Chem-Bio Informatics Journal, Vol. 4, No. 1, pp.1-14(2004)

Inference of S-system Models of Genetic Networks from Noisy Time-series Data

Shuhei Kimura*, Mariko Hatakeyama and Akihiko Konagaya

RIKEN Genomic Sciences Center,
1-7-22 Suehiro-cho, Tsurumi, Yokohama, 230-0045, Japan

**E-mail: skimura@gsc.riken.jp*

(Received December 24, 2003; accepted February 26, 2004; published online March 31, 2004)

Abstract

In this paper, we propose a new method for the inference of S-system models of large-scale genetic networks from the observed time-series data of gene expression patterns. The proposed method employs a technique to decompose the genetic network inference problem into several subproblems. The S-system parameters are estimated by solving these decomposed subproblems. In addition, the proposed method estimates the initial levels of the gene expression. The estimation of the initial gene expression levels is necessary when the noisy time-series data are given. We verify the effectiveness of the proposed method through the genetic network inference problems.

Key Words: Genetic network, S-system, DNA microarray, Real-coded GA, Problem decomposition

Area of Interest: Genome Wide Experimental Data Analysis

1. Introduction

Advancements in technologies such as DNA microarrays now allow us to measure gene expression patterns on a genomic scale [1]. A great number of researchers have taken an interest in the inference of underlying genetic networks from the observed time-series data of gene expression patterns, and this has become one of the major topics in the bioinformatics field [2]. Numerous models have been proposed to describe networks, and numerous algorithms based on individual models have been proposed for the inference of genetic networks [2][3][4][5][6][7][19].

The S-system model is considered an ideal choice for inferring genetic networks, as the model is rich enough in structure to capture various dynamics and some methods are available for analyzing it [8][9]. The S-system model is a set of non-linear differential equations of the form

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{N} X_j^{g_{i,j}} - \beta_i \prod_{j=1}^{N} X_j^{h_{i,j}} , \quad (i = 1, \cdots, N), \tag{1}$$

where X_i is the state variable and *N* is the number of components in the network. In a genetic network, X_i is the expression level of the *i*-th gene and *N* is the number of genes in the network. α_i and β_i are multiplicative parameters called rate constants, and $g_{i,j}$ and $h_{i,j}$ are exponential parameters called kinetic orders.

Several network inference algorithms based on the S-system model have been proposed [8][9][10]. These algorithms estimate all of the S-system parameters ($\alpha_i, \beta_i, g_{i,j}$ and $h_{i,j}$) from observed time-series data of the gene expression patterns simultaneously. Since the number of S-system parameters is proportional to the square of the number of network components, the algorithms must estimate a large number of S-system parameters when we try to infer large-scale network systems containing many network components. This is why inference algorithms based on the S-system model have been applied only to small-scale networks of less than five genes. For the purpose of resolving the high-dimensionality of the genetic network inference problem in the S-system model, a problem decomposition strategy, that divides the original problem into several subproblems, has been proposed [11][12].

The problem decomposition approach enables us to infer S-system models of larger-scale genetic networks. However, existing inference algorithms are still incapable of inferring realistic genetic networks, as they were designed without considering the noise in the observed time-series data. In this paper, we extend the problem decomposition approach to a realistic noisy environment. The proposed method estimates the initial levels of the gene expression, as well as the S-system parameters, in order to enhance the probability of finding the correct interaction between genes. The initial gene expression level and the set of S-system parameters are estimated alternately in each decomposed subproblem. We verify the effectiveness of the proposed method by applying it to the genetic network inference problems containing 5 and 30 genes, respectively.

2. Genetic Network Inference Problem

2.1 Canonical problem definition

The canonical genetic network inference problem is defined as a function optimization problem to minimize the following sum of the squared relative error [8][9].

$$f = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{X_{\mathrm{cal},i,t} - X_{\mathrm{exp},i,t}}{X_{\mathrm{exp},i,t}} \right)^2 , \qquad (2)$$

where $X_{\mathrm{exp},i,t}$ is an experimentally observed gene expression level at time *t* of the *i*-th gene, $X_{\mathrm{cal},i,t}$ is a numerically calculated one acquired by solving a system of differential equations (1), *N* is the number of genes in the network, and *T* is the number of sampling points of observed data.

Since *2N(N+1)* S-system parameters must be determined in order to solve the set of differential equations (1), this function optimization problem is *2N(N+1)* dimensional. This is too many dimensions for non-linear function optimizers in cases where the algorithms are used to infer S-system models of large-scale genetic networks containing many network components [11][12].

2.2 Problem decomposition

As mentioned shortly before, the high dimensionality of large-scale network systems makes these systems difficult to infer using algorithms based on the S-system model. Very recently, a

problem decomposition strategy has been proposed as a means for resolving this high dimensionality problem [11][12].

The problem decomposition strategy divides the genetic network inference problem into several subproblems, each of which corresponds to each gene. The objective function of the subproblem corresponding to the *i*-th gene is

$$f_i = \sum_{t=1}^{T} \left(\frac{X_{\mathrm{cal},i,t} - X_{\mathrm{exp},i,t}}{X_{\mathrm{exp},i,t}} \right)^2 , \tag{3}$$

where $X_{\mathrm{cal},i,t}$ is the numerically calculated gene expression level at time *t* of the *i*-th gene, just as described in the previous subsection. In contrast to the previous subsection, however, $X_{\mathrm{cal},i,t}$ is acquired by solving the following differential equation.

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{N} Y_j^{g_{i,j}} - \beta_i \prod_{j=1}^{N} Y_j^{h_{i,j}} , \tag{4}$$

where

$$Y_j = \begin{cases} X_j , & \text{if } j = i, \\ \hat{X}_j , & \text{otherwise.} \end{cases} \tag{5}$$

$\hat{X}_j$ is an estimated gene expression level obtained not by solving any differential equation, but by making a direct estimation from the observed time-series data. Two methods are used to estimate $\hat{X}_j$ in this paper. When the data are assumed to be observed with no measurement error, $\hat{X}_j$ is estimated using the spline interpolation [13] of the observed gene expression level. When the absence of measurement error cannot be confirmed, the value is estimated using local linear regression [14], a data smoothing technique.

The equation (4) is solvable when *2(N+1)* S-system parameters (i.e., $\alpha_i , \beta_i , g_{i,1}, \cdots, g_{i,N} , h_{i,1}, \cdots, h_{i,N}$) are given. Thus, this decomposition strategy divides a *2N(N+1)* dimensional network inference problem into *N* individual *2(N+1)* dimensional subproblems. The solutions of the equation (4) ($i = 1, \cdots, N$) completely coincide with the solution of the set of equations (1) only when accurate curves are given as the observed gene expression patterns.

2.3 Use of a priori knowledge

The genetic network inference problem based on the S-system model may have multiple optima due to the high degree-of-freedom and pollution of the time-series data by the measurement error. In this study, we try to improve our chances of finding a correct solution by introducing a priori knowledge about the genetic network into the objective function.

The connectivity of the genetic network has been demonstrated to be sparse [15]. When there is no interaction between two genes, the S-system parameter values corresponding to the interaction ($g_{i,j}$ and $h_{i,j}$) are zero. Kikuchi and his colleagues incorporated this knowledge into the objective function using a penalty term called the pruning term [9]. This turns out to be an imperfect solution, however, as the pruning term pushes all the kinetic orders down to zero, a condition that may make prevent the model from finding the existing interactions. To avoid this, we introduce another penalty term into the objective function (3), as follows [12].

$$F_i = \sum_{t=1}^{T} \left(\frac{X_{\mathrm{cal},i,t} - X_{\mathrm{exp},i,t}}{X_{\mathrm{exp},i,t}} \right)^2 + c \sum_{j=1}^{N-I} \left(\left| G_{i,j} \right| + \left| H_{i,j} \right| \right), \tag{6}$$

where $G_{i,j}$ and $H_{i,j}$ are given by rearranging $g_{i,j}$ and $h_{i,j}$, respectively, in descending order of their absolute values (i.e., $\left|G_{i,1}\right| \le \left|G_{i,2}\right| \le \cdots \le \left|G_{i,N}\right|$ and $\left|H_{i,1}\right| \le \left|H_{i,2}\right| \le \cdots \le \left|H_{i,N}\right|$). N is the number of genes in the network and I is a maximum indegree. The maximum indegree determines the maximum number of genes that affect the i-th gene directly. c is a penalty coefficient.

The first term on the right-hand side of the equation (6) is the same as that of the equation (3). The second term on the r.h.s. of the equation (6) is a penalty term that forces most of the kinetic orders ($g_{i,j}$ and $h_{i,j}$) down to zero, thereby causing most of the genes to disconnect when this penalty term is applied. This term will not penalize, however, when the number of genes that directly affect the i-th gene is lower than the maximum indegree I. Thus, the optimum solutions to the objective functions (3) and (6) are the same when the number of genes that affect the focused (i-th) gene is lower than the maximum indegree. In this paper, we use the equation (6) as the objective function that should be minimized.

3. Estimation of Initial Gene Expression Level

When we try to solve the decomposed subproblem, the differential equation (4) must be solved. In order to solve the differential equation (4), we need to input the initial expression level of the gene (the initial state value for the differential equation), in addition to the S-system parameters. The initial gene expression level is obtainable from the observed time-series data if they contain no measurement error. However, since the data are generally polluted by the measurement noise when the gene expression patterns are actually measured, the initial gene expression level should be estimated together with the S-system parameters.

As mentioned just above, the initial gene expression level needs to be estimated when the inference algorithm is applied to a realistic genetic network inference problem. However, the simultaneous estimation of the initial gene expression level and the S-system parameters makes the function optimization problem higher dimensional, and this is inconvenient for function optimizers. Therefore, we estimate the initial gene expression level and the set of the S-system parameters alternately, i.e., when the initial expression level of the i-th gene is estimated, the S-system parameters are fixed to the values of some candidate solution for the i-th subproblem. In this study, the fixed S-system parameter values are obtained from the best candidate solution that has ever been found. Since the initial expression level of the i-th gene is a unique variable and the rest of the model parameters are fixed, the estimation of the initial expression level of the i-th gene is formulated as a single dimensional function minimization problem. The objective function of this estimation problem is

$$F_i^{\mathrm{adj}} = \sum_{t=1}^{T} \gamma^{t-1} \left(\frac{X_{\mathrm{cal},i,t} - X_{\mathrm{exp},i,t}}{X_{\mathrm{exp},i,t}} \right)^2, \tag{7}$$

where $X_{\mathrm{cal},i,t}$ is acquired by solving the equation (4), and $\gamma\,(0 \le \gamma \le 1)$ is a discount parameter. As the fixed S-system parameters obtained from the best candidate are not always optimal, the calculated gene expression curve may differ greatly from the actual curve. When the calculated curve is incorrect, the algorithm should not be able to fit the curve, especially the latter half of the curve, into the observed data. Therefore, in this study, we introduce the discount parameter γ.

When the noisy time-series data are given as the observed gene expression patterns, we should estimate both the set of the S-system parameters and the initial gene expression level. To estimate them, we minimize the objective functions (6) and (7) alternately in this study.

4. Method for the Inference of Genetic Networks

In this section, we propose a new method for the inference of S-system models of genetic networks.

4.1 Parameter estimators

In the proposed method, different parameter estimators, described below, are used to optimize the objective functions (6) and (7), respectively. The proposed method applies these parameter estimators in order to infer genetic networks.

4.1.1 Parameter estimator for the S-system parameters

In order to estimate the S-system parameters, we must optimize the objective function (6) mentioned in the section 2. Any type of function optimizer can be applied to this problem. In this study, we use GLSDC (a Genetic Local Search with distance independent Diversity Control) [16], a method based on the real-coded genetic algorithm. GLSDC has two features important for the inference of large-scale genetic networks: it works well on multimodal function optimization problems with high-dimensionality and can be suitably applied to parallel computation. GLSDC uses the modified Powell's method [13] as a local search operator. Two genetic operators, ENDX (an Extended Normal Distribution Crossover) [16][17] and MGG (Minimal Generation Gap model) [18] are also used in GLSDC. The following is an algorithm of GLSDC.

[Algorithm: GLSDC]

1. [Initialize] As an initial population, create n_p individuals randomly. Set Generation=0 and the iteration number of the converging operations $N_{\text{iter}} = N_0$.
2. [Local Search Phase] Apply the modified Powell's method to all individuals in the population.
3. [Adaptation of N_{iter}] If the best individual in the population shows no improvement over the previous generation, set $N_{\text{iter}} \leftarrow N_{\text{iter}} + N_0$. Otherwise, set $N_{\text{iter}} = N_0$.
4. [Converging Phase] Execute the exchange of individuals according to the genetic operators, i.e., ENDX and MGG, N_{iter} times. There is no need to calculate the fitness value at this time.
5. [Termination] Stop if the halting criteria are satisfied. Otherwise, Generation $\leftarrow$ Generation+ 1 and go to step 2.

Readers can find more detailed information on GLSDC in [16].

As mentioned in the section 3, the initial expression level of the *i*-th gene is required when GLSDC solves the *i*-th subproblem. The initial level is fixed to the value obtained by minimizing the objective function (7), while GLSDC is solving this problem.

4.1.2 Parameter estimator for the initial gene expression level

The initial gene expression level is estimated by minimizing the objective function (7) mentioned in the section 3. When this problem is solved, the rest of the model parameters are fixed

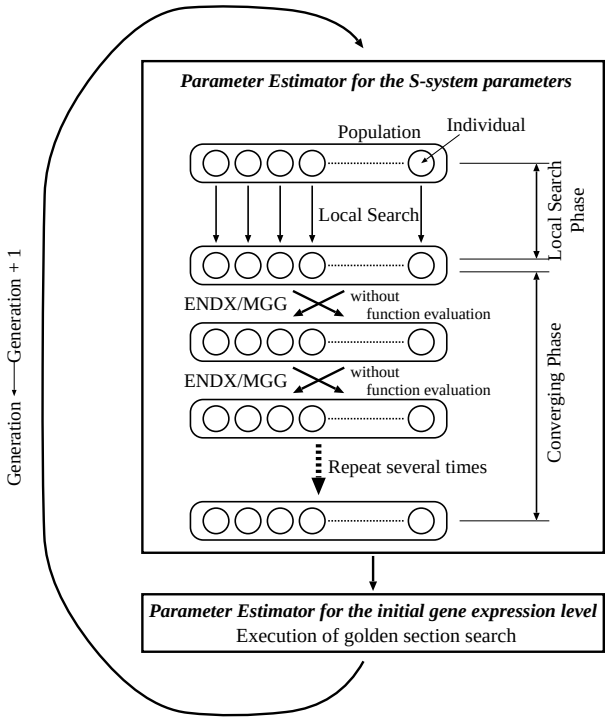

Figure 1. The framework of the proposed method

to the best candidate solution that has ever been found by GLSDC. As this problem is one dimensional, we attempt to solve it with the use of the golden section search [13]. When multiple sets of time-series data are given as the observed data, the one-dimensional search is applied to all of the sets.

4.2 Algorithm

The simultaneous estimation of the set of the S-system parameters and the initial gene expression level makes the function optimization problem higher dimensional, and this is inconvenient for function optimizers. Therefore, we estimate them alternately. On the basis of this idea, we propose a new method for the inference of genetic networks. In the proposed method, the two parameter estimators described above are applied alternately, i.e., the estimation of the initial gene expression level is performed at the end of the every cycle (generation) of GLSDC (see Figure 1).

When GLSDC estimates the S-system parameters, the initial gene expression level is fixed to the value obtained by minimizing the objective function (7), as mentioned above. However, the initial level is unavailable at the first generation, because it is estimated after the estimation of the S-system parameters. Therefore, only at the first generation, we use the value obtained directly from the observed time-series data as the initial gene expression level.

5. Numerical Experiments

We conducted a series of numerical experiments to demonstrate the effectiveness of our proposed method. Lacking any actual biological data, we used artificial genetic network models as a case study in this paper.

Table 1. S-system parameters of the target model A

i	α_i	$g_{i,1}$	$g_{i,2}$	$g_{i,3}$	$g_{i,4}$	$g_{i,5}$	β_i	$h_{i,1}$	$h_{i,2}$	$h_{i,3}$	$h_{i,4}$	$h_{i,5}$
1	5.0	0.0	0.0	1.0	0.0	-1.0	10.0	2.0	0.0	0.0	0.0	0.0
2	10.0	2.0	0.0	0.0	0.0	0.0	10.0	0.0	2.0	0.0	0.0	0.0
3	10.0	0.0	-1.0	0.0	0.0	0.0	10.0	0.0	-1.0	2.0	0.0	0.0
4	8.0	0.0	0.0	2.0	0.0	-1.0	10.0	0.0	0.0	0.0	2.0	0.0
5	10.0	0.0	0.0	0.0	2.0	0.0	10.0	0.0	0.0	0.0	0.0	2.0

Table 2. 15 sets of the initial gene expression levels used in the experiment of the target model A

Set	X_1	X_2	X_3	X_4	X_5
1	1.655967e+00	1.868416e+00	1.032173e-01	2.730268e-01	1.562687e+00
2	7.862766e-01	5.474855e-01	9.287958e-01	3.894443e-01	9.344040e-01
3	3.468547e-01	1.994981e+00	1.532913e+00	1.761393e+00	1.264981e+00
4	8.020131e-01	8.949262e-01	3.135082e-01	7.610533e-02	1.269706e+00
5	9.590725e-01	2.805737e-01	5.507401e-01	1.694232e+00	5.744767e-01
6	3.992936e-01	1.849408e+00	2.912736e-01	1.144217e+00	9.988814e-01
7	1.055713e-02	5.114093e-02	8.495855e-01	1.740444e+00	1.969969e-01
8	1.489803e+00	9.168820e-01	1.707836e+00	1.827741e+00	2.824051e-01
9	1.842769e-01	1.589055e+00	6.668454e-01	4.727903e-01	1.265678e+00
10	1.285646e+00	8.995862e-01	1.994967e-01	8.811659e-01	1.723054e+00
11	1.336863e-01	4.233753e-01	4.168260e-01	4.823942e-01	5.539923e-01
12	1.652500e+00	1.744966e+00	3.904404e-01	1.584671e+00	4.339247e-01
13	1.562800e+00	1.164151e+00	1.391469e+00	6.808265e-01	1.090292e+00
14	3.271505e-01	1.147837e+00	1.576167e-01	8.645541e-01	2.591408e-01
15	5.522177e-01	4.220327e-01	1.084436e+00	1.994388e+00	1.050098e+00

5.1 Experiment 1: small-scale network inference with noise-free data

In this experiment, we confirm that the proposed method has an ability to infer a correct S-system model of the genetic network when a sufficient amount of noise-free data is given.

5.1.1 Experimental setup

As a target network, we used the S-system model of the small-scale genetic network consisting of 5 genes (N=5) [9]. We call this model the target model A, and list its parameters in Table 1. By applying the problem decomposition strategy mentioned in the section 2.2, this genetic network inference problem was decomposed into 5 subproblems.

If an insufficient amount of time-series data is given as observed gene expression patterns, the high degree-of-freedom of S-system models ensures that many candidate solutions may be found. To enhance the probability of finding the correct solution, we used 15 sets of time-series data, each covering all 5 genes, as a sufficient amount of observed gene expression data. The sets of time-series data were obtained by solving the set of differential equations (1) on the target model A. The initial values of these sets are generated randomly, and they are listed in Table 2. In a practical application, these sets of time-series data could be obtained by actual biological experiments of different experimental conditions. 11 sampling points for the time-series data were assigned on each gene in each set, hence the observed time-series data on each gene consisted of $15 \times 11 = 165$

Table 3. Estimated S-system parameters for the target model A

i	α_i	$g_{i,1}$	$g_{i,2}$	$g_{i,3}$	$g_{i,4}$	$g_{i,5}$	β_i	$h_{i,1}$	$h_{i,2}$	$h_{i,3}$	$h_{i,4}$	$h_{i,5}$
1	4.802	-0.025	-0.002	1.017	-0.002	-1.017	9.790	2.032	-0.005	-0.006	0.001	0.005
2	10.045	1.982	0.008	0.001	0.005	-0.001	10.044	0.002	1.989	-0.001	0.002	0.005
3	10.081	-0.002	-0.991	0.009	-0.003	-0.003	10.078	-0.002	-0.991	1.994	-0.007	-0.002
4	8.086	-0.010	0.004	1.991	0.005	-0.990	10.118	0.000	0.003	0.013	1.978	0.008
5	9.478	0.004	0.010	0.034	2.066	-0.087	9.505	0.007	0.012	0.009	-0.031	2.024

Table 4. Estimated initial gene expression levels for the target model A

Set	X_1	X_2	X_3	X_4	X_5
1	1.657616e+00	1.866675e+00	1.031946e-01	2.729209e-01	1.562868e+00
2	7.863174e-01	5.475139e-01	9.290369e-01	3.897167e-01	9.338474e-01
3	3.466258e-01	1.995005e+00	1.532000e+00	1.760710e+00	1.264359e+00
4	8.014521e-01	8.953308e-01	3.135446e-01	7.614276e-02	1.272565e+00
5	9.596057e-01	2.805703e-01	5.509538e-01	1.693683e+00	5.737853e-01
6	3.990045e-01	1.849312e+00	2.913377e-01	1.144084e+00	1.002049e+00
7	1.056477e-02	5.135667e-02	8.495751e-01	1.745016e+00	1.969867e-01
8	1.491036e+00	9.191003e-01	1.707857e+00	1.825848e+00	2.827157e-01
9	1.843174e-01	1.588871e+00	6.669493e-01	4.722400e-01	1.262910e+00
10	1.284880e+00	9.005177e-01	1.995071e-01	8.813032e-01	1.719754e+00
11	1.336483e-01	4.230248e-01	4.169177e-01	4.828245e-01	5.532325e-01
12	1.652757e+00	1.743340e+00	3.903545e-01	1.582928e+00	4.327569e-01
13	1.563244e+00	1.163774e+00	1.392009e+00	6.811343e-01	1.090419e+00
14	3.270106e-01	1.147585e+00	1.576615e-01	8.634238e-01	2.591978e-01
15	5.520035e-01	4.222403e-01	1.084154e+00	1.991653e+00	1.050043e+00

sampling points. Spline interpolation was used to obtain $\hat{X}_j$'s. Spline interpolated curves of these observed data were used as $\hat{X}_j$'s in the equation (5), as mentioned in the section 2.2.

We used the following recommended parameters in GLSDC applied here [16]: the population size n_p is 3*n*, where *n* is the dimension of the search space; the number of children generated by the crossover per selection is 10; and the number of applying the converging operations N_0 is $2n_p$. Five runs were carried out for each subproblem, and each run was continued until the number of fitness evaluations reached 1.0×10^6. The search regions of the S-system parameters were [0.0, 20.0] for α_i and β_i, and [-3.0, 3.0] for $g_{i,j}$ and $h_{i,j}$. The search regions of the initial gene expression levels were set to $\pm30\%$ of the observed ones (i.e., $[0.7X_{\exp,i,0}, 1.3X_{\exp,i,0}]$). The maximum indegree *I* was 5, the penalty coefficient *c* was 1.0, and the discount parameter γ was 0.75.

The structure skeletalizing technique was introduced in order to reduce the computational cost [8]. This technique assigns a value of zero to the kinetic orders ($g_{i,j}$ and $h_{i,j}$) whose absolute values are less than a given threshold δ_s. Structure skeletalizing effectively reduces the computational cost, because the exponential calculation of the equation (4) is omissible when the kinetic orders are zero. In this paper, we used the threshold $\delta_s = 1.0\times10^{-3}$.

As this network model contains 5 genes, we have to estimate $2\times5\times(5+1)=60$ S-system parameters in order to infer the network. In addition, we have to estimate all of the initial levels of

the gene expression, which total $5 \times 15 = 75$. Therefore, 60 + 75 = 135 parameters must be estimated in this problem.

5.1.2 Results

The S-system parameters and the initial gene expression levels estimated by the proposed method are listed in Tables 3 and 4, respectively. As can be seen from the tables, our method was unable to estimate the parameter values with perfect precision. However, they were precise enough to biologically interpret the network.

Our method running on a single-CPU personal computer (Pentium III 1GHz) required about 58.8 minutes to optimize each subproblem. This is far less computing time, from a comparison with the method earlier (PEACE1; Predictor by Evolutionary Algorithms and Canonical Equations 1) [9]. PEACE1 running on a PC cluster (Pentium III 933MHz$\times$1040CPUs) reportedly took more than 10 hours to estimate the S-system parameters.

In this experiment, the effectiveness of the proposed method was confirmed by estimating both the S-system parameters and the initial gene expression levels. In practical, however, it is not necessary to estimate the initial gene expression levels when the observed data seem to contain no measurement error. When the initial gene expression levels do not need to be estimated, the estimated parameters are more precise since the problem contains fewer unknown parameters.

5.2 Experiment 2: large-scale network inference with noisy data

Next, we test the performance of our method in a noisy real-world setting by conducting the experiment with the sets of noisy time-series data.

5.2.1 Experimental setup

The larger-scale S-system model containing 30 genes (N=30) was used as a target model here. The network structure and the S-system parameters of the model are given in Figure 2 and Table 5, respectively [3]. This model is referred to as the target model B in this paper. The problem decomposition strategy divided the original inference problem into 30 subproblems.

The observed gene expression data included 20 sets of time-series data, each covering all 30 genes. The sets of time-series data began from randomly generated initial values in [0.0,2.0], and were obtained by solving the set of differential equations (1) on the target model B. We added 10% Gaussian noise to the time-series data in order to simulate the measurement noise that often corrupts the observed data obtained from actual measurements of gene expression patterns. 11 sampling points for the time-series data were assigned on each gene in each set. Local linear regression [14] was used to obtain $\hat{X}_j$'s. Figure 3 shows an example of the estimated gene expression curve $\hat{X}_j$ used in this experiment.

Five runs were carried out for each subproblem. Each run was continued until the number of fitness evaluations reached 4.0×10^7. The search regions were [0.0, 3.0] for α_i and β_i, [-3.0, 3.0] for $g_{i,j}$ and $h_{i,j}$, and $[0.7X_{\exp,i,0}, 1.3X_{\exp,i,0}]$ for the initial gene expression levels. All of the other experimental conditions were the same as those in the previous subsection.

In this experiment, $2 \times 30 \times (30+1) = 1860$ S-system parameters and $30 \times 20 = 600$ levels of initial gene expression should be estimated.

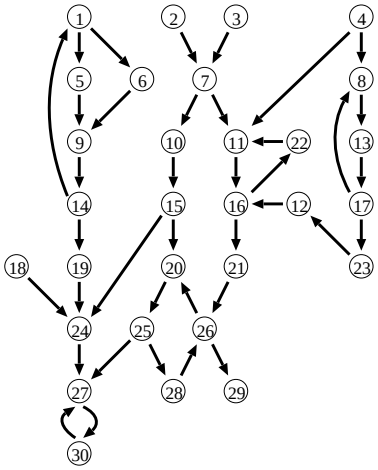

Figure 2. Network structure of the target model B

Table 5. S-system parameters of the target model B

α_i	1.0
β_i	1.0
$g_{i,j}$	$g_{1,14}=-0.1$, $g_{5,1}=1.0$, $g_{6,1}=1.0$, $g_{7,2}=0.5$, $g_{7,3}=0.4$, $g_{8,4}=0.2$, $g_{8,17}=-0.2$, $g_{9,5}=1.0$, $g_{9,6}=-0.1$, $g_{10,7}=0.3$, $g_{11,4}=0.4$, $g_{11,7}=-0.2$, $g_{11,22}=0.4$, $g_{12,23}=0.1$, $g_{13,8}=0.6$, $g_{14,9}=1.0$, $g_{15,10}=0.2$, $g_{16,11}=0.5$, $g_{16,12}=-0.2$, $g_{17,13}=0.5$, $g_{19,14}=0.1$, $g_{20,15}=0.7$, $g_{20,26}=0.3$, $g_{21,16}=0.6$, $g_{22,16}=0.5$, $g_{23,17}=0.2$, $g_{24,15}=-0.2$, $g_{24,18}=-0.1$, $g_{24,19}=0.3$, $g_{25,20}=0.4$, $g_{26,21}=-0.2$, $g_{26,28}=0.1$, $g_{27,24}=0.6$, $g_{27,25}=0.3$, $g_{27,30}=-0.2$, $g_{28,25}=0.5$, $g_{29,26}=0.4$, $g_{30,27}=0.6$, other $g_{i,j}=0.0$
$h_{i,j}$	1.0, if $i=j$, 0.0, otherwise

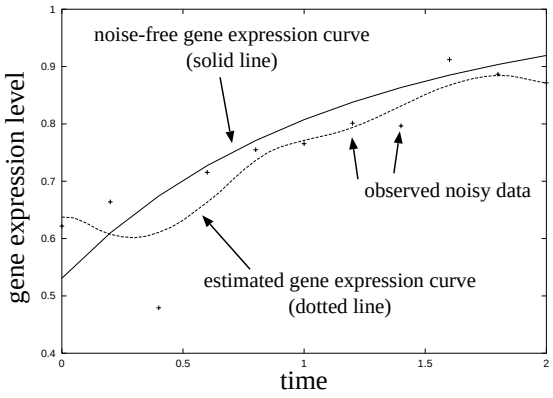

Figure 3. Example of the estimated gene expression curve

The figure shows a gene expression curve estimated by the local linear regression (dotted line), a noise-free curve calculated on the target model B (solid line), and given noisy data.

Table 6. Estimated S-system parameters for the target model B

α_1 = 0.484,	$g_{1,1}$	= -0.916,	$g_{1,16}$	= 0.571,	$g_{1,20}$	= 0.367,	$g_{1,23}$	= 0.479,	$g_{1,25}$	= 0.396,	β_1 = 0.437,	$h_{1,1}$	= 2.552,	$h_{1,2}$	= 0.315,	$h_{1,16}$	= 0.686,	$h_{1,25}$	= 0.675,	$h_{1,25}$	= 0.211
α_2 = 0.510,	$g_{2,2}$	= -0.465,	$g_{2,16}$	= 0.131,	$g_{2,19}$	= 0.235,	$g_{2,20}$	= 0.215,	$g_{2,23}$	= 0.153,	β_2 = 0.492,	$h_{2,2}$	= 1.486,	$h_{2,3}$	= -0.177,	$h_{2,6}$	= 0.351,	$h_{2,16}$	= 0.225,	$h_{2,19}$	= 0.320
α_3 = 0.527,	$g_{3,1}$	= 0.322,	$g_{3,3}$	= -0.461,	$g_{3,13}$	= 0.216,	$g_{3,20}$	= 0.069,	$g_{3,27}$	= -0.089,	β_3 = 0.529,	$h_{3,3}$	= 1.252,	$h_{3,6}$	= -0.117,	$h_{3,13}$	= 0.347,	$h_{3,14}$	= -0.058,	$h_{3,25}$	= 0.200
α_4 = 0.360,	$g_{4,2}$	= 0.247,	$g_{4,4}$	= -0.629,	$g_{4,10}$	= 0.231,	$g_{4,14}$	= 0.153,	$g_{4,30}$	= 0.369,	β_4 = 0.362,	$h_{4,4}$	= 2.401,	$h_{4,15}$	= -0.289,	$h_{4,19}$	= -0.276,	$h_{4,27}$	= 0.491,	$h_{4,30}$	= 0.279
α_5 = 0.492,	$g_{5,1}$	= 1.785,	$g_{5,2}$	= -0.300,	$g_{5,5}$	= -0.448,	$g_{5,12}$	= 0.103,	$g_{5,15}$	= 0.129,	β_5 = 0.493,	$h_{5,1}$	= -0.526,	$h_{5,5}$	= 1.525,	$h_{5,11}$	= -0.242,	$h_{5,16}$	= 0.271,	$h_{5,29}$	= 0.212
α_6 = 0.490,	$g_{6,1}$	= 1.118,	$g_{6,6}$	= -0.353,	$g_{6,14}$	= -0.157,	$g_{6,17}$	= 0.378,	$g_{6,30}$	= -0.172,	β_6 = 0.490,	$h_{6,1}$	= -0.559,	$h_{6,4}$	= -0.074,	$h_{6,6}$	= 1.475,	$h_{6,13}$	= -0.185,	$h_{6,30}$	= -0.315
α_7 = 0.798,	$g_{7,2}$	= 0.663,	$g_{7,3}$	= 0.187,	$g_{7,4}$	= -0.112,	$g_{7,7}$	= -0.090,	$g_{7,9}$	= -0.117,	β_7 = 0.850,	$h_{7,3}$	= -0.297,	$h_{7,7}$	= 1.173,	$h_{7,10}$	= 0.092,	$h_{7,13}$	= 0.179,	$h_{7,30}$	= 0.120
α_8 = 0.374,	$g_{8,4}$	= 0.540,	$g_{8,8}$	= -0.684,	$g_{8,13}$	= 0.390,	$g_{8,15}$	= 0.393,	$g_{8,30}$	= 0.331,	β_8 = 0.345,	$h_{8,2}$	= 0.413,	$h_{8,8}$	= 2.164,	$h_{8,16}$	= -0.230,	$h_{8,18}$	= 0.329,	$h_{8,27}$	= 0.163
α_9 = 0.536,	$g_{9,5}$	= 1.490,	$g_{9,9}$	= -0.475,	$g_{9,10}$	= 0.140,	$g_{9,21}$	= 0.279,	$g_{9,28}$	= 0.223,	β_9 = 0.578,	$h_{9,5}$	= -0.238,	$h_{9,9}$	= 1.468,	$h_{9,22}$	= 0.243,	$h_{9,23}$	= -0.151,	$h_{9,27}$	= 0.206
α_{10} = 0.386,	$g_{10,7}$	= 0.688,	$g_{10,10}$	= -0.959,	$g_{10,11}$	= -0.246,	$g_{10,25}$	= 0.280,	$g_{10,27}$	= 0.589,	β_{10} = 0.408,	$h_{10,2}$	= 0.500,	$h_{10,4}$	= -0.386,	$h_{10,10}$	= 1.796,	$h_{10,25}$	= 0.972,	$h_{10,27}$	= 0.583
α_{11} = 0.261,	$g_{11,4}$	= 0.710,	$g_{11,7}$	= -0.553,	$g_{11,11}$	= -0.842,	$g_{11,19}$	= 0.206,	$g_{11,22}$	= 0.420,	β_{11} = 0.261,	$h_{11,4}$	= -0.759,	$h_{11,8}$	= 0.115,	$h_{11,11}$	= 1.888,	$h_{11,22}$	= -0.366,	$h_{11,27}$	= 0.426
α_{12} = 0.549,	$g_{12,4}$	= 0.146,	$g_{12,12}$	= -0.352,	$g_{12,16}$	= 0.160,	$g_{12,23}$	= 0.107,	$g_{12,27}$	= 0.209,	β_{12} = 0.525,	$h_{12,12}$	= 1.610,	$h_{12,14}$	= 0.161,	$h_{12,16}$	= 0.194,	$h_{12,18}$	= 0.063,	$h_{12,26}$	= -0.251
α_{13} = 0.406,	$g_{13,8}$	= 1.316,	$g_{13,11}$	= 0.120,	$g_{13,13}$	= -0.505,	$g_{13,19}$	= 0.241,	$g_{13,22}$	= 0.187,	β_{13} = 0.374,	$h_{13,13}$	= 2.639,	$h_{13,17}$	= -0.513,	$h_{13,18}$	= 0.695,	$h_{13,23}$	= 0.327,	$h_{13,28}$	= 0.290
α_{14} = 0.346,	$g_{14,3}$	= 0.473,	$g_{14,6}$	= 0.253,	$g_{14,9}$	= 2.205,	$g_{14,14}$	= -0.904,	$g_{14,28}$	= 0.400,	β_{14} = 0.379,	$h_{14,9}$	= -0.610,	$h_{14,14}$	= 1.914,	$h_{14,17}$	= 0.496,	$h_{14,18}$	= 0.239,	$h_{14,29}$	= 0.312
α_{15} = 0.350,	$g_{15,1}$	= 0.627,	$g_{15,15}$	= -0.621,	$g_{15,18}$	= -0.149,	$g_{15,20}$	= 0.314,	$g_{15,25}$	= 0.051,	β_{15} = 0.362,	$h_{15,4}$	= 0.379,	$h_{15,5}$	= 0.390,	$h_{15,10}$	= -0.125,	$h_{15,15}$	= 2.559,	$h_{15,29}$	= 0.543
α_{16} = 0.373,	$g_{16,1}$	= 0.235,	$g_{16,7}$	= 0.162,	$g_{16,11}$	= 0.730,	$g_{16,12}$	= -0.310,	$g_{16,16}$	= -0.555,	β_{16} = 0.373,	$h_{16,8}$	= -0.264,	$h_{16,11}$	= -0.506,	$h_{16,16}$	= 1.960,	$h_{16,20}$	= -0.137,	$h_{16,27}$	= 0.258
α_{17} = 0.303,	$g_{17,2}$	= 0.254,	$g_{17,6}$	= -0.534,	$g_{17,13}$	= 1.276,	$g_{17,17}$	= -0.957,	$g_{17,22}$	= 0.587,	β_{17} = 0.363,	$h_{17,9}$	= 0.399,	$h_{17,15}$	= 0.641,	$h_{17,17}$	= 2.715,	$h_{17,22}$	= 0.914,	$h_{17,24}$	= 0.594
α_{18} = 0.420,	$g_{18,18}$	= -1.266,	$g_{18,19}$	= 0.276,	$g_{18,22}$	= 0.707,	$g_{18,25}$	= -0.433,	$g_{18,27}$	= 0.449,	β_{18} = 0.445,	$h_{18,3}$	= 0.870,	$h_{18,16}$	= 0.563,	$h_{18,18}$	= 2.278,	$h_{18,24}$	= 0.569,	$h_{18,30}$	= -0.188
α_{19} = 0.361,	$g_{19,6}$	= 0.460,	$g_{19,14}$	= 0.745,	$g_{19,16}$	= 0.384,	$g_{19,19}$	= -1.082,	$g_{19,30}$	= 0.380,	β_{19} = 0.391,	$h_{19,2}$	= -0.571,	$h_{19,14}$	= 0.567,	$h_{19,15}$	= 0.477,	$h_{19,19}$	= 2.510,	$h_{19,29}$	= -0.215
α_{20} = 0.576,	$g_{20,15}$	= 0.991	$g_{20,17}$	= 0.179,	$g_{20,20}$	= -0.246,	$g_{20,26}$	= 0.389,	$g_{20,28}$	= -0.171,	β_{20} = 0.552,	$h_{20,2}$	= 0.127,	$h_{20,4}$	= -0.389,	$h_{20,7}$	= -0.229,	$h_{20,17}$	= -0.234,	$h_{20,20}$	= 1.358
α_{21} = 0.590,	$g_{21,8}$	= 0.300,	$g_{21,16}$	= 1.279,	$g_{21,20}$	= 0.231,	$g_{21,21}$	= -0.642,	$g_{21,22}$	= 0.262,	β_{21} = 0.630,	$h_{21,3}$	= 0.211,	$h_{21,10}$	= -0.114,	$h_{21,21}$	= 1.728,	$h_{21,23}$	= -0.472,	$h_{21,29}$	= 0.280
α_{22} = 0.586,	$g_{22,16}$	= 0.872,	$g_{22,20}$	= 0.314,	$g_{22,22}$	= -0.752,	$g_{22,25}$	= 0.103,	$g_{22,29}$	= 0.165,	β_{22} = 0.624,	$h_{22,11}$	= 0.480,	$h_{22,16}$	= -0.914,	$h_{22,20}$	= 0.915,	$h_{22,22}$	= 1.124,	$h_{22,23}$	= 0.393
α_{23} = 0.375,	$g_{23,7}$	= 0.201,	$g_{23,8}$	= -0.417,	$g_{23,17}$	= 0.409,	$g_{23,23}$	= -0.598,	$g_{23,24}$	= -0.325,	β_{23} = 0.424,	$h_{23,4}$	= 0.102,	$h_{23,18}$	= 0.282,	$h_{23,20}$	= 0.148,	$h_{23,23}$	= 1.963,	$h_{23,26}$	= 0.105
α_{24} = 0.336,	$g_{24,10}$	= -0.214,	$g_{24,18}$	= -0.379,	$g_{24,24}$	= -0.696,	$g_{24,27}$	= 0.489,	$g_{24,28}$	= 0.187,	β_{24} = 0.359,	$h_{24,15}$	= 1.279,	$h_{24,17}$	= 1.423,	$h_{24,19}$	= -0.490,	$h_{24,24}$	= 2.609,	$h_{24,29}$	= 0.435
α_{25} = 0.912,	$g_{25,7}$	= 0.289,	$g_{25,10}$	= -0.109,	$g_{25,20}$	= 0.407,	$g_{25,25}$	= -0.047,	$g_{25,30}$	= 0.190,	β_{25} = 0.938,	$h_{25,7}$	= 0.208,	$h_{25,12}$	= 0.065,	$h_{25,25}$	= 1.303,	$h_{25,28}$	= -0.131,	$h_{25,30}$	= 0.188
α_{26} = 0.490,	$g_{26,4}$	= 0.288,	$g_{26,11}$	= -0.084,	$g_{26,17}$	= 0.238,	$g_{26,21}$	= -0.430,	$g_{26,26}$	= -0.201,	β_{26} = 0.518,	$h_{26,1}$	= -0.176,	$h_{26,3}$	= 0.247,	$h_{26,19}$	= -0.214,	$h_{26,26}$	= 1.840,	$h_{26,27}$	= 0.232
α_{27} = 0.565,	$g_{27,3}$	= -0.158,	$g_{27,10}$	= 0.342,	$g_{27,24}$	= 0.698,	$g_{27,25}$	= 0.284,	$g_{27,27}$	= -0.452,	β_{27} = 0.545,	$h_{27,10}$	= 0.374,	$h_{27,11}$	= -0.182,	$h_{27,24}$	= -0.295,	$h_{27,27}$	= 1.089,	$h_{27,30}$	= 0.183
α_{28} = 0.652,	$g_{28,1}$	= -0.096,	$g_{28,10}$	= 0.076,	$g_{28,25}$	= 0.647,	$g_{28,26}$	= 0.093,	$g_{28,28}$	= -0.194,	β_{28} = 0.692,	$h_{28,2}$	= 0.210,	$h_{28,6}$	= 0.131,	$h_{28,17}$	= 0.052,	$h_{28,24}$	= 0.177,	$h_{28,28}$	= 1.443
α_{29} = 0.281,	$g_{29,9}$	= -0.337,	$g_{29,11}$	= 0.455,	$g_{29,13}$	= 0.285,	$g_{29,26}$	= 1.447,	$g_{29,29}$	= -1.312,	β_{29} = 0.260,	$h_{29,2}$	= 0.559,	$h_{29,13}$	= 0.931,	$h_{29,14}$	= -0.279,	$h_{29,18}$	= 0.796,	$h_{29,29}$	= 2.831
α_{30} = 0.656,	$g_{30,3}$	= 0.386,	$g_{30,5}$	= -0.252,	$g_{30,23}$	= -0.044,	$g_{30,27}$	= 0.884,	$g_{30,30}$	= -0.325,	β_{30} = 0.682,	$h_{30,14}$	= -0.066,	$h_{30,23}$	= 0.040,	$h_{30,24}$	= 0.125,	$h_{30,25}$	= -0.213,	$h_{30,30}$	= 1.515
other $g_{i,j}$ = 0.000, $h_{i,j}$ = 0.000																					

5.2.2 Results

When inferring the whole model structure, the proposed method inferred an average of 60.0 ± 1.9 true-positive interactions and 240.0 ± 1.9 false-positive interactions. The number of the false-negative interactions was 8.0 ± 1.9. Table 6 shows an example of the estimated S-system parameters. As shown in the table, many interactions were inferred. The number of the inferred interactions corresponded to the maximum indegree *I* mentioned in the section 2.3. In this experiment, many false-positive interactions with absolute parameter values too large to disregard were inferred. We suggest, however, that in cases with a few number of the false-negative interactions, the inference of false-positive interactions may not constitute a serious impediment, given that the inferred model is intended mainly for the use of biologists as a tool for generating hypotheses and facilitating experimental designs. The necessary interactions that were not correctly inferred should be added, and the wrong interactions should be removed in either of two ways, by using more sets of time-series data obtained from additional biological experiments, or by using further a priori knowledge about the genetic network. The computational time required for solving each decomposed subproblem averaged about 73.8 hours on a singe-CPU personal computer (Pentium III 1GHz), and the subproblems were solved simultaneously on parallel computers.

In order to show the effectiveness of the method for estimating the initial levels of the gene expression, we also performed an experiment without the estimation of them. We used the values obtained directly from the given time-series data as the initial expression levels, and then estimated only the S-system parameters. In the experiment without the estimation of the initial gene expression levels, the number of true-positive, false-positive and false-negative interactions averaged about 58.0 ± 0.7, 242.0 ± 0.7 and 10.0 ± 0.7, respectively. According to this result, the estimation of the initial gene expression levels slightly increased our chances of finding the correct interactions.

Table 7. Estimated S-system parameters in the experiment where noise-free data are given

α_1 = 0.975,	$g_{1,1}$ = -0.009,	$g_{1,8}$ = 0.001,	$g_{1,14}$ = -0.102,	$g_{1,24}$ = -0.003,	$g_{1,29}$ = -0.004,	β_1 = 0.975,	$h_{1,1}$ = 1.017,	$h_{1,8}$ = 0.002,	$h_{1,11}$ = -0.002,	$h_{1,22}$ = -0.006,	$h_{1,24}$ = -0.004									
α_2 = 0.836,	$g_{2,1}$ = 0.012,	$g_{2,2}$ = -0.077,	$g_{2,5}$ = -0.004,	$g_{2,17}$ = -0.002,	$g_{2,26}$ = -0.001,	β_2 = 0.835,	$h_{2,1}$ = 0.010,	$h_{2,2}$ = 1.115,	$h_{2,11}$ = 0.005,	$h_{2,16}$ = -0.004,	$h_{2,21}$ = 0.007									
α_3 = 0.912,	$g_{3,3}$ = -0.036,	$g_{3,7}$ = 0.002,	$g_{3,8}$ = 0.009,	$g_{3,11}$ = -0.006,	$g_{3,29}$ = 0.015,	β_3 = 0.911,	$h_{3,3}$ = 1.059,	$h_{3,8}$ = 0.006,	$h_{3,11}$ = -0.005,	$h_{3,25}$ = -0.003,	$h_{3,29}$ = 0.013									
α_4 = 0.936,	$g_{4,4}$ = -0.027,	$g_{4,8}$ = -0.009,	$g_{4,16}$ = 0.004,	$g_{4,18}$ = 0.010,	$g_{4,29}$ = -0.028,	β_4 = 0.934,	$h_{4,4}$ = 1.036,	$h_{4,8}$ = -0.012,	$h_{4,13}$ = 0.006,	$h_{4,16}$ = 0.005,	$h_{4,18}$ = 0.013									
α_5 = 0.945,	$g_{5,1}$ = 1.022,	$g_{5,5}$ = -0.023,	$g_{5,22}$ = -0.013,	$g_{5,24}$ = -0.002,	$g_{5,25}$ = -0.003,	β_5 = 0.943,	$h_{5,1}$ = -0.029,	$h_{5,5}$ = 1.030,	$h_{5,16}$ = 0.003,	$h_{5,18}$ = 0.006,	$h_{5,22}$ = -0.012									
α_6 = 0.972,	$g_{6,1}$ = 1.017,	$g_{6,6}$ = -0.012,	$g_{6,7}$ = -0.002,	$g_{6,17}$ = 0.001,	$g_{6,24}$ = -0.005,	β_6 = 0.972,	$h_{6,1}$ = -0.006,	$h_{6,5}$ = -0.001,	$h_{6,6}$ = 1.014,	$h_{6,24}$ = 0.008										
α_7 = 0.909,	$g_{7,2}$ = 0.516,	$g_{7,3}$ = 0.408,	$g_{7,7}$ = -0.032,	$g_{7,12}$ = 0.024,	$g_{7,29}$ = -0.005,	β_7 = 0.909,	$h_{7,2}$ = -0.034,	$h_{7,3}$ = -0.028,	$h_{7,7}$ = 1.066,	$h_{7,12}$ = 0.020,	$h_{7,29}$ = -0.004									
α_8 = 0.920,	$g_{8,4}$ = 0.208,	$g_{8,8}$ = -0.037,	$g_{8,17}$ = -0.213,	$g_{8,18}$ = 0.002,	$g_{8,29}$ = -0.004,	β_8 = 0.920,	$h_{8,4}$ = -0.010,	$h_{8,8}$ = 1.048,	$h_{8,18}$ = 0.003,	$h_{8,22}$ = -0.003										
α_9 = 0.879,	$g_{9,5}$ = 1.077,	$g_{9,6}$ = -0.120,	$g_{9,9}$ = -0.066,	$g_{9,16}$ = -0.035,	$g_{9,19}$ = -0.016,	β_9 = 0.878,	$h_{9,5}$ = -0.051,	$h_{9,6}$ = -0.009,	$h_{9,9}$ = 1.062,	$h_{9,16}$ = -0.045,	$h_{9,19}$ = -0.017									
α_{10} = 0.855,	$g_{10,7}$ = 0.315,	$g_{10,10}$ = -0.066,	$g_{10,17}$ = 0.001,	$g_{10,23}$ = 0.003,	$g_{10,25}$ = 0.002,	β_{10} = 0.855,	$h_{10,3}$ = 0.006,	$h_{10,7}$ = -0.033,	$h_{10,10}$ = 1.110,	$h_{10,18}$ = -0.003,	$h_{10,22}$ = -0.008									
α_{11} = 0.900,	$g_{11,4}$ = 0.403,	$g_{11,7}$ = -0.218,	$g_{11,11}$ = -0.038,	$g_{11,22}$ = 0.427,	$g_{11,29}$ = -0.004,	β_{11} = 0.899,	$h_{11,2}$ = 0.005,	$h_{11,4}$ = -0.035,	$h_{11,11}$ = 1.065,	$h_{11,22}$ = -0.015										
α_{12} = 0.914,	$g_{12,8}$ = 0.002,	$g_{12,11}$ = 0.008,	$g_{12,12}$ = -0.038,	$g_{12,19}$ = 0.005,	$g_{12,23}$ = 0.113,	β_{12} = 0.913,	$h_{12,10}$ = 0.005,	$h_{12,11}$ = 0.013,	$h_{12,12}$ = 1.062,	$h_{12,17}$ = 0.003,	$h_{12,29}$ = -0.009									
α_{13} = 0.885,	$g_{13,4}$ = -0.003,	$g_{13,8}$ = 0.625,	$g_{13,13}$ = -0.044,	$g_{13,29}$ = -0.022,		β_{13} = 0.885,	$h_{13,4}$ = 0.002,	$h_{13,8}$ = -0.044,	$h_{13,13}$ = 1.078,	$h_{13,26}$ = 0.006,	$h_{13,29}$ = -0.024									
α_{14} = 0.966,	$g_{14,3}$ = -0.002,	$g_{14,5}$ = 0.005,	$g_{14,9}$ = 1.017,	$g_{14,14}$ = -0.013,	$g_{14,24}$ = 0.001,	β_{14} = 0.966,	$h_{14,5}$ = 0.003,	$h_{14,9}$ = -0.015,	$h_{14,10}$ = 0.004,	$h_{14,14}$ = 1.019,	$h_{14,18}$ = 0.009									
α_{15} = 0.911,	$g_{15,7}$ = -0.002,	$g_{15,10}$ = 0.202,	$g_{15,15}$ = -0.032,	$g_{15,17}$ = 0.001,	$g_{15,26}$ = -0.006,	β_{15} = 0.909,	$h_{15,10}$ = -0.016,	$h_{15,15}$ = 1.060,	$h_{15,18}$ = 0.004,	$h_{15,26}$ = -0.005,	$h_{15,29}$ = -0.005									
α_{16} = 0.939,	$g_{16,9}$ = -0.004,	$g_{16,11}$ = 0.516,	$g_{16,12}$ = -0.201,	$g_{16,16}$ = -0.021,	$g_{16,22}$ = -0.015,	β_{16} = 0.939,	$h_{16,9}$ = -0.004,	$h_{16,11}$ = -0.017,	$h_{16,12}$ = 0.013,	$h_{16,16}$ = 1.047,	$h_{16,22}$ = -0.018									
α_{17} = 0.914,	$g_{17,12}$ = 0.014,	$g_{17,13}$ = 0.510,	$g_{17,17}$ = -0.047,	$g_{17,29}$ = -0.015,		β_{17} = 0.913,	$h_{17,1}$ = 0.002,	$h_{17,12}$ = 0.015,	$h_{17,13}$ = -0.031,	$h_{17,17}$ = 1.048,	$h_{17,29}$ = -0.014									
α_{18} = 0.866,	$g_{18,6}$ = 0.019,	$g_{18,8}$ = 0.010,	$g_{18,18}$ = -0.083,	$g_{18,24}$ = 0.001,	$g_{18,29}$ = 0.001,	β_{18} = 0.865,	$h_{18,6}$ = 0.016,	$h_{18,8}$ = 0.002,	$h_{18,13}$ = 0.003,	$h_{18,18}$ = 1.074,	$h_{18,26}$ = 0.002									
α_{19} = 0.930,	$g_{19,14}$ = 0.107,	$g_{19,19}$ = -0.028,	$g_{19,21}$ = -0.016,	$g_{19,29}$ = 0.003,	$g_{19,30}$ = -0.002,	β_{19} = 0.929,	$h_{19,9}$ = -0.001,	$h_{19,10}$ = -0.002,	$h_{19,19}$ = 1.043,	$h_{19,25}$ = 0.001										
α_{20} = 0.941,	$g_{20,15}$ = 0.717	$g_{20,20}$ = -0.021,	$g_{20,26}$ = 0.307,	$g_{20,28}$ = 0.003,	$g_{20,29}$ = -0.005,	β_{20} = 0.940,	$h_{20,15}$ = -0.021,	$h_{20,17}$ = -0.003,	$h_{20,20}$ = 1.040,	$h_{20,26}$ = -0.007,	$h_{20,29}$ = -0.008									
α_{21} = 0.870,	$g_{21,6}$ = 0.009,	$g_{21,10}$ = -0.003,	$g_{21,16}$ = 0.645,	$g_{21,21}$ = -0.061,	$g_{21,22}$ = -0.004,	β_{21} = 0.869,	$h_{21,6}$ = 0.009,	$h_{21,16}$ = -0.045,	$h_{21,17}$ = 0.006,	$h_{21,21}$ = 1.091,	$h_{21,29}$ = -0.003									
α_{22} = 0.731,	$g_{22,3}$ = 0.003,	$g_{22,11}$ = -0.011,	$g_{22,16}$ = 0.607,	$g_{22,22}$ = -0.168,	$g_{22,29}$ = 0.007,	β_{22} = 0.753,	$h_{22,11}$ = -0.013,	$h_{22,16}$ = -0.071,	$h_{22,22}$ = 1.188,	$h_{22,25}$ = 0.001,	$h_{22,29}$ = 0.010									
α_{23} = 1.006,	$g_{23,8}$ = -0.004,	$g_{23,17}$ = 0.198,	$g_{23,23}$ = 0.002,	$g_{23,25}$ = -0.002,		β_{23} = 1.005,	$h_{23,6}$ = -0.002,	$h_{23,8}$ = -0.004,	$h_{23,22}$ = 0.002,	$h_{23,23}$ = 0.995,	$h_{23,25}$ = -0.002									
α_{24} = 0.977,	$g_{24,13}$ = 0.001,	$g_{24,15}$ = -0.202,	$g_{24,18}$ = -0.100,	$g_{24,19}$ = 0.300,	$g_{24,24}$ = -0.008,	β_{24} = 0.977,	$h_{24,15}$ = 0.002,	$h_{24,18}$ = -0.001,	$h_{24,19}$ = -0.003,	$h_{24,24}$ = 1.010,	$h_{24,30}$ = -0.003									
α_{25} = 0.882,	$g_{25,1}$ = 0.006,	$g_{25,20}$ = 0.428,	$g_{25,25}$ = -0.055,	$g_{25,28}$ = 0.004,	$g_{25,29}$ = 0.004,	β_{25} = 0.882,	$h_{25,1}$ = 0.004,	$h_{25,13}$ = -0.003,	$h_{25,18}$ = -0.001,	$h_{25,20}$ = -0.021,	$h_{25,25}$ = 1.076									
α_{26} = 0.861,	$g_{26,17}$ = 0.003,	$g_{26,18}$ = -0.009,	$g_{26,21}$ = -0.214,	$g_{26,26}$ = -0.060,	$g_{26,28}$ = 0.119,	β_{26} = 0.860,	$h_{26,11}$ = 0.009,	$h_{26,12}$ = -0.005,	$h_{26,18}$ = -0.014,	$h_{26,21}$ = 0.019,	$h_{26,26}$ = 1.099									
α_{27} = 0.877,	$g_{27,3}$ = 0.002,	$g_{27,24}$ = 0.649,	$g_{27,25}$ = 0.327,	$g_{27,27}$ = -0.066,	$g_{27,30}$ = -0.215,	β_{27} = 0.878,	$h_{27,24}$ = -0.028,	$h_{27,25}$ = -0.014,	$h_{27,27}$ = 1.070,	$h_{27,29}$ = -0.005,	$h_{27,30}$ = 0.015									
α_{28} = 0.942,	$g_{28,8}$ = 0.005,	$g_{28,19}$ = 0.003,	$g_{28,25}$ = 0.514,	$g_{28,28}$ = -0.021,	$g_{28,29}$ = 0.016,	β_{28} = 0.942,	$h_{28,18}$ = 0.005,	$h_{28,22}$ = -0.007,	$h_{28,25}$ = -0.013,	$h_{28,28}$ = 1.040,	$h_{28,29}$ = 0.016									
α_{29} = 0.639,	$g_{29,1}$ = 0.001,	$g_{29,12}$ = 0.012,	$g_{29,24}$ = 0.006,	$g_{29,26}$ = 0.509,	$g_{29,29}$ = -0.298,	β_{29} = 0.639,	$h_{29,4}$ = -0.004,	$h_{29,12}$ = 0.011,	$h_{29,24}$ = 0.003,	$h_{29,26}$ = -0.107,	$h_{29,29}$ = 1.254									
α_{30} = 0.926,	$g_{30,5}$ = -0.001,	$g_{30,18}$ = -0.002,	$g_{30,24}$ = 0.003,	$g_{30,27}$ = 0.625,	$g_{30,30}$ = -0.032,	β_{30} = 0.925,	$h_{30,13}$ = 0.002,	$h_{30,22}$ = -0.001,	$h_{30,27}$ = -0.017,	$h_{30,28}$ = -0.002,	$h_{30,30}$ = 1.047									
other $g_{i,j}$ = 0.000, $h_{i,j}$ = 0.000																				

When the differential equation (4) was solved, the gene expression curves $\hat{X}_j$ were directly estimated in this study. This estimation of the gene expression curves is necessary for the problem decomposition strategy mentioned in the section 2.2. However, these estimated curves will cause erroneous results if they fail to resemble the true ones. To increase the probability of finding the correct interactions, therefore, all subproblems should be executed simultaneously, and the gene expression curves should be updated when better ones are obtained as solutions of the differential equations (4). Our group has been working to extend our method for these purposes.

5.3 Experiment 3: large-scale network inference with noise-free data

In the experiment of the previous subsection, the proposed method failed to infer the correct network structure because of noise in the given data. However, the proposed method has an ability to infer the target model B correctly when noise-free data are available.

Table 7 shows the best results in the experiment where 15 sets of the noise-free time-series data, began from randomly generated initial values in [0.0, 2.0], were given. In this experiment, the estimation of the initial gene expression levels was omitted, since the noise-free data were given. Spline interpolation was used to obtain $\hat{X}_j$'s, and the rest of the experimental conditions were the same as those in the subsection 5.2. As shown in the table, our method was failed to estimate precise parameter values. The numbers of the true-positive, false-positive and false-negative interactions were 68.0, 225.0 and 0.0, respectively. While all of the interactions existing in the target model B were inferred correctly, many false-positive interactions were inferred. However, most of the false-positive interactions were omissible, as the absolute parameter values of these interactions were much smaller than those of the correct interactions. When we omitted these interactions, we obtained the almost correct network structure.

Even when the noise-free data were given as the observed gene expression patterns and the

estimation of the initial gene expression levels were omitted, the proposed method was unable to estimate the S-system parameter values with perfect precision. One of the factors compromising the precision of our method may have been the implicit noise. In order to calculate the fitness value of the *i*-th subproblem, the gene expression pattern of the *i*-th gene is acquired by solving the equation (4). To solve this equation, we estimate the gene expression curves of the other genes $\hat{X}_j$ directly from the observed time-series data, as mentioned in the section 2.2. These directly estimated curves may contain implicit noise, as it is difficult to estimate accurate curves from a finite number of sampling points even when the sampling points are entirely free of noise. Note that the implicit noise is unavoidable for the problem decomposition approach even when the observed data contain no measurement error.

6. Conclusion

In this paper, we extended the problem decomposition approach to a realistic noisy environment. The proposed method employs a technique to decompose the genetic network inference problem into several subproblems, and then solves each subproblem using GLSDC. In addition, our method estimates the initial levels of the gene expression in order to enhance the probability of finding the correct interaction between genes. Our experiments demonstrated that this method slightly increased our chances of inferring the correct interactions when the realistic noisy time-series data were given.

Even if the problem decomposition strategy is applied, our method is still incapable of inferring real genetic networks containing many hundreds or thousands of genes. When attempting to infer these larger-scale genetic networks, we should combine the use of our method based on the S-system model with other methods based on other models, as proposed in the paper by Maki and colleagues [3].

References

[1] J.L. DeRisi, V.R. Iyer and P.O. Brown, *Science*, **278**, 680-686 (1997).
[2] E. Sakamoto and H. Iba, Proc. of 2001 Congress on Evolutionary Computation (CEC), 720-726 (2001).
[3] Y. Maki, D. Tominaga, M. Okamoto, S. Watanabe and Y. Eguchi, Proc. of Pacific Symposium on Biocomputing (PSB), **6**, 446-458 (2001).
[4] P. D'haeseleer, S. Liang and R. Somogyi, *Bioinformatics*, **16**, 707-726 (2000).
[5] T. Akutsu, S. Miyano and S. Kuhara, *Bioinformatics*, **16**, 727-734 (2000).
[6] T. Chen, H.L. He and G.M. Church, Proc. of PSB, **4**, 29-40 (1999).
[7] P. D'haeseleer, X. Wen, S. Fuhrman and R. Somogyi, Proc. of PSB, **4**, 41-52 (1999).
[8] D. Tominaga, N. Koga and M. Okamoto, Proc. of Genetic and Evolutionary Computation Conference 2000, 251-258 (2000).
[9] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi and M. Tomita, *Bioinformatics*, **19**, 643-650 (2003).
[10] T. Ueda, I. Ono and M. Okamoto, *Genome Informatics*, **13**, 386-387 (2002).
[11] Y. Maki, T. Ueda, M. Okamoto, N. Uematsu, Y. Inamura and Y. Eguchi, *Genome Informatics*, **13**, 382-383 (2002).
[12] S. Kimura, M. Hatakeyama and A. Konagaya, Proc. of 2003 CEC, 631-638 (2003).

[13] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, Numerical Recipes in C second edition, Cambridge University Press (1995).
[14] W.S. Cleveland, *J. of American Statistical Association*, **79**, 829-836 (1979).
[15] D. Thieffry, A.M. Huerta, E. Perez-Rueda and J. Collado-Vides, *BioEssays*, **20**, 433-440 (1998).
[16] S. Kimura and A. Konagaya, Proc. of 2003 IEEE Conf. on Systems, Man & Cybernetics, 335-342 (2003).
[17] S. Kimura, I. Ono, H. Kita and S. Kobayashi, *Trans. of the Society of Instrument and Control Engineers*, **36**, 1162-1171 (2000, in Japanese).
[18] H. Satoh, M. Yamamura and S. Kobayashi, Proc. of the 4th Int. Conf. on Fuzzy Logic, Neural Networks and Soft Computing, 494-497 (1996).
[19] S. Ando and H. Iba, Proc. of 2001 CEC, 712-719 (2001).