Mathematical Biosciences 282 (2016) 21–33

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

A new parametric method to smooth time-series data of metabolites in metabolic networks

Atsuko Miyawaki[a], Kansuporn Sriyudthsak[b], Masami Yokota Hirai[b], Fumihide Shiraishi[a,*]

[a] Section of Bio-Process Design, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, 6-10-1, Hakozaki, Higashi-Ku, Fukuoka 820-8581, Japan
[b] RIKEN Center for Sustainable and Resource Science, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

ARTICLE INFO

Article history:
Received 17 March 2016
Revised 29 August 2016
Accepted 20 September 2016
Available online 28 September 2016

Keywords:
Data smoothing
Parameter estimation
Biochemical systems theory
Convergent calculation
Metabolite concentration

ABSTRACT

Mathematical modeling of large-scale metabolic networks usually requires smoothing of metabolite time-series data to account for measurement or biological errors. Accordingly, the accuracy of smoothing curves strongly affects the subsequent estimation of model parameters. Here, an efficient parametric method is proposed for smoothing metabolite time-series data, and its performance is evaluated. To simplify parameter estimation, the method uses S-system-type equations with simple power law-type efflux terms. Iterative calculation using this method was found to readily converge, because parameters are estimated stepwise. Importantly, smoothing curves are determined so that metabolite concentrations satisfy mass balances. Furthermore, the slopes of smoothing curves are useful in estimating parameters, because they are probably close to their true behaviors regardless of errors that may be present in the actual data. Finally, calculations for each differential equation were found to converge in much less than one second if initial parameters are set at appropriate (guessed) values.

1. Introduction

Analytical instruments have greatly improved in the past two decades, and it is now possible to comprehensively measure intracellular metabolite concentrations [1,2]. If these metabolome data are then used to construct a mathematical model, one will be able to simulate actual metabolic reactions in cells, and thereby elucidate reaction mechanisms. However, the data are invariably influenced by measurement errors and individual differences [3]. To overcome this problem, the data can initially be smoothed and the smoothing curves can be used to extract data points to construct a mathematical model [4,5]. If a simple structure is contained in a network the equation determined by the smoothing method can partly be adopted in the finally constructed mathematical model. Unfortunately, the mathematical function used to smooth data is in most cases entirely irrelevant to the phenomenon of interest. In any case, smoothing curves would probably be close to true values when measured data are not widely dispersed. However, highly scattered data provide innumerable possible curves. Furthermore, when data for each metabolite are smoothed independently, the resulting curves usually do not satisfy mass balances. Consequently, the use of smoothed values can be questionable.

Time-dependent metabolite concentrations can be represented approximately in terms of an S-system or a GMA-system within the framework of biochemical systems theory [6–8]. Iwata et al. [9] examined how much a simplified S-system equation with an efflux term agrees with the time course of each metabolite when the values of two unknown parameters, *i.e.*, rate constant and kinetic order, are adjusted. This is based on the fact that the time course of each metabolite depends on its concentration in a pool from which it outflows. Results showed that the efflux term, expressed as a function of the concentration of the relevant metabolite, can generate satisfactory agreement between calculated and measured time courses. This finding suggests that the calculated time courses can thus be used as smoothing curves. However, they do not contain any other effects such as feedback inhibition and, therefore, may be incorrect biochemically. Nevertheless, the smoothing method is physicochemically meaningful, as it is based on the S-system representation. Moreover, since the time course of metabolites reflects network structures, S-system equations with parameters determined from time-series data would by extension also generate smoothing curves that reflect network structures.

In this study, we attempt to establish a data smoothing method that creates calculated values that satisfy mass balances, and also enables easy estimation of rate constants and kinetic orders. The fundamental equation used in this method is an S-system representation with an efflux term in the form of a simple power

* Corresponding author. Fax: +81 92 642 7603.
E-mail address: fumishira@brs.kyushu-u.ac.jp (F. Shiraishi).

http://dx.doi.org/10.1016/j.mbs.2016.09.011
0025-5564/

law. Moreover, we investigate a convergent calculation by which several parameters in the efflux terms are estimated rapidly and efficiently by least-squares. To evaluate performance, the proposed smoothing method is applied to several network models.

2. Theory

2.1. S-system representation

The rate of change of metabolite concentrations can be described as an S-system-type differential equation:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n} X_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n} X_j^{h_{i,j}} = V_i - V_{-i} \ (i = 1, 2, \ldots, n), \tag{1}$$

where X_i (i=1, …, n) represents metabolite concentrations, V_i and V_{-i} are influx and efflux terms, α_i and β_i (i=1, …, n) are the rate constants in influx and efflux terms, $g_{i,j}$ and $h_{i,j}$ (i,j=1, …, n) are the kinetic orders of influx and net efflux, and n is the number of dependent variables. Eq. (1) is the canonical representation generalized for a metabolic reaction network, and ordinary network systems can be written in simpler forms. Nevertheless, even such simple systems have a number of parameters that are not easy to determine from time-series data.

The present study proposes a data smoothing method based on Eq. (1). However, the number of parameters to be estimated should be reduced as much as possible. As efflux from a metabolite pool depends on the metabolite concentration in that pool, we can express each differential equation simply as

$$\frac{dX_i}{dt} = V_i - \beta_i X_i^{h_{i,i}} \quad (i = 1, 2, \ldots, n). \tag{2}$$

Eq. (2), a simpler S-system form, is reasonable, because efflux is expressed as a function of the concentration of the relevant metabolite, and a large body of research indicates that the S-system model (or power-law function) can suitably capture the behavior of metabolites. Our data-smoothing method determines the parameters in each differential equation stepwise, from upstream to downstream metabolites. Finally, to gain a high probability of convergence, we assume in Eq. (2) that the influx V_i is known, while both β_i and $h_{i,i}$ are unknown.

If the downstream metabolite X_k imposes feedback inhibition strongly on an enzyme that catalyzes an efflux, its concentration may decrease or oscillate. Eq. (2) cannot capture such a behavior, and therefore we use

$$\frac{dX_i}{dt} = V_i - \beta_i X_i^{h_{i,i}} X_k^{h_{i,k}} \quad (i = 1, 2, \ldots, n), \tag{3}$$

where β_i, $h_{i,i}$, and $h_{i,k}$ are unknown.

Hereafter, we use Method I to refer to the proposed data-smoothing method, and we use Method II to refer to the least-squares method of estimating parameters in a polynomial equation.

2.2. Pathway types and a method for convergent calculation

Let us consider metabolic reaction pathways consisting of elementary network structures shown in Fig. 1. These include a linear pathway with feedback inhibition, as well as one-metabolite, two-metabolite, diverging-branch, and converging-branch pathways. We now describe the procedure for estimating parameters in each network structure.

2.2.1. Linear pathway with a constant influx

When a highly abundant metabolite is the starting substrate in a network (Fig. 1(a)), its concentration changes very slowly and can be approximated as constant. Thus, the S-system equation is given as

$$\frac{dX_1}{dt} = \alpha_1 - \beta_1 X_1^{h_{1,1}}, \tag{4}$$

$$\frac{dX_2}{dt} = \beta_1 X_1^{h_{1,1}} - \beta_2 X_2^{h_{2,2}}. \tag{5}$$

where α_1 in the influx term is constant and is assumed to be known *a priori*. This parameter, for example, may represent uptake of glucose, a flux that can be measured easily. Let us now attempt to determine β_1 and $h_{1,1}$ so that $X_{1,i}$ ($i = 1, \ldots, N$) values calculated using Eq. (4) at t_i ($i = 1, \ldots, N$) agree with the measured values ${X_{1,i}}^{\exp}$ ($i = 1, \ldots, N$), where N represents the number of calculated or measured values. Least-squares provides the sum of the squared differences between calculated and measured values, $S(\beta_1, h_{1,1})$, according to

$$S(\beta_1, h_{1,1}) = \sum_{i=1}^{N} \left(X_{1,i} - {X_{1,i}}^{\exp}\right)^2. \tag{6}$$

We adopt the Newton–Raphson method as a root-finding technique to determine unknown parameters. This method is a very simple algorithm that converges rapidly [10–12]. Eq. (6) is partially differentiated with respect to β_1 and $h_{1,1}$, and the resulting equations are set to zero [13]. Consequently, the following equations are derived:

$$f_1(\beta_1, h_{1,1}) = \sum_{i=1}^{N} \frac{\partial X_{1,i}}{\partial \beta_1} X_{1,i} - \sum_{i=1}^{N} \frac{\partial X_{1,i}}{\partial \beta_1} {X_{1,i}}^{\exp} \tag{7}$$

$$f_2(\beta_1, h_{1,1}) = \sum_{i=1}^{N} \frac{\partial X_{1,i}}{\partial h_{1,1}} X_{1,i} - \sum_{i=1}^{N} \frac{\partial X_{1,i}}{\partial h_{1,1}} {X_{1,i}}^{\exp}. \tag{8}$$

When $f_1(\beta_1, h_{1,1})$ and $f_2(\beta_1, h_{1,1})$ are set to zero, Eqs. (7) and (8) provide algebraic equations, which can be solved with respect to β_1 and $h_{1,1}$ by the Newton–Raphson method. The convergent calculation is carried out using

$$\begin{bmatrix} {\beta_1}^{(p)} \\ {h_{1,1}}^{(p)} \end{bmatrix} = \begin{bmatrix} {\beta_1}^{(p-1)} \\ {h_{1,1}}^{(p-1)} \end{bmatrix} - \begin{bmatrix} \dfrac{\partial f_1\left({\beta_1}^{(p-1)}, {h_{1,1}}^{(p-1)}\right)}{\partial \beta_1} & \dfrac{\partial f_1\left({\beta_1}^{(p-1)}, {h_{1,1}}^{(p-1)}\right)}{\partial h_{1,1}} \\ \dfrac{\partial f_2\left({\beta_1}^{(p-1)}, {h_{1,1}}^{(p-1)}\right)}{\partial \beta_1} & \dfrac{\partial f_2\left({\beta_1}^{(p-1)}, {h_{1,1}}^{(p-1)}\right)}{\partial h_{1,1}} \end{bmatrix}^{-1} \begin{bmatrix} f_1\left({\beta_1}^{(p-1)}, {h_{1,1}}^{(p-1)}\right) \\ f_2\left({\beta_1}^{(p-1)}, {h_{1,1}}^{(p-1)}\right) \end{bmatrix}, \tag{9}$$

where the pth estimates are obtained from the $(p-1)$th estimates. This iterative calculation provides solutions to β_1 and $h_{1,1}$. It should be noted that this calculation requires the values of

$$\frac{\partial X_{1,i}}{\partial \beta_1}, \frac{\partial X_{1,i}}{\partial h_{1,1}}, \frac{\partial}{\partial \beta_1}\left(\frac{\partial X_{1,i}}{\partial \beta_1}\right), \frac{\partial}{\partial h_{1,1}}\left(\frac{\partial X_{1,i}}{\partial \beta_1}\right),$$

$$\text{and } \frac{\partial}{\partial h_{1,1}}\left(\frac{\partial X_{1,i}}{\partial h_{1,1}}\right) \ (i = 1, \ldots, N)$$

at each time point t_i (i=1, … , N). These values must be obtained from the following equations, which are derived by partially differentiating Eq. (4) with respect to β_1 and $h_{1,1}$:

$$\frac{d}{dt}\frac{\partial X_1}{\partial \beta_1} = -{X_1}^{h_{1,1}} - \beta_1 h_{1,1} {X_1}^{h_{1,1}-1} \frac{\partial X_1}{\partial \beta_1} \tag{10}$$

$$\frac{d}{dt}\frac{\partial X_1}{\partial h_{1,1}} = -\beta_1 \frac{\partial {X_1}^{h_{1,1}}}{\partial h_{1,1}} \tag{11}$$