

$$y_n = Hx_n + w_n, \quad \text{(Observation model)}, \tag{2}$$

where F is the state transition matrix ($k \times k$ matrix), H is the observation matrix ($l \times k$ matrix). $v_n \sim N_k(0_k, Q)$ and $w_n \sim N_l(0_l, R)$ are the system noise and the observation noise, respectively. The initial state vector x_0 is assumed to be a Gaussian random vector with mean vector μ_0 and covariance matrix Σ_0, i.e., $x_0 \sim N_k(\mu_0, \Sigma_0)$. If one knows the values for the parameters H, F, R, Q, μ_0, and Σ_0, in the model, the conventional Kalman smoothing estimators, $x_{n|N}$, $V_{n|N}$, and $V_{n,n-1|N}$ can be calculated as conditional expectations, that is,

$$x_{n|N} = E\{x_n \mid y_1, \ldots, y_N\}, \tag{3}$$

$$V_{n|N} = E\{(x_n - x_{n|N})(x_n - x_{n|N})' \mid y_1, \ldots, y_N\}, \tag{4}$$

$$V_{n,n-1|N} = E\{(x_n - x_{n|N})(x_{n-1} - x_{n-1|N})' \mid y_1, \ldots, y_N\}, \tag{5}$$

by using the Kalman filter (Kalman, 1960) and the fixed-lag smoother algorithm (Kitagawa and Gersch, 1996) efficiently. However, in this case, these parameters that determine the system are unknown and thus they should be estimated to identify the system. In this study, we estimate these parameters by the maximum likelihood method using EM algorithm. The estimation methods are explained in the following section.

2.2 Maximum likelihood estimation with EM algorithm

In order to obtain the maximum likelihood estimator of the parameters in the model, we use EM algorithm (Dempster et al., 1977). The EM algorithm for state-space models (equations (1) and (2)) is formulated by Shumway (2000) and Shumway and Stoffer (1982) as follows.

Let $\{Y_N, X_N\}$ call the complete data, where $Y_N = \{y_1, \ldots, y_N\}$ is the set of observation data and $X_N = \{x_0, \ldots, x_N\}$ is the set of state variables (unobserved data). Then the joint likelihood for the complete data is given by

$$P(Y_N, X_N; \theta) = P(x_0)\prod_{n=1}^{N} P(x_n \mid x_{n-1})P(y_n \mid x_n), \tag{6}$$

where $\theta = \{H, F, R, Q, \mu_0\}$ is the parameter vector in the model. $\sum_0$ is assumed to be known (Shumway and Stoffer, 1982). The probability densities $P(x_0)$, $P(x_n|x_{n-1})$ and $P(y_n|x_n)$ are given by

$$P(x_0) = \frac{\exp\left\{(-1/2)(x_0 - \mu_0)'\sum_0^{-1}(x_0 - \mu_0)\right\}}{(2\pi)^{k/2} \mid \sum_0 \mid^{1/2}}, \tag{7}$$

$$P(x_n \mid x_{n-1}) = \frac{\exp\left\{(-1/2)(x_n - Fx_{n-1})'Q^{-1}(x_n - Fx_{n-1})\right\}}{(2\pi)^{k/2} \mid Q \mid^{1/2}}, \tag{8}$$

$$P(y_n \mid x_n) = \frac{\exp\left\{(-1/2)(y_n - Hx_n)'R^{-1}(y_n - Hx_n)\right\}}{(2\pi)^{l/2} \mid R \mid^{1/2}}. \tag{9}$$

Thus the joint log-likelihood of the complete data becomes

$$\begin{aligned}\log P(Y_N, X_N; \theta) = &-\frac{1}{2}\log \mid \sum_0 \mid -\frac{1}{2}(x_0 - \mu_0)'\sum_0^{-1}(x_0 - \mu_0) \\ &-\frac{N}{2}\log \mid Q \mid -\frac{1}{2}\sum_{n=1}^{N}(x_n - Fx_{n-1})'Q^{-1}(x_n - Fx_{n-1}) \\ &-\frac{N}{2}\log \mid R \mid -\frac{1}{2}\sum_{n=1}^{N}(y_n - Hx_n)'R^{-1}(y_n - Hx_n) \\ &-\frac{k + N(k+l)}{2}\log 2\pi.\end{aligned} \tag{10}$$

In EM algorithm, to estimate the maximum likelihood parameter vector $\hat{\theta}$, the conditional expectation of the joint log-likelihood of the complete data

$$q(\theta \mid \theta^{\dagger}) = E[\log P(Y_N, X_N; \theta) \mid Y_N, \theta^{\dagger}], \tag{11}$$

is iteratively maximised as a function of θ until convergence, where $\theta^{\dagger}$ is the parameter vector obtained in the previous iteration. It is well known that the log-likelihood calculated with the $(i + 1)$-th iterative estimated parameters is larger than that with the i-th iterative estimated parameters. An iteration of EM algorithm consists of two steps

called the expectation step (E-step) and the maximization step (M-step), respectively. Each step in the $(i + 1)$th iteration is shown as follows.

In E-step, $q(\theta|\theta_i)$ of equation (11) is calculated by

$$\begin{aligned} q(\theta \mid \theta_i) &= E[\log P(Y_N, X_N \mid \theta) \mid Y_N, \theta_i] \\ &= -\frac{1}{2}\log|\Sigma_0| - \frac{1}{2}\operatorname{trace}\left\{\Sigma_0^{-1}(V_{0|N} + (x_{0|N} - \mu_0)(x_{0|N} - \mu_0)')\right\} \\ &\quad -\frac{N}{2}\log|Q| - \frac{1}{2}\operatorname{trace}\{Q^{-1}(C - BF' - FB' + FAF')\} \\ &\quad -\frac{N}{2}\log|R| \\ &\quad -\frac{1}{2}\operatorname{trace}\left\{R^{-1}\sum_{n=1}^{N}[(y_n - Hx_{n|N})(y_n - Hx_{n|N})' + HV_{n|N}H']\right\} \\ &\quad -\frac{k + N(k+l)}{2}\log 2\pi, \end{aligned} \tag{12}$$

where $\theta_i = \{H(i), F(i), R(i), Q(i), \mu_0(i)\}$ is the parameter vector estimated in the ith iteration, and

$$A = \sum_{n=1}^{N}(V_{n-1|N} + x_{n-1|N}x'_{n-1|N}), \tag{13}$$

$$B = \sum_{n=1}^{N}(V_{n,n-1|N} + x_{n|N}x'_{n-1|N}), \tag{14}$$

$$C = \sum_{n=1}^{N}(V_{n|N} + x_{n|N}x'_{n|N}), \tag{15}$$

In the above equation, the conventional Kalman smoothing estimators $x_{n|N}$, $V_{n|N}$, and $V_{n,n-1|N}$ (equations (3)–(5)) can be calculated by using the Kalman filter and the fixed-lag smoother algorithm as already mentioned in the previous section. It is noted that the log-likelihood

$$\log L(Y_N \mid \theta_i) = \log \int P(X_N, Y_N|\theta)\mathrm{d}X_N, \tag{16}$$

is also obtained as a byproduct of the Kalman filter.

In M-step, θ_i is updated to θ_{i+1} to be $\theta_{i+1} = \arg\max_\theta q(\theta \mid \theta_i)$ by $\partial_H q(\theta \mid \theta_i) = 0$, $\partial_F q(\theta \mid \theta_i) = 0$, $\partial_R q(\theta \mid \theta_i) = 0$, $\partial_Q q(\theta \mid \theta_i) = 0$ and $\partial_{\mu_0} q(\theta \mid \theta_i) = 0$. Thus $\theta_{i+1} = \{H(i + 1), F(i + 1), R(i + 1), Q(i + 1), \mu_0(i + 1)\}$ is obtained by

$$H(i+1) = \left(\sum_{n=1}^{N} E\{y_n x'_n \mid Y_N\}\right)C^{-1}, \tag{17}$$

$$F(i+1) = BA^{-1}, \tag{18}$$

$$R(i+1) = N^{-1}\sum_{n=1}^{N}[(y_n - Hx_{n|N})(y_n - Hx_{n|N})' + HV_{n|N}H'], \tag{19}$$

$$Q(i+1) = N^{-1}(C - BA^{-1}B'), \tag{20}$$

$$\mu_0(i+1) = x_{0|N}. \tag{21}$$

The procedure to obtain the maximum likelihood estimator of the parameter vector $\hat{\theta}$ is summarised as below.

P1: Select initial values of $\theta_0 = \{H(0), F(0), R(0), Q(0), \mu(0)\}$ and some reasonable baseline level of $\sum_0$. The conventional Kalman smoothing estimators $x_{n|N}$, $V_{n|N}$, $V_{n,n-1|N}$ (equations (3)–(5)) can be recursively calculated by Kalman filter and the fixed lag smoother with the upper initial parameters.

P2: Calculate the conditional expectation of the log likelihood with equation (12). (E-step)

P3: Calculate equations (17)–(21) and obtain the next iterative estimated parameters that maximise conditional expectation of the log likelihood. (M-step)

P4: Insert estimated parameters to the state-space equations (1) and (2), and calculate the conventional Kalman smoothing estimators.

P5: Repeat the upper procedures P2–P4 until the log likelihood is converged.

The EM algorithm for the maximum likelihood estimation may fall into a local maximum. Therefore, the global maximum must be chosen by comparing the results from several sets of the initial values. We note that the above algorithm can be extended to deal with missing values in observation values naturally (Shumway, 2000; Shumway and Stoffer, 1982).

2.3 Identification of dimension of the state variable

The dimension of the state vectors k is not yet identified. Bayesian Information Criterion (BIC; Schwarz, 1978) is introduced in this section in order to solve this problem. BIC for a model, which have k-dimensional state vector is as follows:

$$\mathrm{BIC}(k) = -2\log L(Y_N \mid \hat{\theta}^{(k)}) + \lambda_p \log v_s, \tag{22}$$

where $\log L(Y_N \mid \hat{\theta}^{(k)})$ is the maximum marginal log-likelihood with the parameter vector $\hat{\theta}^{(k)}$ estimated in EM algorithm (equation (16)). λ_p is the number of parameters to be estimated. v_s is the number of samples. In this case, $v_s = N$: the number of time points.

We consider the dimension of the state vectors that has the minimum BIC, i.e., $k = \arg\min_k \mathrm{BIC}(k)$, as the optimal one.

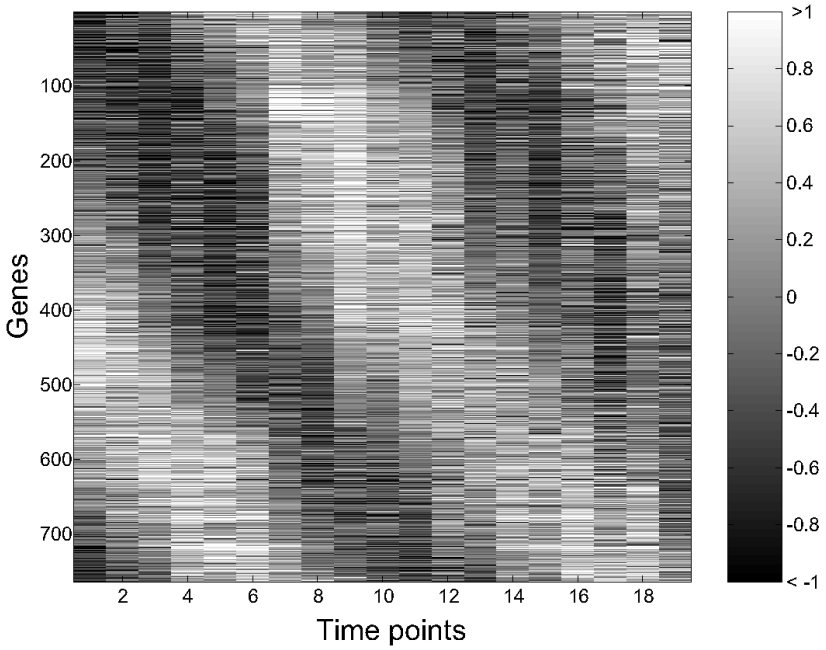
3 Analysis

In order to estimate the optimal dimension of state vectors for real dynamical biological systems, we applied the above method to a publicly available cDNA microarray time-course data set obtained for studying the cell-cycle regulated genes of yeast

(*Saccaromyces cerevisiae*) (Spellman et al., 1998). The dataset is available at *http://cellcycle-www.stanford.edu*.

Spellman et al. (1998) identified 800 genes that meet an objective minimum criterion for cell cycle regulation. They obtained the microarray data using samples from yeast cultures synchronised by three independent methods: α factor arrest, elutriation and arrest of a *cdc15* temperature-sensitive mutant. Here we used the *cdc15*'s time-course data of the 800 genes. The time course of a gene includes evenly spaced 19 time points. The observation interval is ten minutes. Figure 1 shows variations of expression levels of genes. Each of them shows rather a good wave form. For the analysis, we selected the time courses in which the number of missing points is less than ten. As a result, the number of selected time courses (i.e., the number of genes) is $l = 763$.

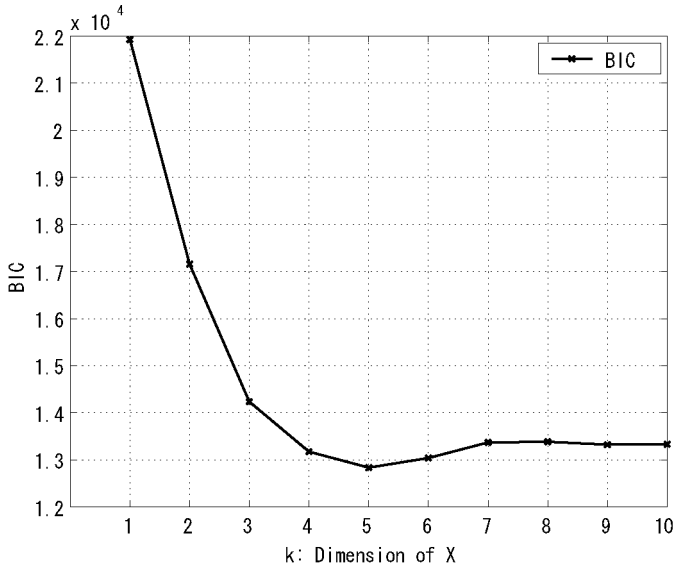
Figure 1 The time-course gene expression data (*cdc15*)



We then applied state-space models in which dimension of x_n are $k \in \{1, \ldots, 10\}$ to the data, and estimated maximum likelihood parameters and BIC for each case. In the analysis, we assumed diagonal matrices for Q and R in the model. We also assumed that all of the diagonal elements of R are the same. For each k, EM algorithm was applied five times with different initial parameter vectors to avoid local maxima of the log-likelihood. A parameter vector for each k yielding the largest log-likelihood among five of them was considered as the maximum likelihood parameter: $\hat{\theta}^{(k)}$. Then we calculated BIC(k) by equation (22) for each k, where $v_s = N = 19$ and $\lambda_p = k(k + 1 + 2) + 1$.

Finally, we obtained $k = 5$ as the optimal dimension of state vectors by $k = \arg\min_k$ BIC(k) (see Figure 2).

Figure 2 Dimension of $x_n(k)$ vs. BIC



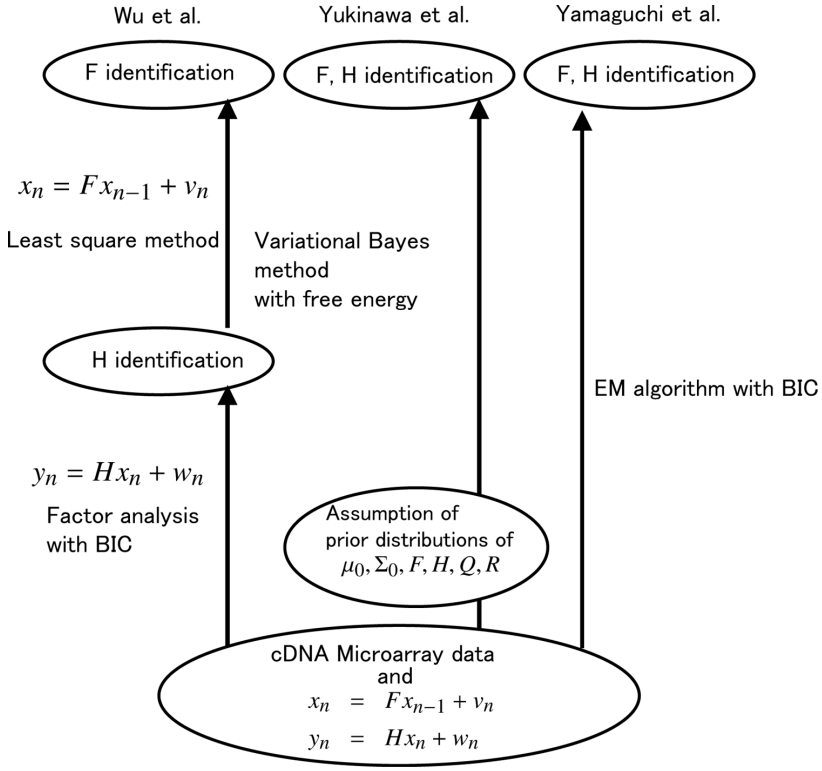
4 Discussion

There exist closely relating studies to this study, which dealt with the problem to determine the dimension of internal variable using almost the same model and data of this paper (Wu et al., 2004; Yukinawa et al., 2005). Here we compare these methods and results with those in this study.

Although there are small differences between them in methods to treat missing values in data, the more significant differences are in methods to estimate parameters and criteria to determine the optimal dimension (Wu et al., 2004), they proposed a two-step manner, at the first step, the state variable x_n and the H are estimated by factor analysis and the dimension of x_n is determined by minimising BIC. Then at the second step F is estimated using a least square method. In Yukinawa et al. (2005), considering Bayesian estimation, they estimated F and H simultaneously using a variational Bayes method assuming prior distributions for the parameters. The dimension of x_n is determined by maximising the variational free energy. In this study, we estimated F and H simultaneously as Yukinawa et al. (2005) but by using a maximum likelihood estimation with EM algorithm and without setting prior distributions for the parameters. We determined the dimension of x_n by BIC. All of their methods are schematically shown in Figure 3.

The estimated optimal dimension of internal state variable in this study was five. This is the same as that estimated in Wu et al. (2004). On the other hand, the estimated dimension by Yukinawa et al. (2005) was two. However, it should be noted that for calculating BICs, Wu et al. (2004) used the number of genes for v_s in equation (22), not the number of time points. For a time-course data, the number of time points is usually used for v_s. Since the number of genes in this data is much larger than that of time points, their obtained BICs have larger penalties for model complexities and thus the resulting dimension may change when the number of time points is used for BICs.

Figure 3 Comparison of procedures of the three papers



This comparison showed that the different estimation methods resulted in the different number of optimal dimension of the internal state variable. At this time it is difficult to answer a question which method is better than the others, that is, which method makes more accurate result, because criteria they used to determine the optimal dimension were different. Thus, it should be needed to investigate performances and tendencies of these methods systematically by using real and artificial data examples which have various numbers of genes and time points.

5 Concluding remarks

In this study, we applied state-space models in order to model the time-course data of gene expression levels of yeast, which have significant features, i.e., high dimension and short length. We used EM algorithm for parameter estimation and BIC for an optimal dimension of internal variable of x_n, i.e., $\hat{k}$. As a result, we obtained $\hat{k} = 5$. Then we compared the result, those of two studies (Wu et al., 2004; Yukinawa et al., 2005), which used almost the same model and the same data but different parameter estimation methods and criteria to determine the optimal dimension. Although our methods chose the same dimension of that in Wu et al. (2004), more systematic performance comparison using various real and artificial dataset would be needed.

We focused on determination of the dimension of internal variable in this study. The next important step is to compare the estimated parameters $\hat{\theta}$ and time course of state variables with the existing biological knowledge. It is especially useful if the internal variable can be interpreted in a real biological system. It is also expected that parameters F and H have important information about the connection between internal variables and genes. Estimating such networks and extracting other useful information by using the estimated parameters is our future work.

Acknowledgements

This research is partially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grant-in-Aid for Scientific Research (A), 17200020, 2005. This work is also partially supported by Kyushu University 21st Century COE Program, Development of Dynamic Mathematics with High Functionality, of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Akutsu, T., Miyano, S. and Kuhara, S. (1999) 'Identification of genetic networks from a small number of gene expression patterns under the Boolean network model', *Proceedings of Pacific Symposium on Biocomputing*, pp.17–28.

Chen, T., He, H.L. and Church, G.M. (1999) 'Modeling gene expression with differential equations', *Proceedings of Pacific Symposium on Biocomputing*, pp.29–40.

Dempster, A., Laird, N. and Rubin, D. (1977) 'Maximum likelihood from incomplete data via the EM algorithm (with discussion)', *Journal of the Royal Statistical Society, Series B*, Vol. 39, pp.1–38.

Drăghici, S. (2003) *Data Analysis Tools for DNA Microarrays,* Mathematical Biology and Medicine Series, Chapman & Hall/CRC.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and The Kalman Filter*, Cambridge University Press, New York.

Jensen, F.V. (2001) *Bayesian Networks and Decision Graphs*, Statistics for engineering and information science, Springer-Verlag, New York.

Kalman, R.E. (1960) 'A new approach to linear filtering and prediction problems', *Trans. Amer. Soc. Mech. Eng., J. Basic Engineering*, Vol. 82, pp.35–45.

Kim, S., Imoto, S. and Miyano, S. (2003) 'Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data', *Proceedings of the 1st Computational Methods in Systems Biology*, Vol. 2602 of *Lecture Note in Computer Science*, Springer-Verlag, pp.104–113.

Kitagawa, G. and Gersch, W. (1996) *Smoothness Priors Analysis of Time Series*, Springer-Verlag, New York.

Murphy, K.P. (2002) *Dynamic Bayesian Networks: Representation, Inference and Learning*, PhD Thesis, U.C. Berkeley.

Rangel, C., Angus, J., Ghahramani, Z., Lioumi, M., Sotheran, E., Gaiba, A., Wild, D.L. and Falciani, F. (2004) 'Modelling T-cell activation using gene expression profiling and state space models', *Bioinformatics*, Vol. 20, pp.1361–1372.

Schmitt Jr., W.A. and Stephanopoulos, G. (2003) 'Prediction of transcriptional profiles of synechocystis PCC6803 by dynamic autoregressive modeling of DNA microarray data', *Biotechnology and Bioengineering*, Vol. 84, No. 7, pp.855–863.

Schwarz, G. (1978) 'Estimating the dimension of a model', *Ann. Statist.*, Vol. 6, pp.461–464.

Shumway, R.H. (2000) 'Dynamic mixed models for irregularly observed time series', *Resenhas- Reviews of the Institute of Mathematics and Statistics*, University of Sao Paulo, USP Press, Brazil, Vol. 4, No. 4, pp.433–456.

Shumway, R.H. and Stoffer, D.S. (1982) 'An approach to time series smoothing and forecasting using the EM algorithm', *J. Time Series Analysis*, Vol. 3, pp.253–264.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B. (1998) 'Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces Cerevisiae by microarray hybridization', *Mol. Biol. of the Cell*, Vol. 9, pp.3273–3297, http://www.molbiolcell.org/cgi/content/abstract/9/12/3273.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*, 2nd ed., Springer-Verlag, New York.

Wu, F.X., Zhang, W.J. and Kusalic, A.J. (2004) 'Modeling gene expression from microarray expression data with state-space equations', *Pacific Symposium on Biocomputing*, Vol. 9, pp.581–592.

Yukinawa, N., Yoshimoto, J., Oba, S. and Ishii, S. (2005) 'System identification of gene expression time-series based on a linear dynamical system model with variational Bayesian estimation', *Information Processing Society of Japan, Trans. on Mathematical Modeling and its Applications*, Vol. 46, No. SIG10 (TOM12), pp.57–65, in Japanese.