Proceedings of the 2007 American Control Conference
Marriott Marquis Hotel at Times Square
New York City, USA, July 11-13, 2007

A Multi-Model Approach to Identification of Biosynthetic Pathways

Mary J. Dunlop, Elisa Franco, and Richard M. Murray

Abstract— **We present an identification framework for biochemical systems that allows multiple candidate models to be compared. This framework is designed to select a model that fits the data while maintaining model simplicity. The model identification task is divided into a parameter estimation stage and a model comparison stage. Model selection is based on calculating Akaike's Information Criterion, which is a systematic method for determining the model that best represents a set of experimental data. Two case studies are presented: a simulated transcriptional control circuit and a system of oscillators that has been built and characterized *in vitro*. In both examples the multi-model framework is able to discriminate between model candidates to select the one that best describes the data.**

I. Introduction and Background

Biochemical pathways are complex systems, often affected by high levels of uncertainty in both reaction rates and pathway connectivity. Reaction rates are model parameters, which may be unknown or difficult to measure. Pathway connectivity determines model structure and describes how reaction components interact with each other. Model structure may be uncertain if chemical species react in unintended ways. Obtaining a reliable mathematical representation is important when trying to characterize or re-design a biological network, as is often the case in synthetic biology applications.

In this paper we present a general architecture for model selection in biological systems. The framework uses experimental data to estimate the parameters of each model, then Akaike's Information Criterion (AIC) is used to select the best model from a set of candidates by processing the residual error between estimated and experimental data.

In the recent literature, several studies have focused on identification and estimation architectures for finding parameters of biochemical pathways. The suitability of different optimization algorithms for biochemical kinetic parameters estimation is discussed in [22], [10], [23]. Identifiability is closely linked to parameter estimation and is of particular importance for complex nonlinear systems. A global algorithm based on differential algebra is proposed in [3], while local identifiability is discussed in [16]; a local sufficient condition is also given in [9]. An iterative scheme to optimize identifiability of biochemical networks is offered in [11], which is an extension of a method introduced in [27].

The AIC has been used successfully for biological model identification in previous studies such as [26] and [12] where the authors focus on identification and reconstruction of gene networks using linear models; [4] uses the AIC as a statistical measure when determining network connectivity from analysis of proteomic and genomic data, but does not develop a mathematical model for the system.

The main contribution of this paper is a method for choosing the best model from a set of candidates that describe a set of experimental data. We discuss parameter estimation in a biological context and summarize the main results associated with Akaike's Information Criterion.

A. Parameter Estimation

Parameter estimation is an important subclass of system identification problems [21]. The goal is to identify a set of parameters $\theta \in \mathbb{R}^P$ that cause the model

$$y = F(\theta)$$

to best fit the measured data y. Biological systems are typically modeled with sets of ordinary differential equations, so the parameter estimation problem becomes finding θ given

$$\begin{aligned} \dot{x} &= f(x, \theta) \\ y &= h(x), \end{aligned}$$

where x is the system state, usually concentrations of DNA, mRNA, and proteins.

Parameter estimation problems are generally approached by posing them as optimization problems, where the goal is to minimize the difference between estimated and experimental data with respect to the model parameters, θ. Biological parameter estimation problems are generally not suitable for gradient-based approaches because they are non-convex and solutions can easily fall into local-minima. In this context, global search algorithms have often proved more successful [23].

Simulated annealing is a well-studied global optimization method for nonlinear cost functions. First introduced in [20] and [8], the method is an analogy to the thermodynamic process of annealing. The cost function's value in parameter space is treated as an energy landscape, where the goal is to find the lowest energy state. At each step of the algorithm a move is made in the parameter space. If this results in a decrease in energy ($\Delta E \leq 0$), the new parameters are accepted. If energy increases, a probabilistic rule is applied to decide whether the new parameters are accepted or not. The probability that an energy-increasing move is accepted, $P(\Delta E) = \exp(-\Delta E/k_B T)$, is a function of a T, a temperature-like variable. T starts high to allow wide exploration of the energy landscape, and decreases as the

Research supported in part by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office.

The authors are in the Division of Engineering and Applied Sciences, California Institute of Technology, Pasadena, CA 91125. {mjdunlop, elisa, murray}@cds.caltech.edu

optimization proceeds to allow for fine tuning. Allowance of occasional energy-increasing moves is what keeps the optimization from getting stuck in local minima.

Simulated annealing has been used successfully in [6] and [22] to estimate the parameters of ODE based models of biochemical systems.

B. The Akaike Information Criterion

AIC is a method for model selection that trades off fitting the data well and maintaining model simplicity. Given a set of experimental data y, and a set of candidate models $\mathcal{M} = \{M_1, ..., M_m\}$ to fit the data, each characterized by a parameter vector θ_i of size P_i, AIC [1] is a systematic procedure for solving the model selection problem. This method bridges the gap between information and likelihood theory by offering an estimate of the Kullback-Leibler (KL) distance [19] based on a maximized log likelihood value. Suppose the true process (which has no parameters) is described by $y = G(z)$, where z is a random variable and y is our measurement set. If $G(z)$ is approximated by model candidate $M(z|\theta)$, which depends on a parameter vector θ, then the loss of information introduced by the approximation can be measured with the KL distance

$$I(G, M) = \int G(z) \log \left(\frac{G(z)}{M(z|\theta)} \right) dz. \tag{1}$$

This distance does not satisfy all the properties of a metric, but represents a well-defined concept of "distance" between the model and reality. If the model parameters need to be estimated and the true process is unknown, (1) cannot be directly computed. Rewriting it as

$$\begin{aligned} I(G, M) = &\int G(z) \log(G(z)) dz \\ &- \int G(z) \log(M(z|\theta) dz, \end{aligned} \tag{2}$$

the first term is constant across models, so an estimate of (1) is based only on the second term, which is a relative KL distance. Denoting the parameter estimates as $\hat{\theta}$, given model M and a set of data y, Akaike [1] found that this relative KL distance is approximated by a biased function of the log likelihood of the estimation process:

$$AIC(G, M) = -2 \log(\mathcal{L}(\hat{\theta}|y)) + 2P, \tag{3}$$

where $\mathcal{L}$ denotes the log likelihood function (the factor 2 was introduced for historical reasons). The term $2P$ is the estimated bias.

If $M(z|\theta)$ is a candidate $M_i(z|\theta_i) \in \mathcal{M}$, the model selection process can be stated as

$$\min_{M_i \in \mathcal{M}} AIC(G, M_i) = -2 \log(\mathcal{L}(\hat{\theta}_i|y)) + 2P_i. \tag{4}$$

The above criterion can be enhanced depending on the size N of the data set y and of the number of parameters. In particular, if $N / \max_i P_i > 40$, a second-order bias correction can be used [13]. For large sample sizes, an improvement to AIC was obtained with Takeuchi's TIC [25] that eliminates the bias-adjustment term.

Based on AIC weighting, there are several ways to proceed to multi-model inference [7], including ranking and scaled relative plausibility of the available models, model-averaged parameter estimates, and estimates of sampling variances not conditioned on any particular model.

It is very important to notice that the AIC value associated with each model does not have an absolute meaning: it is rather its size relative to the minimum AIC in the set of candidates that allows model ranking. For each model, the quantity $\Delta_i = AIC(G, M_i) - \min_i AIC(G, M_i)$ will be examined.

The relative likelihood $\mathcal{L}(\hat{\theta}_i|y)$ of a model M_i, given the data, is proportional to $\exp(-\Delta_i/2)$; normalization over all sets of models yields the so-called Akaike weights:

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{j=1}^{m} \exp(-\Delta_j/2)}. \tag{5}$$

Each w_i weights the evidence in favor of model i being the actual KL best model for the situation.

II. Methods

We integrate parameter estimation and Akaike's model selection criterion. For a given set of data, parameter estimation is performed for each candidate model M_i. Next, the AIC is calculated for all candidate models M_1, ..., M_m. The model with the lowest AIC value is the one that best describes the experimental data, in the KL sense, according to (4). Fig. 1a shows the multi-model identification method.

Parameter estimation is performed by using Adaptive Simulated Annealing [14] to solve the optimization problem

$$\begin{aligned} \max_{\tilde{\theta}} \ & \sum_i \|y_i^{est}(\tilde{\theta}) - y_i^{exp}\|_2 \\ \text{subject to} \quad & \theta_L \leq \tilde{\theta} \leq \theta_U. \end{aligned} \tag{6}$$

$y_i^{exp} \in \mathbb{R}^N$ is a vector of experimental data at N time points associated with output i. The optimization variables for the simulated annealing algorithm are the estimated parameters $\tilde{\theta}$, which are constrained to physically realistic values by lower bounds $\theta_L \in \mathbb{R}^P$ and upper bounds $\theta_U \in \mathbb{R}^P$. The estimated data y_i^{est} are found by numerically integrating the model ODEs with the estimated parameters. The algorithm is described in Fig. 1b.

The calculation of AIC for a certain model M_i is in general straightforward [21], [7]. If the estimated parameter vector resulting from the optimization problem (6) is $\hat{\theta}_i$, and the associated error $\epsilon_i = y_i^{est}(\hat{\theta}) - y_i^{exp}$ is assumed to be Gaussian with constant variance, AIC is given by:

$$\begin{aligned} AIC_i = &\frac{1}{2} \log \left[\sum_{j=1}^{N} \epsilon_i(j)^T \epsilon_i(j) \right] + \frac{1}{2} \\ &+ \frac{1}{2} \log 2\pi + \frac{P_i}{N}. \end{aligned}$$

Different expressions for AIC can be found when the error is assumed to have other probability distributions. This is an important consideration for biological noise sources, which

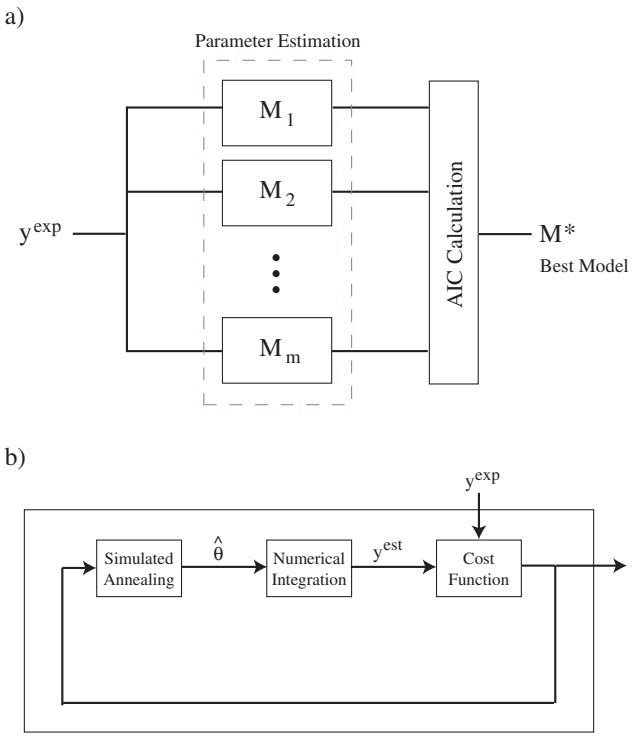

Fig. 1. Overview of the Multi-Model Approach. a) Experimental data is compared with the results of the parameter estimation for each of the m models. The error between the estimated and experimental data is used to calculate the AIC for each model. The model with the smallest AIC value is the best of the candidate models. b) Parameter estimation is performed by minimizing the error between estimated and experimental data.

may not be accurately represented by Gaussian distributions. For example, it was found in [24] that cellular processes are best modeled with log-normal noise. The AIC method can be used not only for model structure selection, but also for determining which stochastic properties best describe the system.

III. Results

The multi-model identification method is applied to two example problems, both involving discrimination between different types of network connectivity. The first example is a simulated transcriptional regulatory network, the second example uses data from a set of oscillators that have been constructed *in vitro*.

A. Three Types of Transcriptional Control

Three types of transcriptional control are possible in genetic regulation: activation, repression, and no regulation. If A is a transcription factor, a protein that can regulate expression of B, then the following equations can be used to describe these three cases:

$$\dot{A} = \alpha_0 - \beta A$$

$$\dot{B} = \begin{cases} \alpha_0 - \beta B & \text{no regulation} \\ \alpha_0 + \frac{\alpha A^n}{1+A^n} - \beta B & \text{activation} \\ \alpha_0 + \frac{\alpha}{1+A^n} - \beta B & \text{repression} \end{cases}$$

α_0 is the basal transcription rate, α is the transcription rate that is regulated by the transcription factor A, n is the Hill coefficient, and β is the protein degradation rate [2]. A and B are protein concentrations. The model states are $x_1 = A$, $x_2 = B$ and we assume that both states are measurable. The parameters are $\theta = [\alpha, \beta, n]$ and we set $\alpha_0 = 0.001$.

The "experimental" data for this example are simulated numerically for the three types of network connections. Gaussian white noise with covariance $V = \text{diag}([0.05^2\ 0.05^2])$ is added to both states to simulate measurement noise. For each set of experimental data we evaluate the three models M_1 = no regulation, M_2 = activator, and M_3 = repressor with the multi-model identification algorithm. The number of parameters associated with each model is $P_1 = 2$, $P_2 = 3$, and $P_3 = 3$.

Fig. 2 shows an example of how the parameter estimates change. Although the estimates tend towards the actual parameter values, there are large variations as the parameter space is explored. These data are from the beginning of the parameter search process.

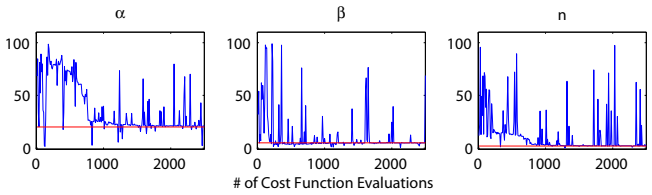

Fig. 2. Example of parameter estimation. The blue line is the parameter estimate at each cost function evaluation. The red line is the actual parameter value used to simulate the experimental data. The three parameters α, β, and n are estimated. Only the first 7% of the cost function evaluations are shown. $\theta_L = [0, 0, 0]$, $\theta_U = [100, 100, 100]$.

Fig. 3 compares experimental and estimated data after parameter estimation is complete. These data are used to calculate the Δ AIC values, which are listed in Table I. The associated weights are given in Table II.

	None	**Act**	**Rep**
M_1 **= None**	0.0	0.88	3.91
M_2 **= Act**	0.0	0.0	2.39
M_3 **= Rep**	0.0	0.85	0.0

TABLE I

Δ AIC VALUES FOR 3 CIRCUIT NETWORK DISCRIMINATION

	None	**Act**	**Rep**
M_1 **= None**	0.334	0.281	0.099
M_2 **= Act**	0.333	0.435	0.210
M_3 **= Rep**	0.333	0.284	0.692

TABLE II

WEIGHT VALUES (w_i) FOR 3 CIRCUIT NETWORK DISCRIMINATION

For the activation and repression experimental data, the respective models are identified as being the most likely candidates based on their weight values w_i. The no regulation

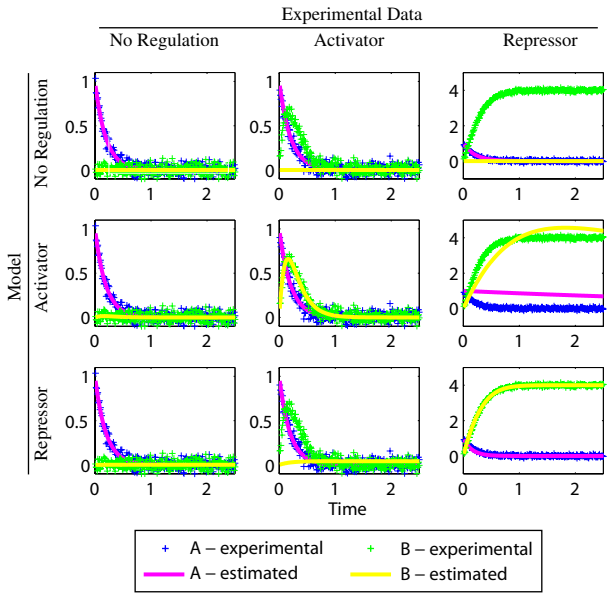

Fig. 3. Estimated and experimental data for 3-circuit network discrimination. Plots shows protein concentration versus time for the two states, A and B, versus time. Each column uses different experimental data. Each row tests a different model.

data can be fit well by all the models. This is because the model M_1 is a subset of models M_2 and M_3 if $\alpha = 0$. Indeed, in the parameter estimation stage the value of α is estimated to be very small.

B. In vitro Oscillators

In vitro circuits are a subclass of synthetic circuits that arose from the need to better understand the regulatory capabilities that nucleic acids have within a cell [17], [18], [15]. Biological parts, such as DNA, RNA, and enzymes are combined in a biochemical reaction that is similar to what happens within a cell, but without the added complexity of interactions with other cellular components. *In vitro* versions of transcriptional circuits can be built by designing short (30-100 base pair) DNA template strands that can interact with their RNA products and with other short DNA molecules.

In this example we consider two versions of a transcriptional oscillator developed by Kim [17]: a basic two node oscillator and a two node oscillator with positive feedback, shown in Fig. 4a and b. The genetic components of the basic two node oscillator are graphically described in Fig. 4c: two DNA templates, denoted x_1 and x_8, are partially incomplete in the promoter region. In order for RNAP to bind and initiate transcription of the two mRNA strands, x_4 and x_{10}, the single stranded DNA activators x_2 and x_9 need to bind and complete the templates. Oscillations arise because x_2 also binds to its complementary molecule x_3. When x_{10} is in excess, x_3 and x_{10} form a hybrid double stranded complex, freeing x_2, which binds to x_1 and allows for transcription of x_4. On the other hand, when x_4 is in excess, the activator x_9 is stripped off of x_8, decreasing the amount of x_{10} in solution. The hybridization reactions (x_3 and x_{10}, x_4 and x_9) are favored in competitive binding. RNaseH is an enzyme that degrades the hybrid complexes, breaking down the RNA and releasing the single stranded DNA.

The oscillations are measured with fluorescent molecules that are integrated within the strands of interest. If a quencher is not in their proximity, the fluorophores emit light in a know emission/absorption spectrum. Fluorescence measurements corresponding to the concentration of the incomplete DNA templates x_1 and x_8 are measured ($y = [x_1, x_8]$).

The two node oscillator can be modeled with the set of ODEs and algebraic equations (7)–(9). The state variables are concentrations of the DNA and RNA molecules. Mass action (hybridization) and Michaelis-Menten (RNAP, RNaseH activity) reactions are present; the kinetic rates are denoted p_i, and are the parameters to be estimated.

$$
\begin{aligned}
\frac{dx_1}{dt} &= -p_7 x_1 x_2 + p_{11} x_5 x_3 \\
\frac{dx_2}{dt} &= -p_7 x_2 (x_1 + x_{16}) - p_{10} x_2 x_3 + p_9 x_6 x_4 \\
\frac{dx_3}{dt} &= \frac{p_6}{p_5} H x_7 - p_{10} x_2 x_3 - p_8 x_4 x_3 - p_{11} x_3 (x_5 + x_{17}) \\
\frac{dx_4}{dt} &= \frac{p_{13}}{p_{12}} R(x_{11} + x_{15}) + \frac{p_{15}}{p_{14}} R x_8 + \frac{p_{29}}{p_{28}} R x_{17} \\
&\quad + \frac{p_{31}}{p_{30}} R x_{16} - p_8 x_4 x_3 - p_9 x_6 x_4 \\
\frac{dx_8}{dt} &= -p_{18} x_8 (x_9 + x_{14}) + p_{20} x_{11} x_{10} \\
\frac{dx_9}{dt} &= -p_{19} x_9 (x_{10} + x_{13}) - p_{18} x_8 x_9 + p_{23} x_{14} \\
\frac{dx_{10}}{dt} &= \frac{p_2}{p_1} R x_5 + \frac{p_4}{p_3} R x_1 - p_{19} x_{10} (x_9 + x_{14}) \\
&\quad - p_{20} x_{11} x_{10} \\
\frac{dx_{13}}{dt} &= -p_{32} x_{13} (x_9 + x_{11}) + p_{23} (x_{14} + x_{15}) \\
&\quad + p_{19} x_{14} x_{10} \\
\frac{dx_{14}}{dt} &= \frac{p_{17}}{p_{16}} H\ x_{12} + p_{32} x_9 x_{13} - p_{23} x_{14} - p_{19} x_{14} x_{10} \\
&\quad - p_{18} x_8 x_{14} \\
\frac{dx_{15}}{dt} &= p_{32} x_{11} x_{13} - p_{23} x_{15} + p_{18} x_8 x_{14} \\
\frac{dx_{16}}{dt} &= -p_{21} x_{16} x_2 + p_{22} x_{17} x_3
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
R &= RNAP_{tot} / (1 + \frac{x_5}{p_1} + \frac{x_1}{p_3} + \frac{(x_{11} + x_{15})}{p_{12}} \\
&\quad + \frac{x_8}{p_{14}} + \frac{x_{17}}{p_{28}} + \frac{x_{16}}{p_{30}}) \\
H &= RNaseH_{tot} / (1 + \frac{x_7}{p_5} + \frac{x_{12}}{p_{16}})
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
0 &= x_{16} + x_{17} \\
0 &= x_1 + x_5 \\
0 &= x_2 + x_5 + x_6 + x_{17} \\
0 &= x_3 + x_6 + x_7 \\
0 &= x_8 + x_{11} + x_{15} \\
0 &= x_9 + x_{11} + x_{12} + x_{14} + x_{15}
\end{aligned} \tag{9}
$$

The reactions are started by adding x_1, x_2, x_3, x_8, and x_9 for the basic oscillator. For the self activating oscillator x_{16} is also added initially. These equations describe both basic and self activating oscillators. The basic oscillator model is a subset of the self activating oscillator model where the reaction terms associated with parameters p_{21}, p_{22}, p_{28}, p_{29}, p_{30}, and p_{31} are eliminated. The basic oscillator has 22 model parameters and the self activating version has 28.

Because of the large number of model parameters and small number of measured outputs, it is important to consider different perturbations to the experiment when determining system parameters. Single data sets will not be sufficient to identify all the parameters in the model. Model parameters are fit to 45 experimental data sets in [17]: 38 data sets for the basic oscillator and 7 for the self activating oscillator. Both experimental data sets are used to fit the 22 parameters common to both switch models, and the self activating switch data is used to fit the final 6 parameters specific to the self activating model. We use these estimated parameters when calculating the AIC.

If the simulated annealing parameter estimation method is applied to individual experimental data sets, the optimization has many parameters to adjust and only two outputs to fit. Consequently, the estimated data match experimental data well in all cases. Since several perturbation experiments exist, it is more realistic to estimate the parameters using all the available data at once, as in [17], since the parameters governing the process remain the same. Determining the experimental perturbations necessary to accurately predict parameters, and not just outputs, is closely linked to identifiabilty of parameters and is discussed further in the Future Work section.

The two models considered are $M_1 =$ basic oscillator and $M_2 =$ self activating oscillator. We use two data sets for each of the oscillators and compute the AIC and corresponding weights for each of the candidate models.

Fig. 5 compares the estimated and experimental data for the two models (rows) and four experimental data sets (columns). Tables III and IV report the ΔAIC and weight values for the four sets of experimental data. The multi-model method successfully determines which type of oscillator generated the data. The weight values suggest that the differences between the two models are not as clear as in the previous example, but these data are from a significantly more complex biological system.

IV. Conclusion and Future Work

A method for model selection that utilizes Akaike's Information Criterion has been presented in this paper. We apply the framework to two problems of identification and parameter estimation in biochemical networks. These network models are typically complex and nonlinear making identification a challenging problem. We present a method that works in two steps: parameter estimation and model selection with AIC. Given several candidate models that may describe experimental data, the multi-model identification framework uses a systematic method based on information and likelihood theory to determine which model best describes the data. Two application examples were considered: a transcriptional control circuit and a pair of *in vitro* oscillators.

In the future it will be important to integrate identifiability tests into the multi-model framework. Although it is often possible to fit model outputs to experimental data, for accurate parameter estimation it is important that the system be sufficiently perturbed. Identifiability tests using the Fisher Information Matrix have proved successful for biological estimation problems [11] and are good candidates for integration into the multi-model framework.

A particularly useful aspect of AIC is that it allows not only for model structure selection, but also for model

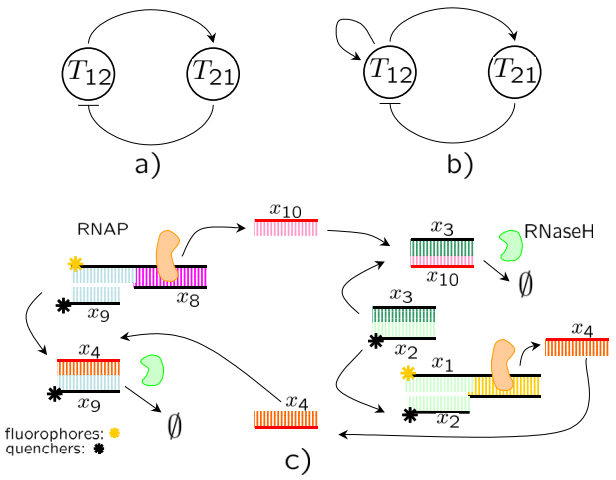

Fig. 4. a) Basic two node oscillator: T_{21} corresponds to the template x_1, while T_{12} corresponds to template x_8. b) Self activating two node oscillator. c) Graphical sketch of the basic two node oscillator mechanism.

	Basic 1	**Basic 2**	**SA 1**	**SA 2**
M_1 **= Basic**	0.0	0.0	0.18	0.42
M_2 **= SA**	0.23	0.40	0.0	0.0

TABLE III

Δ AIC VALUES FOR OSCILLATOR NETWORKS (SA = SELF ACTIVATING)

	Basic 1	**Basic 2**	**SA 1**	**SA 2**
M_1 **= Basic**	0.529	0.550	0.477	0.447
M_2 **= SA**	0.471	0.450	0.523	0.553

TABLE IV

WEIGHT VALUES (w_i) FOR OSCILLATOR NETWORKS (SA = SELF ACTIVATING)

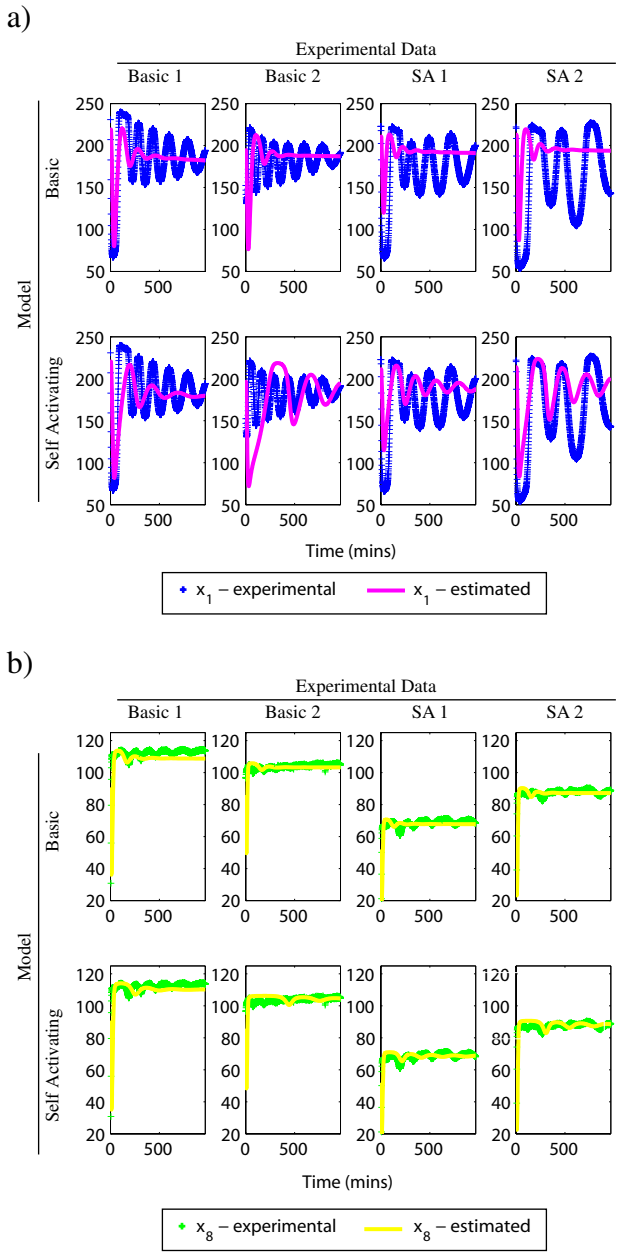

Fig. 5. Estimated and experimental data for the *in vitro* oscillators: a) state x_1, b) state x_8. Both figures show protein concentration (nM) versus time (minutes). Different experimental data sets are shown in each column; rows show different model candidates.

statistical selection. Characterizing the stochastic properties of biological noise is an important area of study where the AIC may be able to extend current understanding.

REFERENCES

[1] H. Akaike. "Information theory and an extension of the maximum likelihood principle." International Symposium on Information Theory, 2nd. pg. 267-281. 1973.

[2] U. Alon. An introduction to systems biology: design principles of biological circuits. Chapman and Hall, pg. 106-108. 2007.

[3] S. Audoly, G. Bellu, L. D'Angio, M.P. Saccomanni, and C. Cobelli. "Global identifiability of nonlinear models of biological systems." IEEE Trans. Biomedical Engineering, Vol. 48. 2001.

[4] J. S. Bader, A. Chaudhuri, J. M. Rothberg, and J. Chant. "Gaining confidence in high-throughput protein interaction networks." Nature Biotechnology, Vol. 22, pg. 78-85. 2003.

[5] H. Bozdogan. "Akaike's information criterion and recent developments in information complexity." Journal of Mathematical Psychology, Vol. 44. 2000.

[6] D. Braun, S. Basu, and R. Weiss. "Parameter estimation for two synthetic gene networks: a case study." IEEE International Conference on Acoustics, Speech, and Signal Processing. 2005.

[7] K.B. Burhnam and D. Anderson. Model Selection and Multi-Model Inference. Springer, 2^{nd} edition. 2003.

[8] V. Cerny. "Themodynamical approach to the traveling salesman problem - an efficient simulation algorithm." Journal of Optimization Theory and Applications. Vol. 45 (1), pg. 41-51 1985.

[9] M. Farina, R. Findeisen, E. Bullinger, S. Bittanti, F. Allgower, and P. Wellstead. "Results towards identifiability properties of biochemical reaction networks." IEEE Conference on Decision and Control. 2006.

[10] X.J. Feng and H. Rabitz. "Optimal identification of biochemical reaction networks." Biophysical Journal, Vol. 86. 2004.

[11] K.G. Gadkar, R. Gunawan, and F. J. Doyle III. "Iterative approach to model identification of biological networks." BMC Bioinformatics, Vol. 6, Issue 155. 2005.

[12] 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." Proceedings of the Pacific Symposium on Biocomputing. 2003.

[13] C.M. Hurvich and C-L. Tsai. "Model selection for extended quasi-likelihood models in small samples." Statistics and Probability Letters. Vol. 27. 1995.

[14] L. Ingber. "Very fast simulated re-annealing." Mathematical and Computer Modelling. Vol. 12 (8), pg. 967-973. 1989.

[15] F.J. Isaacs, D.J. Dwyer, and J.J. Collins. "RNA synthetic biology." Nature Biotechnology. Vol. 24, pg. 545-554. 2006.

[16] J.A. Jacquez and T. Perry. "Parameter estimation: local identifiability of parameters." American Journal of Physiology. Vol. 258. 1990.

[17] J. Kim. "*In vitro* synthetic transcriptional networks." PhD thesis, Caltech, Dec 2006.

[18] J. Kim, K.S. White, and E. Winfree. "Construction of an *in vitro* bistable circuit from synthetic transcriptional switches." Molecular Systems Biology. 2006.

[19] S. Kullback and R.A. Leibler. "On information and sufficiency." The Annals of Mathematical Statistics. Vol. 22, pg. 70-86. 1951.

[20] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. "Optimization by simulated annealing." Science Vol. 220, pg. 671-680. 1983.

[21] L. Ljung. System identification: theory for the user. Prentice Hall, Upper Saddle River, N.J. 1987.

[22] P. Mendes and D. B. Kell. "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation." Bioinformatics, Vol.14. 1998.

[23] C.G. Moles, P. Mendes, and J.R. Banga. "Parameter estimation in biochemical pathways: a comparison of global optimization methods." Genome Research. Vol. 13, pg. 2467-2474. 2003.

[24] N. Rosenfeld, J.W. Young, U. Alon, P.S. Swain, and M.B. Elowitz. "Gene regulation at the single-cell level." Science. Vol. 307, pg. 1962-5. 2005.

[25] K. Takeuchi. Distribution of informational statistics and criterion of model fitting. *Suri-Kagaku* (Mathematic Sciences), Vol. 153 (In Japanese).1976.

[26] M. Xiong, J. Li, and A. Fang. "Identification of genetic networks." Genetics. Vol. 166. 2004.

[27] K.Z. Yao, B.M. Shaw, B. Kou, K.B. McAuley, and D.W. Bacon. "Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design." Polymer Reaction Engineering. Vol. 11, pg. 563-588. 2003.