Available online at www.sciencedirect.com
ScienceDirect
Journal of the Chinese Institute of Chemical Engineers 39 (2008) 399–406
JOURNAL OF THE CHINESE INSTITUTE OF CHEMICAL ENGINEERS
www.elsevier.com/locate/jcice

# Inverse problems of biological systems using multi-objective optimization

Pang-Kai Liu, Feng-Sheng Wang *

*Department of Chemical Engineering, National Chung Cheng University, 168 University Road, Ming-Hsiung, Chia-yi 621-02, Taiwan*

Received 21 December 2007; received in revised form 7 May 2008; accepted 7 May 2008

**Abstract**

Mathematical modeling for dynamic biological systems is a central theme in systems biology. There are still many challenges in using time-course data to obtain an inverse problem of nonlinear dynamic biological systems. In this study, a multi-objective optimization technique is introduced to determine kinetic parameter values of biochemical reaction systems. The multi-objective parameter estimation was converted into the minimax problem through the satisfying trade-off method. The aspiration value was assigned as the minimum solution to the corresponding single objective estimation. The aim of this trade-off estimation was to obtain a compromised result by simultaneously minimizing both concentration and slope error criteria. Hybrid differential evolution was applied to solve the minimax problem and to yield a global estimation.


*Keywords:* Biochemical systems theory; Parameter estimation; Collocation method; Min–max optimization; Dynamic model; Hybrid differential evolution

## 1. Introduction

Mathematical models describing the behavior of micro-biological systems are attracting increasing attention in the post-genomic era (Voit, 2002). Such mathematical models provide a description of mechanism of biological systems which are required for analysis, design, optimization and control. The ultimate goal of mathematical modeling is to obtain an expression that quantitatively describes the dynamic behaviors of the system under consideration. The generality of a model depends upon several factors, which include the system's complexity and available information. Biochemical Systems Theory (BST) comes in two variants: generalized mass action (GMA) and S-system forms (Savageau, 1976; Voit, 2000). These offer alternative options for representing complex metabolic pathway systems. For the identification of structure from time-course data the S-system model is particularly useful, especially if limited additional information about the biological system is available. The flux aggregation is applied in S-systems to model branch and reverse pathways so some difficulties, such as mass conservation at a branch point, arise when dealing with networks of interacting biochemical reactions that feature branch and reverse pathways. The GMA model is better able to capture realistic reaction information such as branch or reverse pathways and greatly increases the mathematical tractability of the system while preserving its nonlinear nature (Marín-Sanguino and Torres, 2003; Polisetty *et al.*, 2006). The aim of this study is to introduce a multi-objective optimization approach to determine kinetic parameter values in GMA/S-system models using time-course observations.

Parameter estimation is an essential step in the verification and subsequent use of a mathematical model in the field of biological systems. There exists, however, no unique method to estimate model parameters for nonlinear power-law models (Chou *et al.*, 2006). Most of the traditional nonlinear regression algorithms based on gradient methods have the possibility of getting trapped at local optima, depending upon the degree of system nonlinearity and the initial starting point (Mendes and Kell, 1998). Alternating regression (Chou *et al.*, 2006) dissects the nonlinear inverse problem of estimating parameter values into iterative steps of linear regression. The branch and bound algorithm (Polisetty *et al.*, 2006) is applied to convert the inverse problem of GMA or S-system into a convex optimization problem in order to obtain a global solution.

Numerical integration for solving differential equations is another issue for inverse problems. It is time-consuming and may break down during the optimization searching. Tsai and Wang (2005) have compared three techniques of parameter estimation for nonlinear dynamic biological systems. Time-course slope information for a dynamic system has been applied to avoid

* Corresponding author. Fax: +886 5 2721206.
*E-mail address:* chmfsw@ccu.edu.tw (F.-S. Wang).

0368-1653/$ – see front matter 
doi:10.1016/j.jcice.2008.05.001

**Nomenclature**

| | |
|---|---|
| $f_i$ | kinetic order for GMA model |
| $g_{ij}$ | kinetic order for generation |
| $h_{ij}$ | kinetic order for degradation |
| $\hat{J}_i$ | aspiration value |
| $K_i$ | kinetic parameters for Michaelis–Menten equation |
| $V_i^{\max}$ | kinetic parameters for Michaelis–Menten equation |
| $X_i$ | state variable |

*Greek symbols*

| | |
|---|---|
| $\alpha_i$ | rate constant for generation |
| $\beta_i$ | rate constant for degradation |
| $\gamma_{ij}$ | rate constant for GMA model |
| $\nu_{ij}$ | influx into pool or efflux out of pool |

numerical integration; using the dynamic model to compute the slope error information can alleviate computational burden (Lall and Voit, 2005; Voit, 2000; Voit and Almeida, 2004). As discussed in Voit (2000), very small error values may be obtained from a time-course slope. However, dynamic profiles of the model may not be adequate at the experimental concentrations. Tsai and Wang (2005) have applied hybrid differential evolution with the modified collocation method to determine the model parameters of the S-system. The modified collocation method is applied to convert ordinary differential equations into algebraic equations in order to yield approximate dynamic profiles. By contrast, the modified collocation approximation may yield a very small concentration error criterion value. However, when the estimated parameters are applied to compute the slope profile, this may yield a large slope error criterion. This indicates that the characteristic for each dynamic function, equivalent to the net rate equations, is different from the experimental observation. Such estimated parameters are found, in general, to have poor predictive ability. In this study, multi-objective parameter estimation is introduced to simultaneously minimize both criteria in order to overcome the drawbacks mentioned above and to obtain a trade-off solution. Hybrid differential evolution is then applied to solve the multi-objective parameter estimation problem and move towards yielding an acceptable solution (Wang and Sheu, 2000).

The paper is organized as follows. The following section introduces single objective parameter estimation and multi-objective approach. To illustrate the effectiveness of the proposed algorithm, two dry-lab case studies and a single wet-lab case study are presented in Section 3. Finally, concluding remarks are made in Section 4.

## 2. Method

### 2.1. Single objective estimation

The metabolic reaction systems of a biological entity can be described mathematically by GMA formulation, as follows (Voit, 2000):

$$\dot{\mathbf{X}} = \mathbf{f}(\mathbf{X}, \mathbf{p}) = \begin{bmatrix} \sum_{k=1}^{n_{g_1}} \alpha_{1k} \prod_{j=1}^{n} X_j^{g_{1jk}} - \sum_{k=1}^{n_{h_1}} \beta_{1k} \prod_{j=1}^{n} X_j^{h_{1jk}} \\ \vdots \\ \sum_{k=1}^{n_{g_n}} \alpha_{nk} \prod_{j=1}^{n} X_j^{g_{njk}} - \sum_{k=1}^{n_{h_n}} \beta_{nk} \prod_{j=1}^{n} X_j^{h_{njk}} \end{bmatrix},$$

$$\mathbf{X}(0) = \mathbf{X}_0 \tag{1}$$

where $\mathbf{X}$ represents $n$-dimensional components or pools, and the parameter vector $\mathbf{p}$ consists of rate constants, $\alpha_{ik}$ and $\beta_{ik}$, and kinetic orders, $g_{ijk}$ and $h_{ijk}$. $\mathbf{f}$ is a vector of net rate equations. Each element of $\mathbf{f}$ is composed of all production and degradation terms for the corresponding component. The parameter estimation is to determine rate constants and kinetic orders so that the dynamic profiles satisfactorily fit the measured observation.

Since we have no information on most of the model parameters, we must infer them from the data. Therefore, we optimize an error criterion which measures the discrepancy of simulated data from the real data. The error criterion in conventional parameter estimation techniques is evaluated from the measured concentrations and dynamic profiles computed using ordinary differential equations. Such a criterion is expressed as

$$J_{1l} = \frac{1}{nN_s} \sum_{i=1}^{n} \sum_{s=1}^{N_s} t_s \frac{(X_{e_i}(t_s) - X_i(t_s))^2}{X_{e_i\max}^2}, \quad l = 1, \ldots, N_{\exp} \tag{2}$$

where $X_{e_i}(t_s)$ is the measured data for the $i$th component at $t = t_s$, $X_i(t_s)$ is the computed concentration for the $i$th component at $t = t_s$, and $X_{e_i\max}$ is the maximum measured concentration of the $i$th component. Here, $N_s$ is the number of sampled data points and $N_{\exp}$ is the number of experiments. The time weighting factor $t_s$ is added in the criterion to highlight that the estimated result is more important as the time progress. Tsai and Wang (2005) have applied the modified collocation method (Wang, 2000) to convert ordinary differential equations into algebraic equations to yield the approximate dynamic profiles as follows:

$$\begin{bmatrix} X_1(t_s) \\ \vdots \\ X_n(t_s) \end{bmatrix} \cong \begin{bmatrix} X_{e_1}(t_{s-1}) \\ \vdots \\ X_{e_n}(t_{s-1}) \end{bmatrix} + 0.5\eta_s \left\{ \begin{bmatrix} f_1(\mathbf{X}_e(t_s), \mathbf{p}) \\ \vdots \\ f_n(\mathbf{X}_e(t_s), \mathbf{p}) \end{bmatrix} + \begin{bmatrix} f_1(\mathbf{X}_e(t_{s-1}), \mathbf{p}) \\ \vdots \\ f_n(\mathbf{X}_e(t_{s-1}), \mathbf{p}) \end{bmatrix} \right\},$$

$$s = 1, \ldots, N_s \tag{3}$$

The approximate profiles can be directly applied to evaluate the error criterion as shown in Eq. (2) to avoid numerical integration for dynamic Eq. (1). The decomposition strategy has been previously applied to solve differential equations in parallel to