Journal of Bioinformatics and Computational Biology
Vol. 1, No. 2 (2003) 231–252

Imperial College Press
www.icpress.co.uk

BAYESIAN NETWORK AND NONPARAMETRIC HETEROSCEDASTIC REGRESSION FOR NONLINEAR MODELING OF GENETIC NETWORK

SEIYA IMOTO[*,†], SUNYONG KIM[*,‡], TAKAO GOTO[*,§],
SACHIYO ABURATANI[¶,‖], KOUSUKE TASHIRO[¶,**], SATORU KUHARA[¶,††]
and SATORU MIYANO[*,‡‡]

[*]*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*
[‡]*sunk@ims.u-tokyo.ac.jp*
[§]*takao@ims.u-tokyo.ac.jp*
[‡‡]*miyano@ims.u-tokyo.ac.jp*
[¶]*Graduate School of Genetic Resources Technology, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan*
[‖]*sachiyo@grt.kyushu-u.ac.jp*
[**]*ktashiro@grt.kyushu-u.ac.jp*
[††]*kuhara@grt.kyushu-u.ac.jp*

Received 15 September 2002
Revised 7 December 2002
Accepted 15 January 2003

We propose a new statistical method for constructing a genetic network from microarray gene expression data by using a Bayesian network. An essential point of Bayesian network construction is the estimation of the conditional distribution of each random variable. We consider fitting nonparametric regression models with heterogeneous error variances to the microarray gene expression data to capture the nonlinear structures between genes. Selecting the optimal graph, which gives the best representation of the system among genes, is still a problem to be solved. We theoretically derive a new graph selection criterion from Bayes approach in general situations. The proposed method includes previous methods based on Bayesian networks. We demonstrate the effectiveness of the proposed method through the analysis of *Saccharomyces cerevisiae* gene expression data newly obtained by disrupting 100 genes.

Keywords: Genetic network; Bayesian network; nonparametric regression; heteroscedasticity.

1. Introduction

Due to the development of the microarray technology, constructing genetic network receives a large amount of attention in the fields of molecular biology and bioinformatics.[3–5,14,15,20,21,25,26,33,35,41,42,45,47] However, the dimensionality and complexity of the data disturb the progress of the microarray gene expression data

analysis. That is to say, the information that we want is buried in a huge amount of the data with noise. In this paper, we propose a new statistical method for constructing a genetic network that make capture even the nonlinear relationships between genes clearer.

A Bayesian network[10,19,34] is an effective method in modeling phenomena through the joint distribution of a large number of random variables. In recent years, some interesting works have been established in constructing genetic networks from microarray gene expression data by using Bayesian networks. Friedman and Goldszmidt[18] discretized the expression values and assumed multinomial distributions as the candidate statistical models. Pe'er *et al.*[42] investigated the threshold value for discretizing. On the other hand, Friedman *et al.*[21] pointed out that the discretizing probably loses information from the data. In fact, the number of discretizing values and the thresholds are unknown parameters, which have to be estimated from the data. The resulting network strongly depends on their values. Then Friedman *et al.*[21] considered fitting linear regression models, which analyze the data in the continuous (see also Heckerman and Geiger[29]). However, the assumption that the parent genes depend linearly on the objective gene is not always guaranteed. Imoto *et al.*[33] proposed the use of nonparametric additive regression models (see also Green and Silverman[23] and Hastie and Tibshirani[27]) for capturing not only linear dependencies but also nonlinear structures between genes. In this paper, we propose a method for constructing the genetic network by using Bayesian networks and the nonparametric heteroscedastic regression, which is more resistant to the effect of outliers. We observe that most gene expression data show heteroscedasticity. Recently, several normalizing or variance stabilizing transformations of gene expression data have been proposed.[17,24,32] So we believe that the modeling, which takes the effects of the heteroscedasticity and outliers into account, is a key to analyzing gene expression data and the extraction of valuable information.

Once we set the graph, we have to evaluate its goodness or closeness to the true graph, which is completely unknown. Hence, the construction of a suitable criterion becomes the center of attention of statistical genetic network modeling. Friedman and Goldszmidt[20] used the BDe criterion, which was originally derived by Cooper and Herskovits[9] for choosing a graph (see also Heckerman *et al.*[30]). The BDe criterion only evaluates the Bayesian network model based on the multinomial distributions and Dirichlet priors. However, Friedman and Goldszmidt[20] kept the unknown hyperparameters in Dirichlet priors and we only set up the values experimentally. We investigate the graph selection problem as a statistical model selection or evaluation problem and theoretically derive a new criterion for choosing a graph using the Bayes approach (see Berger[7]). The proposed criterion automatically optimizes all parameters in the model and gives the optimal graph when we can score all candidate graphs. However, for inferring a genetic network with a lot of genes, it is difficult or often impossible to score all candidate graphs. Then we employ the greedy hill-climbing algorithm for obtaining better genetic networks. In addition, our proposed method includes the previous methods for constructing

genetic network based on continuous data and Bayesian networks. To show the effectiveness of the proposed method, we use the Monte Carlo simulation method. We also analyze gene expression data of *Saccharomyces cerevisiae* newly obtained by disrupting 100 genes.

2. Bayesian Network and Nonparametric Heteroscedastic Regression Model

2.1. *Nonlinear Bayesian network model*

Let $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ be a p-dimensional random variable vector. The notation $\boldsymbol{a}^T$ denotes the transpose of $\boldsymbol{a}$. We consider a gene as a random variable. Under the Bayesian network framework, we consider a directed acyclic graph G and Markov assumption between nodes. The joint probability is then decomposed into the product of conditional probabilities, that is, $P(X_1, \ldots, X_p) = P(X_1|\boldsymbol{P}_1) \times \cdots \times P(X_p|\boldsymbol{P}_p)$, where $\boldsymbol{P}_j = (P_1^{(j)}, \ldots, P_{q_j}^{(j)})^T$ is a q_j-dimensional vector of parent variables of X_j in the graph G. This decomposition holds when we use densities instead of probability measure, $f(X_1, \ldots, X_p) = f_1(X_1|\boldsymbol{P}_1) \times \cdots \times f_p(X_p|\boldsymbol{P}_p)$. Hence, an essential point for constructing a genetic network based on a Bayesian network is the construction of each conditional density. In general, the density is specified by infinite dimensional parameters. We then parameterize the conditional densities as $f_j(X_j|\boldsymbol{P}_j; \boldsymbol{\theta}_j)$ for $j = 1, \ldots, p$ and the issue of the construction of $f_j(X_j|\boldsymbol{P}_j)$ is recast as the estimation of its parameter $\boldsymbol{\theta}_j$.

Suppose that we have n sets of array data $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of p genes, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$. We model the probabilistic system generating the data by the parametric model

$$f(x_{i1}, \ldots, x_{ip}; \boldsymbol{\theta}_G) = \prod_{j=1}^{p} f_j(x_{ij}|\boldsymbol{p}_{ij}; \boldsymbol{\theta}_j)\,, \quad i = 1, \ldots, n\,, \tag{1}$$

where $\boldsymbol{p}_{ij} = (p_{i1}^{(j)}, \ldots, p_{iq_j}^{(j)})^T$ are q_j-dimensional parent observation vectors of x_{ij} in the graph G and $\boldsymbol{\theta}_G = (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_p^T)^T$. For example, when gene_2 and gene_3 are parent genes of gene_1, we see $\boldsymbol{p}_{i1} = (x_{i2}, x_{i3})^T$ for $i = 1, \ldots, n$. We extract information from the data based on this probabilistic model.

Imoto *et al.*[33] proposed the use of nonparametric regression strategy for capturing the nonlinear relationships between x_{ij} and $\boldsymbol{p}_{ij}$ and suggested that there are many nonlinear relationships between genes. The linear model therefore hardly achieves a sufficient result. In many cases, this method can capture the objective relationships very well. When the data, however, contain outliers especially near the extreme values of the parent genes, nonparametric regression models sometimes lead to unsuitable smoothed estimates, i.e. the estimated curve exhibits some spurious waviness due to the effects of the outliers. Since what is estimated is the system of a living nature, a too complicated relationship is unsuitable. In fact, this inappropriate case unfortunately sometimes occurs in the analysis of real data.

To avoid this problem, we consider fitting a nonparametric regression model with heterogeneous error variances

$$x_{ij} = m_{j1}(p_{i1}^{(j)}) + \cdots + m_{jq_j}(p_{iq_j}^{(j)}) + \varepsilon_{ij}\,, \tag{2}$$

where ε_{ij} depends independently and normally on mean 0 and variance σ_{ij}^2 and $m_{jk}(\cdot)$ is a smooth function from R to R. Here R denotes a set of real numbers. This model includes Imoto *et al.*'s model[33] and, clearly, the linear regression model as special cases. In general, each smooth function $m_{jk}(\cdot)$ is characterized by the n values $m_{jk}(p_{1k}^{(j)}), \ldots, m_{jk}(p_{nk}^{(j)})$ and the system (2) contains $(n \times q_j + n)$ parameters. Then the number of the parameters in the model is much larger than the number of observations and it has a tendency toward unstable parameter estimates. In this paper, we construct the smooth function $m_{jk}(\cdot)$ by the basis functions approach

$$m_{jk}(p_{ik}^{(j)}) = \sum_{m=1}^{M_{jk}} \gamma_{mk}^{(j)} b_{mk}^{(j)}(p_{ik}^{(j)})\,, \quad k = 1, \ldots, q_j\,,$$

where $\gamma_{1k}^{(j)}, \ldots, \gamma_{M_{jk}k}^{(j)}$ are unknown coefficient parameters and $b_{1k}^{(j)}(\cdot), \ldots, b_{M_{jk}k}^{(j)}(\cdot)$ are basis functions. From this representation, the n parameters $m_{jk}(p_{1k}^{(j)}), \ldots, m_{jk}(p_{nk}^{(j)})$ are reparameterized by the M_{jk} coefficient parameters $\gamma_{1k}^{(j)}, \ldots, \gamma_{M_{jk}k}^{(j)}$.

We strongly recommend the use of nonparametric regression instead of linear regression, because linear regression cannot decide the direction of the Bayes causality or leads to the wrong direction in many cases. We show the advantage of the proposed model compared with linear regression through a simple example. Suppose that we have data of gene_1 and gene_2 in Fig. 1(a). We consider the two models $\text{gene}_1 \to \text{gene}_2$ and $\text{gene}_2 \to \text{gene}_1$, and obtain the smoothed estimates shown in Figs. 1(b) and 1(c), respectively. We decide that the model (b: $\text{gene}_1 \to \text{gene}_2$) is better than (c: $\text{gene}_2 \to \text{gene}_1$) by the proposed criterion, which is derived in a later section (the scores of the models are (b) 120.6 (c) 134.8). Since we generated this data from the true graph $\text{gene}_1 \to \text{gene}_2$, our method yields the correct result. Note that in this case the true model can be represented exactly by Eq. (2) by choosing parameters suitably. In addition, the true system underlying between gene_1 and gene_2 is almost the same as the curve shown in Fig. 1(b). On the other hand, if we fit the linear regression model to this data, the model (c) is chosen (the scores are (b) 156.0 (c) 135.8). The method, which is based on linear regression, yields an incorrect result in this case. Consider the case that the relationship is almost linear. Our method and linear regression can fit the data appropriately. However, it is clearly difficult to decide the direction of Bayes causality. In such a case, the direction is not strict.

In the error variances, σ_{ij}^2, we assume the structures,

$$\sigma_{ij}^2 = w_{ij}^{-1}\sigma_j^2\,, \quad i = 1, \ldots, n;\ j = 1, \ldots, p\,, \tag{3}$$

where $w_{1j}, \ldots, w_{nj}$ are constants and σ_j^2 is an unknown parameter. By setting up the constants $w_{1j}, \ldots, w_{nj}$ in reflecting the feature of the error variances, we can

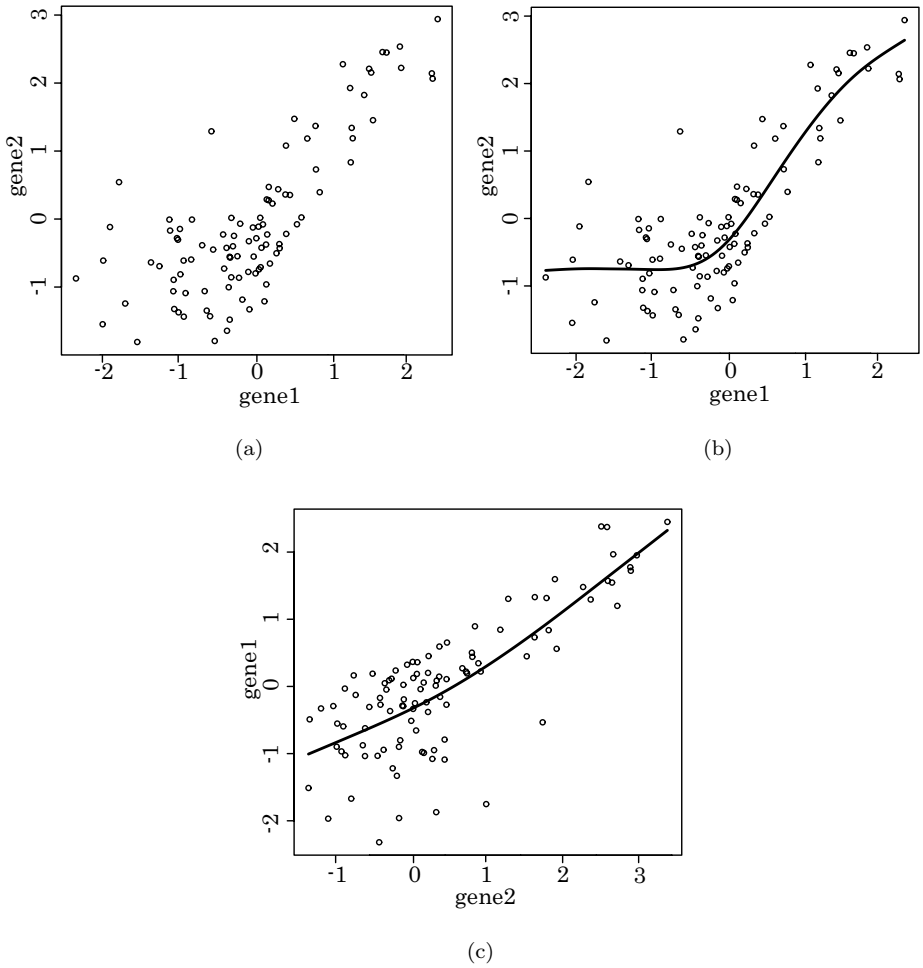

Fig. 1. Simulated data: The true causality is $\text{gene}_1 \rightarrow \text{gene}_2$. (a) Scatter plot of the simulated data. (b) Smoothed curve of the graph $\text{gene}_1 \rightarrow \text{gene}_2$. (c) Smoothed curve of the graph $\text{gene}_2 \rightarrow \text{gene}_1$. These curves are obtained by the proposed method.

represent the heteroscedasticity of the data. Combining (2) and (3), we obtain a nonparametric regression model with heterogeneous error variances

$$f_j(x_{ij}|\boldsymbol{p}_{ij};\boldsymbol{\gamma}_j,\sigma_j^2) = \left(\frac{w_{ij}}{2\pi\sigma_j^2}\right)^{1/2} \exp\left[-\frac{w_{ij}}{2\sigma_j^2}\left\{x_{ij} - \sum_{k=1}^{q_j} \boldsymbol{\gamma}_{jk}^T \boldsymbol{b}_{jk}(p_{ik}^{(j)})\right\}^2\right], \qquad (4)$$

where $\boldsymbol{\gamma}_{jk}$ and $\boldsymbol{b}_{jk}(p_{ik}^{(j)})$ are M_{jk}-dimensional vectors given by, respectively, $\boldsymbol{\gamma}_{jk} = (\gamma_{1k}^{(j)}, \ldots, \gamma_{M_{jk}k}^{(j)})^T$ and $\boldsymbol{b}_{jk}(p_{ik}^{(j)}) = (b_{1k}^{(j)}(p_{ik}^{(j)}), \ldots, b_{M_{jk}k}^{(j)}(p_{ik}^{(j)}))^T$. If the jth gene has no parent genes in the graph, we specify f_j by the normal distribution with mean μ_j and variance σ_j^2. Hence, we define the Bayesian network and nonparametric

heteroscedastic regression model by replacing each conditional density in (1) with (4) or normal density with mean μ_j and variance σ_j^2.

2.2. *Criterion for choosing graph*

Once we set a graph, the statistical model based on the Bayesian network and nonparametric heteroscedastic regression can be constructed and be estimated by a suitable procedure. However, the problem that still remains to be solved is how we can choose the optimal graph, which gives a best approximation of the system underlying the data. Notice that we cannot use the likelihood function as a model selection criterion, because the value of likelihood becomes large in a more complicated model. Hence, we need to consider the statistical approach based on the generalized or predictive error, Kullback–Leibler information, Bayes approach and so on (see e.g. Akaike,[1] Burnham and Anderson,[8] Konishi[37] and Konishi and Kitagawa[38] for the statistical model selection problem). In this section, we construct a criterion for evaluating a graph based on our model from Bayes approach.

The posterior probability of the graph $\pi(G|\boldsymbol{X}_n)$ is obtained by the product of the prior probability of the graph π_G and the marginal probability of the data divided by the standardizing constant. By removing the standardizing constant, the posterior probability of the graph is proportional to

$$\pi(G|\boldsymbol{X}_n) \propto \pi_G \int \prod_{i=1}^{n} f(\boldsymbol{x}_i;\boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})d\boldsymbol{\theta}_G\,, \tag{5}$$

where $\boldsymbol{X}_n = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^T$ is an $n \times p$ gene profile matrix, $\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})$ is the prior distribution on the parameter $\boldsymbol{\theta}_G$ satisfying $\log \pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda}) = O(n)$ and $\boldsymbol{\lambda}$ is the hyperparameter vector. Under Bayes approach, we can choose the optimal graph such that $\pi(G|\boldsymbol{X}_n)$ is maximum. A crucial problem for constructing a criterion based on the posterior probability of the graph is the computation of the high dimensional integration (5). Heckerman and Geiger[29] used the conjugate priors for solving the integral and gave a closed-form solution. To compute this high dimensional integration, we use Laplace's approximation[12,28,46] for the integrals

$$\int \prod_{i=1}^{n} f(\boldsymbol{x}_i;\boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})d\boldsymbol{\theta}_G = \frac{(2\pi/n)^{r/2}}{|J_\lambda(\hat{\boldsymbol{\theta}}_G)|^{1/2}} \exp\{n l_\lambda(\hat{\boldsymbol{\theta}}_G|\boldsymbol{X}_n)\}\{1+O_p(n^{-1})\}\,,$$

where r is the dimension of $\boldsymbol{\theta}_G$, $l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n) = \sum_{i=1}^{n} \log f(\boldsymbol{x}_i;\boldsymbol{\theta}_G)/n + \log \pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})/n$, $J_\lambda(\boldsymbol{\theta}_G) = -\partial^2\{l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n)\}/\partial\boldsymbol{\theta}_G\partial\boldsymbol{\theta}_G^T$ and $\hat{\boldsymbol{\theta}}_G$ is the mode of $l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n)$. Then we define the Bayesian network and nonparametric heteroscedastic regression criterion, named BNRC_{hetero}, for selecting a graph

$$\begin{aligned}\mathrm{BNRC}_{hetero}(G) &= -2\log\left\{\pi_G \int \prod_{i=1}^{n} f(\boldsymbol{x}_i;\boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})d\boldsymbol{\theta}_G\right\} \\ &\approx -2\log\pi_G - r\log(2\pi/n) + \log|J_\lambda(\hat{\boldsymbol{\theta}}_G)| - 2n l_\lambda(\hat{\boldsymbol{\theta}}_G|\boldsymbol{X}_n)\,. \end{aligned} \tag{6}$$

The optimal graph is chosen such that the criterion BNRC_{hetero} (6) is minimal. The merit of the use of the Laplace method is that it is not necessary to consider the use of the conjugate prior distribution. Hence the modeling in the larger classes of distributions of the model and prior is attained.

Suppose that the parameter vectors $\boldsymbol{\theta}_j$ are independent of one another. The prior distribution can then be decomposed into $\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda}) = \prod_{j=1}^{p} \pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j)$. Therefore, $\log|J_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n)|$ and $nl_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n)$ in (6) result in, respectively,

$$\log|J_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n)| = \sum_{j=1}^{p} \log\left|-\frac{\partial^2 l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)}{\partial\boldsymbol{\theta}_j \partial\boldsymbol{\theta}_j^T}\right|,$$

$$l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}_n) = \sum_{j=1}^{p} l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n),$$

where $l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n) = \log f_j(x_{ij}|\boldsymbol{p}_{ij};\boldsymbol{\theta}_j)/n + \log\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j)/n$. Here $\boldsymbol{\lambda}_j$ is the hyperparameter vector. Hence by defining

$$\mathrm{BNRC}_{hetero}^{(j)} = -2\log\left\{\int \pi_{L_j} \prod_{i=1}^{n} f_j(x_{ij}|\boldsymbol{p}_{ij};\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j)d\boldsymbol{\theta}_j\right\},$$

where π_{L_j} are prior probabilities satisfying $\sum_{j=1}^{p}\log\pi_{L_j} = \log\pi_G$, the BNRC_{hetero} score is given by the sum of the local scores

$$\mathrm{BNRC}_{hetero}(G) = \sum_{j=1}^{p} \mathrm{BNRC}_{hetero}^{(j)}. \tag{7}$$

The smoothed estimates based on nonparametric heteroscedastic regression are obtained by replacing the parameters $\boldsymbol{\gamma}_j$ by $\hat{\boldsymbol{\gamma}}_j$. Noticed that we derive the criterion, BNRC_{hetero}, under the assumption $\log\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda}) = O(n)$. If we use the prior density satisfying $\log\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda}) = O(1)$, the BNRC_{hetero} score results in Schwarz's criterion known as BIC or SIC[44]. In such case, the mode $\hat{\boldsymbol{\theta}}_G$ is equivalent to the maximum likelihood estimate. When $l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)$ has a complex surface, the use of the Laplace approximation is inappropriate. For our model, we checked the surface of $l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)$ as much as possible and we observed that the surface is convex. We presume the reason is the use of basis function approach. Due to the basis function approach, the number of parameters is relatively small compared with other nonlinear modeling techniques such as neural networks.

3. Estimating Genetic Network

3.1. *Nonparametric regression*

In this section we present the method for constructing genetic network in practice based on the proposed method described above. First we would like to mention the nonparametric regression model. In the additive model, we construct each smooth function $m_{jk}(\cdot)$ by B-splines.[13,33] Figure 2 is an example of B-splines smoothed

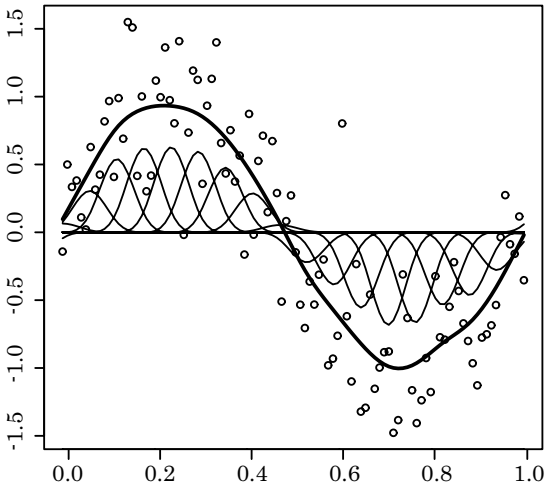

Fig. 2. The fitted curve to simulated data: The thin curves are B-splines that are weighted by coefficients and the thick curve is the smoothed estimate that is obtained by the linear combination of the weighted B-splines.

curve. The thin curves are B-splines that are weighted by coefficients and thick line is a smoothed curve that is obtained by the linear combination of weighted B-splines. Radial basis functions[6] form an alternative choice for the basis function. Radial basis function models can describe the interaction of the parent genes naturally.

In the error variances, we consider the heteroscedastic regression model and assume the structure (3). Choosing constants $w_{1j}, \ldots, w_{nj}$ is an important problem for capturing the heteroscedasticity of the data. In this paper, we set the weights

$$w_{ij} = g(\boldsymbol{p}_{ij}; \rho_j) = \exp\{-\rho_j \|\boldsymbol{p}_{ij} - \bar{\boldsymbol{p}}_j\|^2 / 2s_j^2\}\,, \tag{8}$$

where ρ_j is a hyperparameter, $\bar{\boldsymbol{p}}_j = \sum_{i=1}^n \boldsymbol{p}_{ij}/n$ and $s_j^2 = \sum_{i=1}^n \|\boldsymbol{p}_{ij} - \bar{\boldsymbol{p}}_j\|^2/nq_j$. If we set $\rho_j = 0$, the weights are $w_{1j} = \cdots = w_{nj} = 1$ and the model has homogeneous error variances. If we use a large value of ρ_j, the error variances of the data, which exist near the extreme values of the parent genes, are large. Hence, if there are outliers near the extreme values of the parent genes, we can reduce their effect and gain the suitable smoothed estimates by using the appropriate value of ρ_j. In addition, several choices are possible for setting $\bar{\boldsymbol{p}}_j$ in (8). For example, the median of $\boldsymbol{p}_{1j}, \ldots, \boldsymbol{p}_{nj}$ can be used and possibly achieve more robustness against the outliers in the parent genes. For capturing the heteroscedasticity, the self organizing state space model, which allows for heterogeneous error variances, is a powerful alternative to deal with a nonparametric smooth curve estimation. For examples, see Kitagawa[36] for a methodological introduction and Higuchi and Kitagawa[31] for an illustrative application suited for the problem dealt with in this paper.

3.2. *Priors*

Suppose that the prior distribution $\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j)$ is factorized as $\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j) = \prod_{k=1}^{q_j} \pi_{jk}(\boldsymbol{\gamma}_{jk}|\lambda_{jk})$, where λ_{jk} are hyperparameters. We use a singular M_{jk} variate normal distribution as the prior distribution on $\boldsymbol{\gamma}_{jk}$,

$$\pi_{jk}(\boldsymbol{\gamma}_{jk}|\lambda_{jk}) = \left(\frac{2\pi}{n\lambda_{jk}}\right)^{-(M_{jk}-2)/2} |K_{jk}|_+^{1/2} \exp\left(-\frac{n\lambda_{jk}}{2}\boldsymbol{\gamma}_{jk}^T K_{jk}\boldsymbol{\gamma}_{jk}\right), \qquad (9)$$

where K_{jk} is an $M_{jk} \times M_{jk}$ symmetric positive semidefinite matrix satisfying $\boldsymbol{\gamma}_{jk}^T K_{jk}\boldsymbol{\gamma}_{jk} = \sum_{\alpha=3}^{M_{jk}} (\gamma_{\alpha k}^{(j)} - 2\gamma_{\alpha-1,k}^{(j)} + \gamma_{\alpha-2,k}^{(j)})^2$.

Next we consider the prior probability of the graph π_G. Friedman and Goldszmit[20] employed the prior based on the MDL encoding of the graph. In our context, the marginal probability of the data is equivalent to the type II likelihood[22] with hyperparameters. Thus we set the prior probability of the graph, π_G,

$$\pi_G = \exp\{-(\text{No. of hyperparameters})\} = \prod_{j=1}^{p} \exp\{-(q_j+1)\} = \prod_{j=1}^{p} \pi_{L_j}\,.$$

The justification of this prior is based on Akaike's Bayesian information criterion, known as ABIC[2], and Akaike's information criterion, AIC[1].

3.3. *Criterion*

We derived the criterion, BNRC_{hetero}, for choosing the graph in a general framework. By using the Eq. (7), the BNRC_{hetero} score of the graph can be obtained by the sum of the local scores, $\text{BNRC}_{hetero}^{(j)}$. The result is summarized in the following theorem.

Theorem 1. *Let $f(\boldsymbol{x}_i;\boldsymbol{\theta}_G)$ be a Bayesian network and nonparametric heteroscedastic regression model given in Sec.* 2.1, *and let $\pi(\boldsymbol{\gamma}_{jk}|\lambda_{jk})$ be the prior densities on the parameters $\boldsymbol{\gamma}_{jk}$ defined by* (9). *Then a criterion for evaluating graph is given by* $\text{BNRC}_{hetero} = \sum_{j=1}^{p} \text{BNRC}_{hetero}^{(j)}$, *where*

$$\begin{aligned}
\text{BNRC}_{hetero}^{(j)} = {} & 2(q_j+1) - \left(\sum_{k=1}^{q_j} M_{jk} + 1\right)\log\left(\frac{2\pi}{n}\right) \\
& - \sum_{i=1}^{n} \log w_{ij} + n\log(2\pi\hat{\sigma}_j^2) + n \\
& + \sum_{k=1}^{q_j} \{\log|\Lambda_{jk}| - M_{jk}\log(n\hat{\sigma}_j^2)\} - \log(2\hat{\sigma}_j^2) \\
& + \sum_{k=1}^{q_j} \left\{(M_{jk}-2)\log\left(\frac{2\pi\hat{\sigma}_j^2}{n\beta_{jk}}\right) - \log|K_{jk}|_+ + \frac{n\beta_{jk}}{\hat{\sigma}_j^2}\hat{\boldsymbol{\gamma}}_{jk}^T K_{jk}\hat{\boldsymbol{\gamma}}_{jk}\right\},
\end{aligned}$$

with

$$\Lambda_{jk} = B_{jk}^T W_j B_{jk} + n\beta_{jk} K_{jk}\,; \quad (M_{jk} \times M_{jk})\,,$$

$$B_{jk} = (\boldsymbol{b}_{1k}^{(j)}(p_{1k}^{(j)}), \ldots, \boldsymbol{b}_{M_{jk}k}^{(j)}(p_{nk}^{(j)}))^T;\ (n \times M_{jk})\,,$$

$$W_j = \mathrm{diag}(w_{1j}, \ldots, w_{nj})\,; \quad (n \times n)$$

$$\hat{\sigma}_j^2 = \sum_{i=1}^{n} w_{ij} \left\{ x_{ij} - \sum_{k=1}^{q_j} \hat{\boldsymbol{\gamma}}_{jk}^T \boldsymbol{b}_{jk}(p_{ik}^{(j)}) \right\}^2 \Big/ n\,.$$

Here we approximate the Hessian matrix by

$$\log \left| -\frac{\partial^2 l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j^T} \right| \approx \sum_{k=1}^{q_j} \log \left| -\frac{\partial^2 l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)}{\partial \boldsymbol{\gamma}_{jk} \partial \boldsymbol{\gamma}_{jk}^T} \right| + \log \left| -\frac{\partial^2 l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)}{\partial (\sigma_j^2)^2} \right|\,.$$

□

3.4. *Learning network*

In the Bayesian network literature, it is shown that determining the optimal network is an NP-hard problem. In this paper, we use the greedy hill-climbing algorithm for learning network as follows:

Step 1. Make the score matrix whose (i, j)th element is the $\mathrm{BNRC}_{hetero}^{(j)}$ score of the graph $\mathrm{gene}_i \to \mathrm{gene}_j$.

Step 2. For each gene, implement one of three procedures for an edge: "add", "remove", "reverse", which gives the smallest BNRC_{hetero}.

Step 3. Repeat Step 2 until the BNRC_{hetero} does not reduce.

Generally, the greedy hill-climbing algorithm has many local minima and the result depends on the computational order of variables. To avoid this problem, we permute the computational order of genes and make many candidate learning orders in Step 3. Another problem of the learning network is that the search space of the parent genes is enormously wide, when the number of genes is large. Then we restrict the set of the candidate parent genes based on the score matrix, which is given by Step 1.

3.5. *Hyperparameters*

Consider the nonparametric regression model defined in (4). The estimate $\hat{\boldsymbol{\theta}}_j$ is a mode of $l_{\lambda_j}(\boldsymbol{\theta}_j|\boldsymbol{X}_n)$ and depends on the hyperparameters. In fact, the hyperparameter plays an essential role for estimating the smoothed curve.

In our model, we construct the nonparametric regression model by 20 B-splines. We confirmed that the differences of the smoothed estimates against the various number of the basis functions cannot be found visually. Because when we use a somewhat large number of the basis functions, the hyperparameters control the

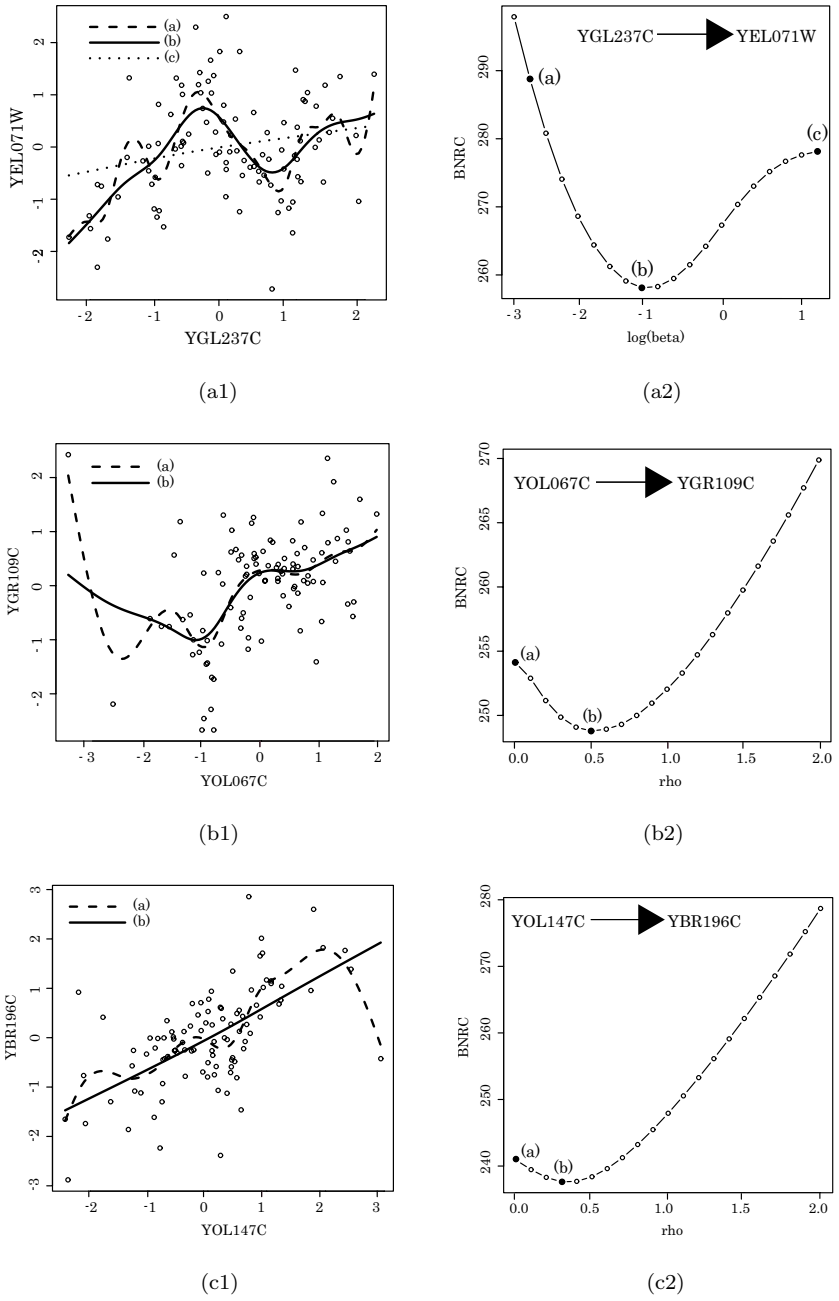

Fig. 3. The smoothed estimates by the various values of the hyperparameters. (a1): The effect of hyperparameter $\beta_{jk} = \hat{\sigma}_j^2 \lambda_{jk}$ in the prior distribution of the coefficients of B-splines. This parameter can control the smoothness of the fitted curve. (b1) and (c1): The effect of hyperparameter ρ_j in the parameter of the error variances. This parameter can capture the heteroscedasticity of the data and can reduce the effects of outliers.

smoothness of the fitted curves. Figure 3(a1) shows the scatter plot of YGL237C and YEL071W with smoothed estimates for 3 different values of the hyperparameters. The details of the data are shown in later section. Clearly, the smoothed estimate strongly depends on the values of the hyperparameters. Figure 3(a2) is the behavior of the BNRC_{hetero} criterion of the two genes in Fig. 3(a1). We can choose the optimal value of the hyperparameter as the minimizer of the BNRC_{hetero} and the optimal smoothed estimate (solid curve in Fig. 3(a1)) can capture the structure between these genes well. The dashed and dotted curves are near the maximum likelihood estimate and the parametric linear fit, respectively.

The effect of the weight constants $w_{1j}, \ldots, w_{nj}$ is shown in Figs. 3(b1) and 3(c1). If we use the nonparametric homoscedastic regression model[33], we obtain the dashed curve, which exhibits some spurious waviness due to the effect of the data in the upper-left corner Fig. 3(b1). By adjusting the hyperparameter ρ_j in (8), the estimated curve results in the solid curve. The optimal value of ρ_j is also chosen by minimizing the BNRC_{hetero} criterion (see Figs. 3(b2) and 3(c2)). Of course, when the smoothed estimate is not affected by outliers, the optimal value of ρ_j tends to zero.

Finally, we show the algorithm for estimating the smoothed curve and optimizing the hyperparameters.

Step 1. Fix the hyperparameter ρ_j.

Step 2. Initialize: $\boldsymbol{\gamma}_{jk} = \mathbf{0}$, $k = 1, \ldots, q_j$.

Step 3. Find the optimal β_{jk} by repeating Steps 3-1 and 3-2.

Step 3-1. Compute:

$$\boldsymbol{\gamma}_{jk} = (B_{jk}^T W_{jk} B_{jk} + n\beta_{jk} K_{jk})^{-1} B_{jk}^T W_{jk} \left(\boldsymbol{x}_{(j)} - \sum_{k' \neq k} B_{jk'} \boldsymbol{\gamma}_{jk'} \right),$$

for fixed β_{jk}.

Step 3-2. Evaluate: Repeat Step 3-1 against the candidate value of β_{jk}, and choose the optimal value of β_{jk}, which minimizes the $\mathrm{BNRC}_{hetero}^{(j)}$.

Step 4. Convergence: Repeat Step 3 for $k = 1, \ldots, q_j, 1, \ldots, q_j, 1, \ldots$ until a suitable convergence criterion is satisfied.

Step 5. Repeat Step 1 to Step 4 against the candidate value of ρ_j, and choose the optimal value of ρ_j, which minimizes the $\mathrm{BNRC}_{hetero}^{(j)}$.

4. Computational Experiments

4.1. *Monte Carlo simulation*

We use the Monte Carlo simulation method to show the effectiveness of our method. The data were generated from the artificial network of Fig. 4(a) with the functional

structures between nodes as follows:

$$
\begin{aligned}
X_1 &= X_2^2 + 2\sin(X_5) - 2X_7 + \varepsilon_1\,, \quad \varepsilon_1 \sim N(0,(4s)^2)\,,\\
X_2 &= \{1+\exp(-4X_3)\}^{-1} + \varepsilon_2\,, \quad \varepsilon_2 \sim N(0,s^2)\,,\\
X_3 &= \varepsilon_3\,, \quad \varepsilon_3 \sim N(0,1)\,,\\
X_4 &= X_5^2/3 + \varepsilon_4\,, \quad \varepsilon_4 \sim N(0,(4s)^2)\,,\\
X_5 &= X_3 - X_6^2 + \varepsilon_5\,, \quad \varepsilon_5 \sim N(0,(4s)^2)\,,\\
X_6 &= \varepsilon_6\,, \quad \varepsilon_6 \sim N(0,1)\,,\\
X_7 &= \begin{cases} -1+\varepsilon_7\,, & (X_8 \le -0.5)\,,\\ X_8+\varepsilon_7\,, & (-0.5 < X_8 \le 0.5)\,,\\ 1+\varepsilon_7\,, & (0.5 < X_8)\,, \end{cases} \quad \varepsilon_7 \sim N(0,(2s)^2),\\
X_8 &= \exp(-X_4-1)/2 + \varepsilon_8\,, \quad \varepsilon_8 \sim N(0,(2s)^2)\,,\\
X_9 &= \varepsilon_9\,, \quad \varepsilon_9 \sim N(0,1)\,,\\
X_{10} &= \cos(X_9) + \varepsilon_{10}\,, \quad \varepsilon_{10} \sim N(0,(4s)^2)\,,
\end{aligned}
$$

where s is a constant. After transforming the observations of the parent variables to mean 0 and variance 1, then the observations of the child variable are generated.

We generate 100 observations from this artificial network and our aim is to rebuild the network in Fig. 4(a) from the simulated data. We use two different

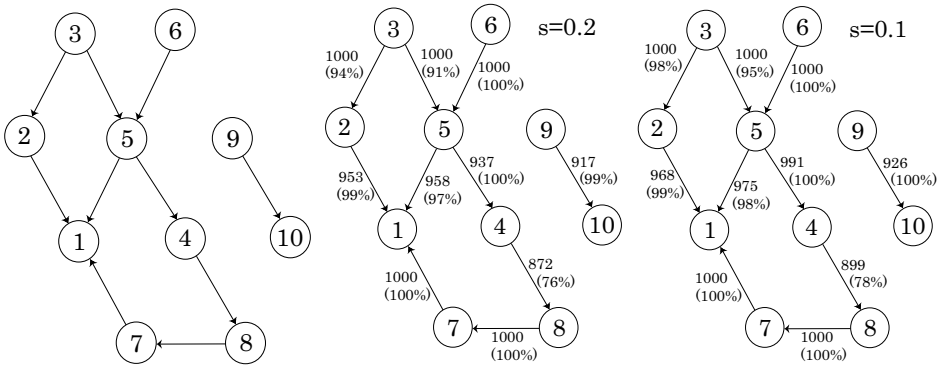

Fig. 4. The results of the Monte Carlo simulations. (a) True network. (b) Result for $s = 0.2$. (c) Result for $s = 0.1$. The number next to edge represents the number of estimated connections from 1000 Monte Carlo experiments. The percentage includes the information of the edge direction. For example, in $s = 0.2$, the connection between X5 and X1 appeared 958 times from 1000 Monte Carlo experiments and those 97% is the correct direction (from X5 to X1).

Table 1. The false positives of the Monte Carlo simulations. The number attached after node name is the number of estimated connection without direction information and the percentage is the direction information. For example, in $s = 0.2$, the proposed method estimated the relationship "X1 → X4" or "X1 ← X4" 15 times from 1000 Monte Carlo simulations and 87 percent of 15 times represents the direction from left to right ("X1 → X4").

$s = 0.2$									
X1	–	X4	15	87%	X6	–	X3	325	81%
		X6	73	99%			X9	190	53%
		X9	72	99%	X7	–	X2	12	58%
		X10	10	70%			X3	32	97%
X2	–	X4	20	65%			X4	24	71%
		X6	71	87%			X5	27	85%
		X8	65	78%			X6	12	98%
		X9	120	92%			X9	68	96%
		X10	27	59%			X10	3	67%
X3	–	X1	31	55%	X8	–	X1	1	100%
		X4	20	90%			X6	139	87%
		X8	66	52%			X9	146	65%
		X9	149	70%	X10	–	X3	71	87%
X4	–	X6	35	51%			X4	8	75%
		X9	137	95%			X6	139	95%
X5	–	X2	9	100%			X8	83	94%
		X8	28	97%					
		X9	108	100%					
		X10	6	100%					
$s = 0.1$									
X1	–	X4	16	100%	X6	–	X3	308	74%
		X6	53	83%			X4	4	50%
		X9	68	100%	X7	–	X2	1	100%
		X10	1	100%			X3	28	100%
X2	–	X4	4	100%			X4	20	95%
		X6	57	91%			X5	18	89%
		X8	31	97%			X6	93	99%
		X9	89	96%			X9	43	98%
		X10	10	80%	X8	–	X1	1	100%
X3	–	X1	23	57%			X6	154	81%
		X4	5	80%			X9	123	77%
		X8	58	52%	X9	–	X6	169	51%
		X9	157	69%	X10	–	X3	81	88%
X4	–	X9	172	99%			X4	9	67%
X5	–	X2	8	75%			X6	156	100%
		X8	27	100%			X7	1	100%
		X9	106	100%			X8	77	96%
		X10	6	100%					

settings of the noise variance, one is $s = 0.2$ and another is $s = 0.1$. The observations from the setting of the noise $s = 0.2$ are experientially similar to the real microarray data. The Monte Carlo simulation was repeated 1000 times and we focused on the number of correct estimations. Figures 4(b) and 4(c) are the results of the Monte Carlo simulations for $s = 0.2$ and $s = 0.1$, respectively.

The results of the Monte Carlo simulations can be summarized as follows: In the setting of noise variance $s = 0.2$, our model can rebuild the target network very well. Table 1 shows the false positives of the Monte Carlo simulations and we can see the percentages of the false positives are almost less than ten percent. Since the simulated data is similar to the real microarray data in the setting $s = 0.2$, we can expect that our network estimation method can work effectively in the real data analysis. From Fig. 4(b) and 4(c) and Table 1, the number of true negatives is much less than the number of false positives. We believe that this tendency is preferred in the exploratory data analysis. Most false positives are related to X_3, X_6 and X_9. Those variables are all independent normals. We presume the reason is similar to the problem of multiple comparisons. It is difficult to obtain a theoretic solution for this problem. A possible solution is the use of a threshold when we compare two models, one model with a parent or parents and one model without a parent. In the setting $s = 0.1$, our model can rebuild the target network more precisely, and the number of false positives decrease compared with the result of $s = 0.2$.

4.2. *Real data analysis*

In this section we show the effectiveness of our proposed method through the analysis of *Saccharomyces cerevisiae* gene expression data, which is newly obtained by disrupting 100 genes. Our research group has installed a systematic experimental method, which observes changes in the expression levels of genes on a microarray by gene disruption. By using this method, we have launched a project whose purpose is to reveal the gene regulatory networks between the 5871 genes of *Saccharomyces cerevisiae*. Many laboratories have also reported similar projects. We have already collected a large number of expression profiles from gene disruption experiments to evaluate genetic regulatory networks. Over 400 mutants are stocked and gene expression profiles are accumulating.

We monitored the transcriptional level of 5871 genes spotted on a microarray by a scanner. The expression profiles of over 400 disruptants were stored in our database. The standard deviation (SD) of the levels of all genes on a microarray was evaluated. The value of SD represents roughly the experimental error. In our data, we estimated the value of 0.5 as the critical point of the accuracy of experiments. We have evaluated the accuracy of those profiles on the base of the standard deviation of the expression ratio of all genes. 107 disruptants including 68 mutants where the transcription factors were disrupted could be selected from 400 profiles.

We used 100 microarrays and constructed a genetic network of 521 genes from the above data. The 94 transcription factors whose regulating genes have been clearly identified were found. The profiles of the 521 genes in control by those 94 factors were selected from 5871 profiles.

Bas1p and Bas2p also activate expression of three genes in the histidine biosynthesis pathway. In a gcn4 background, mutations that abolish the *BAS1* or *BAS2* function lead to a histidine auxotrophy. Previous investigation indicated that Bas1p and Bas2p are DNA binding proteins required for transcription of *HIS4* and these

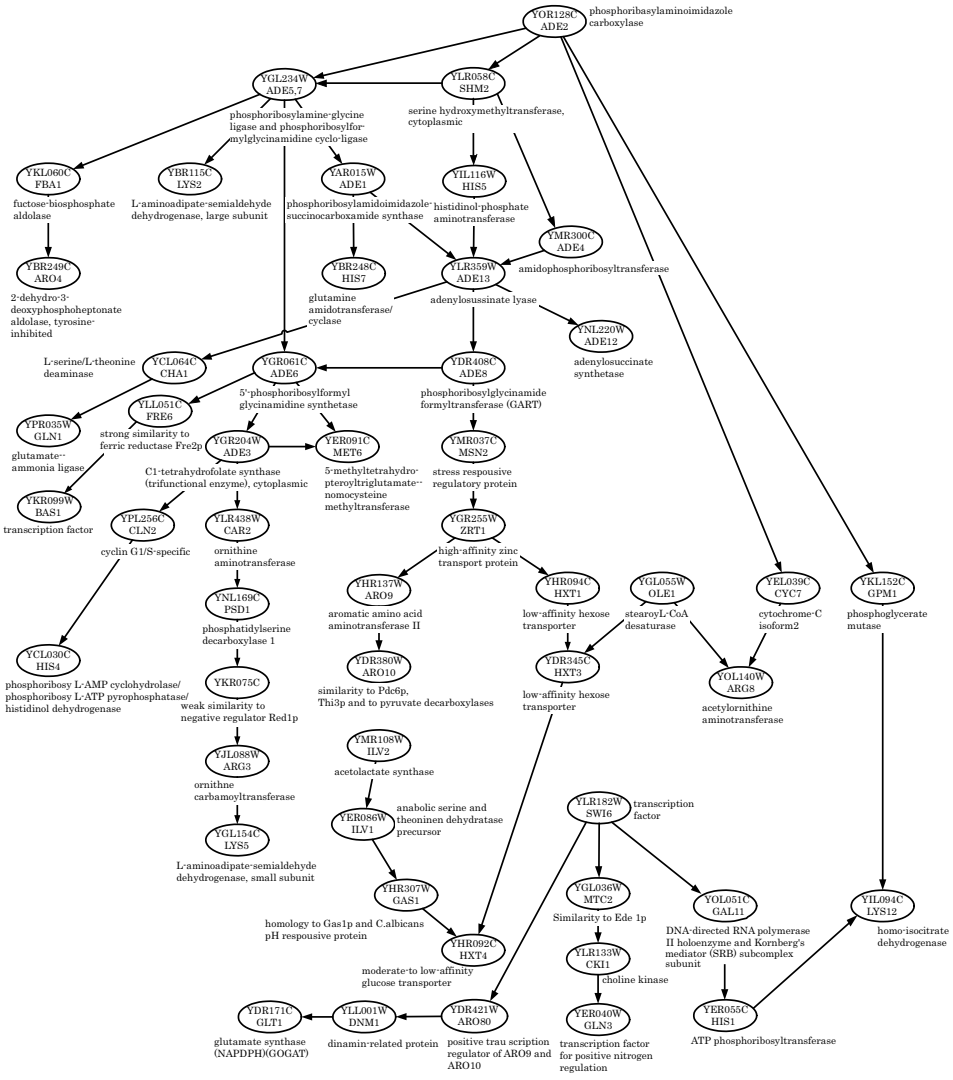

Fig. 5. The resulting partial network of the analysis of 521 *Saccharomyces cerevisiae* genes.

ADE genes like *GCN4*.[11,16,43] In this paper, we made clear that both genetic relation. Figure 5 indicates that those *ADE* genes and histidine biosynethesis genes are related with *BAS1* more directly than *GCN4*. Unfortunately, the direction of the edges between *BAS1* and *ADE* genes are reverse. We observe that, however, this direction is ambiguous and switch easily. In fact, the score of the model includes the edge from *BAS1* to *ADE* genes is almost equal to that of the selected model. In such a case, the model averaging technique[26] might be useful. The ribose component of purine ribonucleotides is derived from ribose 5-P, an inter mediate of the pentose phosphate cycle. The atoms of the base moiety are contributed by many compounds. They are added step wise to the preformed ribose. There exist striking interrelationships with the pathway for histidine synthesis.

Studies on the regulation of the purine biosynthesis pathway in *Saccharomyces cerevisiae* revealed that all the genes encoding enzymes required for AMP de novo biosynthesis are repressed at transcriptional level by the presence of extracellular purines. *ADE* genes are transcriptionally activated as well as some histidine biosynthesis genes. Especially the fact that expression of *HIS4* is related with *ADE* genes were known. In our regulated network, *HIS4* were related with some *ADE* genes closely, and some *HIS* genes are related with *ADE* genes like *HIS4*. The biosynthesis of the essential amino acid histidine shows in *Saccharomyces cerevisiae* shows close connection to purine metabolism, and our result satisfied this fact.

5. Conclusion

In this paper we proposed a new statistical method for estimating a genetic network from microarray gene expression data by using a Bayesian network and nonparametric regression. The key idea of our method is the use of nonparametric heteroscedastic regression models for capturing nonlinear relationships between genes and heteroscedasticity of the expression data. If we have a network that represents the causal relationship among genes, we can simulate the genetic system on the computer, e.g. Genomic Object Net.[39,40] In this stage, it is required that the relationships between genes are suitably estimated. In this sense, the proposed heteroscedastic model can give an essential improvement, because the previous models sometimes lead to unsuitable estimates of the systems. We consider the simulation of biological system as a future work.

An essential problem for network construction is the evaluation of the graph. We investigated this problem as a statistical model selection or evaluation problem and derived the new criterion for selecting graph from Bayes approach. Our method covers the previous methods for constructing genetic networks by using Bayesian networks and improves them in the theoretical and methodological senses. The proposed method successfully extracts the effective information and we can find these information in the resulting genetic network visually. We use the simple greedy algorithm for learning network. However, this algorithm needs much time for determining the optimal graph. Hence, the development of a better algorithm

is one of the important problems and we would like to discuss it in a future paper.

We showed the effectiveness of our method through the Monte Carlo simulations and the analysis of *Saccharomyces cerevisiae* gene expression data and evaluated the resulting network by comparing with biological knowledge. We construct the genetic network without using biological information. Nevertheless, the resulting network includes many important connections, which agree with the biological knowledge. Hence, we expect that our method can demonstrate its power in the analysis of a completely unknown system, like the human genome.

Acknowledgments

The authors would like to thank two referees for their helpful comments and suggestions.

References

1. H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds.), Akademiai Kiadó, Budapest, 267–281, 1973.
2. H. Akaike, "Likelihood and the Bayes procedure," in *Bayesian Statistics*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.), Univ. Press, Valencia, 143–203, 1980.
3. T. Akutsu, S. Miyano and S. Kuhara, "Identification of genetic networks from a small number of gene expression patterns under the Boolean network model," *Proc. Pacific Symposium on Biocomputing* **4**, 17–28 (1999).
4. T. Akutsu, S. Miyano and S. Kuhara, "Inferring qualitative relations in genetic networks and metabolic pathways," *Bioinformatics* **16**, 727–734 (2000).
5. T. Akutsu, S. Miyano and S. Kuhara, "Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function," *J. Comp. Biol.* **7**, 331–344 (2000).
6. T. Andou, S. Imoto and S. Konishi, "Estimating nonlinear regression models based on radial basis function networks (in Japanese with English Abstract)," *Japanese J. Appl. Statist.* **30**, 19–35 (2001).
7. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).
8. K. P. Burnham and D. R. Anderson, *Model Selection and Inference, a Practical Information–Theoretical Approach* (Springer-Verlag, New York, 1998).
9. G. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Mach. Learn.* **9**, 309–347 (1992).
10. R. Cowell, A. Dawid, S. Lauritzen and D. Spiegelhalter, *Probabilistic Networks and Expert Systems* (Springer-Verlag, New York, 1999).
11. B. Daignan-Fornier and G. R. Fink, "Coregulation of purine and histidine biosynthesis by the transcription activator BAS1 and BAS2," *Proc. Natl. Acad. Sci. USA* **89**, 6746–6750 (1992).
12. A. C. Davison, "Approximate predictive likelihood," *Biometrika* **73**, 323–332 (1986).
13. C. De Boor, *A Practical Guide to Splines* (Springer-Verlag, Berlin, 1978).
14. M. J. L. de Hoon, S. Imoto and S. Miyano, "Inferring gene regulatory networks from time-ordered gene expression data using differential equations," *Proc. 5th*

International Conference on Discovery Science, Lecture Note in Artificial Intelligence, Springer-Verlag, 267–274, 2002.
15. 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," *Proc. Pacific Symposium on Biocomputing* **8**, 17–28 (2003).
16. V. Denis, H. Boucherie, C. Monribot and B. Daignan-Fornier, "Role of the Myb-like protein Bas1p in *Saccharomyces cerevisiae*: a proteome analysis," *Mol. Microbiol.* **30**, 556–566 (1998).
17. B. P. Durbin, J. S. Hardin, D. M. Hawkins and D. M. Rocke, "A variance-stabilizing transformation from gene-expression microarray data," *Bioinformatics* **18**, Suppl. 1 (ISMB 2002), 105–110 (2002).
18. N. Friedman and M. Goldszmidt, "Discretizing continuous attributes while learning Bayesian networks," *Proc. 13th International Conference on Machine Learning*, 157–165 (1996).
19. N. Friedman and M. Goldszmidt, "Learning Bayesian networks with local structure," *Proc. 12th Conf. on Uncertainty in Artificial Intelligence*, 252–262 (1996).
20. N. Friedman and M. Goldszmidt, "Learning Bayesian networks with local structure," in *Learning and Inference in Graphical Models*, M. I. Jordan (ed.), Kluwer Academic Publisher, 1998.
21. N. Friedman, M. Linial, I. Nachman and D. Pe'er, "Using Bayesian network to analyze expression data," *J. Comp. Biol.* **7**, 601–620 (2000).
22. I. J. Good, *The Estimation of Probabilities* (MIT Press, Cambridge, Massachusetts, 1996).
23. P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models* (Chapman and Hall, 1994).
24. A. J. Hartemink, D. K. Gifford, T. S. Jaakkola and R. A. Young, "Maximum likelihood estimation of optimal scaling factors for expression array normalization," *SPIE BiOS 2001, San Jose, California*, 2001.
25. A. J. Hartemink, D. K. Gifford, T. S. Jaakkola and R. A. Young, "Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks," *Proc. Pacific Symposium on Biocomputing* **6**, 422–433 (2001).
26. A. J. Hartemink, D. K. Gifford, T. S. Jaakkola and R. A. Young, "Combining location and expression data for principled discovery of genetic regulatory network models," *Proc. Pacific Symposium on Biocomputing* **7**, 437–449 (2002).
27. T. Hastie and R. Tibshirani, *Generalized Additive Models* (Chapman and Hall, 1990).
28. D. Heckerman, " A tutorial on learning with Bayesian networks," in *Learning and Inference in Graphical Models*, M. I. Jordan (ed.) (Kluwer Academic Publisher, 1998).
29. D. Heckerman and D. Geiger, "Learning Bayesian networks: a unification for discrete and Gaussian domains," *Proc. Eleventh Conf. on Uncertainty in Artificial Intelligence*, 274–284 (1995).
30. D. Heckerman, D. Geiger and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learn.* **20**, 197–243 (1995).
31. T. Higuchi and G. Kitagawa, "Knowledge discovery and self-organizing state space model," *IEICE T. Inf. Syst.* **E83-D**, No. 1, 36–43 (2000).
32. W. Huber, A. von Heydebreck, H. Sueltmann, A. Poustka and M. Vingron, "Variance stablization applied to microarray data calibration and to quantification of differential expression," *Bioinformatics* **18**, Suppl. 1 (ISMB 2002), 96–104 (2002).
33. S. Imoto, T. Goto and S. Miyano, "Estimation of genetic networks and functional

structures between genes by using Bayesian networks and nonparametric regression," *Proc. Pacific Symposium on Biocomputing* **7**, 175–186 (2002).
34. F. V. Jensen," *An Introduction to Bayesian Networks* (University College London Press, 1996).
35. 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," *Proc. International Workshop on Computational Methods in Systems Biology*, Lecture Note in Computer Science, Springer-Verlag, 2003, in press.
36. G. Kitagawa, "Self-organizing state space model," *J. Am. Stat. Assoc.* **93**, 1203–1215 (1998).
37. S. Konishi, "Statistical model evaluation and information criteria," in *Multivariate Analysis, Design of Experiments and Survey Sampling*, S. Ghosh (ed.), Marcel Dekker, 1999.
38. S. Konishi and G. Kitagawa, "Generalised information criteria in model selection," *Biometrika* **83**, 875–890 (1996).
39. H. Matsuno, A. Doi, M. Nagasaki and S. Miyano, "Hybrid petri net representation of gene regulatory network," *Proc. Pacific Symposium on Biocomputing* **5**, 338–349 (2000).
40. H. Matsuno, A. Doi, Y. Hirata and S. Miyano, "XML documentation of Biopathways and their simulations in Genomic Object Net," *Genome Informatics* **12** 54–62 (2001).
41. K. Murphy and S. Mian, "Modelling gene expression data using dynamic Bayesian networks," *Technical report*, *Computer Science Division*, *University of California*, *Berkeley*, *CA*, 1999.
42. D. Pe'er, A. Regev, G. Elidan and N. Friedman, "Inferring subnetworks from perturbed expression profiles," *Bioinformatics* **17**, Suppl. 1 (ISMB 2001), 215–224 (2001).
43. R. J. Rolfes and A. G. Hinnebusch, "Translation of the yeast transcriptional activator GCN4 is stimulated by purine limitation: implications for activation of the protein kinase GCN2," *Mol. Cell Biol.* **13**, 5099–5111 (1993).
44. G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.* **6**, 461–464 (1978).
45. V. A. Smith, E. D. Jarvis and A. J. Hartemink, "Evaluating functional network inference using simulations of complex biological systems," *Bioinformatics* **18**, Suppl. 1 (ISMB 2002), 216–224 (2002).
46. L. Tinerey and J. B. Kadane, "Accurate approximations for posterior moments and marginal densities," *J. Am. Stat. Assoc.* **81**, 82–86 (1986).
47. H. Toh and K. Horimoto, "Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling," *Bioinformatics* **18**, 287–297 (2002).

Seiya Imoto is currently a research associate of the laboratory of DNA analysis, Human Genome Center, Institute of Medical Science, University of Tokyo. He received B.S., M.S., and Ph.D. in Mathematics from Kyushu University in 1996, 1998 and 2001, respectively. His current research interests cover analysis of high dimensional data with complex structure by using nonlinear statistical methods as well as development model selection criteria from an information theoretic or a Bayesian statistics approach, and analysis of cDNA microarray gene expression data.

Sunyong Kim is currently a master course student of the laboratory of DNA analysis, Human Genome Center, Institute of Medical Science, University of Tokyo. She received her B.S. in Computer Science from the University of Tokyo in 2002. Her current research interests include development of computational methods/software and establishment of statistical models for inferring gene regulatory pathways.

Takao Goto is currently a student of the Faculty of Medicine, University of Tokyo. He was graduated from the Department of Information of the University of Tokyo in 1999 and received B.S. in Computer Science. His current research interests cover the analysis of gene networks from cDNA microarray gene expression data.

Sachiyo Aburatani is currently an assistant professor of Laboratory of Biostatistics at Human Genome Center, Institute of Medical Science, University of Tokyo. She received her Ph.D. degree in Agriculture from Kyushu University in 2003. She has 5 peer-reviewed publications. Her current research interests cover the inference of interaction between biological molecules by combining information from various disparate resources.

Kousuke Tashiro is currently an associate professor in Laboratory of Molecular Technics, Department of System Biology at Kyushu University. He received his Ph.D. degree in molecular biology from Kyushu University in 1981. He has more than 70 reviewed publications. His current research interests are the analyses of the mechanisms the gene regulatory networks in various biological phenomena such as cell proliferation, cell differentiation and cell adaptation against outer circumstances, using the gene expression profiles obtained by DNA-chip and the computational methods.

Satoru Kuhara is currently a professor of Faculty of Agriculture Kyushu University. He received his Ph.D. degree in Biochemistry from the Kyushu University in 1980. He has more than 60 peer reviewed publications. His current research interests cover modeling of expression control of cells based on microarray experiments, modeling of developmental process in gene expression and generalization of relationship between structure and function of proteins.

Satoru Miyano is a Professor of Human Genome Center, Institute of Medical Science, University of Tokyo. He obtained B.S. in 1977, M.S. in 1979, and Ph.D. in Mathematics from Kyushu University. His current research interests include computational gene network inference methods, modeling and simulation of biological systems, and computational knowledge discovery. He is on the Editorial Board of *Bioinformatics*, *Journal Bioinformatics and Computational Biology*, *Theoretical Computer Science* and the Chief Editor of *Genome Informatics*.