Journal of Biotechnology 149 (2010) 132–140

Contents lists available at ScienceDirect

Journal of Biotechnology

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

Parameter identifiability of power-law biochemical system models

Sridharan Srinath, Rudiyanto Gunawan*

Department of Chemical and Biomolecular Engineering, National University of Singapore, Blk E5, 4 Engineering Drive 4, #02-16, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 27 July 2009
Received in revised form 19 February 2010
Accepted 22 February 2010

Keywords:
Identifiability analysis
Power-law models
Inverse modeling
Confidence region
Biochemical Systems Theory

ABSTRACT

Mathematical modeling has become an integral component in biotechnology, in which these models are frequently used to design and optimize bioprocesses. Canonical models, like power-laws within the Biochemical Systems Theory, offer numerous mathematical and numerical advantages, including built-in flexibility to simulate general nonlinear behavior. The construction of such models relies on the estimation of unknown case-specific model parameters by way of experimental data fitting, also known as inverse modeling. Despite the large number of publications on this topic, this task remains the bottleneck in canonical modeling of biochemical systems. The focus of this paper concerns with the question of identifiability of power-law models from dynamic data, that is, whether the parameter values can be uniquely and accurately identified from time-series data. Existing and newly developed parameter identifiability methods were applied to two power-law models of biochemical systems, and the results pointed to the lack of parametric identifiability as the root cause of the difficulty faced in the inverse modeling. Despite the focus on power-law models, the analyses and conclusions are extendable to other canonical models, and the issue of parameter identifiability is expected to be a common problem in biochemical system modeling.

1. Introduction

Mathematical modeling has become an indispensable tool in biotechnology and biological studies with myriad applications from metabolic engineering to cancer therapy. The procedure of model construction is generally iterative, in which wet-lab experiments generate the biological observations and data needed for model formulation and identification, while *in silico* simulations are used to (in)validate models and to design the most informative experiments (Chou and Voit, 2009; Kitano, 2002a,b). Biochemical models are typically formulated according to physicochemical laws, such as mass or molar balance, and mathematical equations are written down to describe the rate of reactions or transformations among different molecules. By using different approximations of the nature of these transformations, numerous mathematical frameworks, e.g. Boolean, ordinary differential equation (ODE), and stochastic chemical master equation, have been used in biochemical system modeling. These models commonly contain unknown model parameters, which are usually determined such that model simulations can reproduce experimental observations. This step, known as inverse modeling, is often the limiting step in biochemical model construction (Banga and Balsa-Canto, 2008; Chou and Voit, 2009; van Riel, 2006).

Among all mathematical frameworks, ODEs are by far the most commonly used and most relevant for biotechnological applications. In general, the mathematical equations are written as $\dot{\mathbf{X}} = f(\mathbf{X}, \boldsymbol{\theta})$, where $\mathbf{X}$ denotes the concentration vector of biomolecules like metabolites, $\boldsymbol{\theta}$ is the parameter vector, and $f(\mathbf{X},\boldsymbol{\theta})$ is a general vector-valued nonlinear equation. While ODE models predict the concentrations as a function of time, i.e. they are dynamical models, steady state models can be directly derived from ODEs by setting $\dot{\mathbf{X}} = 0$. In metabolic networks, $f(\mathbf{X},\boldsymbol{\theta})$ is usually expanded into $\mathbf{N}\mathbf{v}(\mathbf{X},\boldsymbol{\theta})$ where the matrix $\mathbf{N}$ gives the stoichiometric relationships of the metabolic transformations with rates or fluxes that are given by the vector $\mathbf{v}(\mathbf{X},\boldsymbol{\theta})$. The functionality of $\mathbf{v}(\mathbf{X},\boldsymbol{\theta})$ should depend on the mechanisms by which the enzymatic transformation from one metabolite to another follows, e.g. Michaelis–Menten law. However, to capture more general nonlinear behavior, power-law approximations have been frequently used as canonical ODE models, such as generalized mass action (GMA) and S-system models. GMA models are generally written according to:

$$\dot{X}_i = \sum_{k=1}^{k_i} \left(\pm\gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{ikj}} \right). \quad (1)$$

where X_i here denotes the ith metabolite, n and m are the number of dependent and independent metabolites, respectively, γ_{ik}s denote the rate constants and f_{ikj}s denote the kinetic order parameters. On the other hand, in S-system models, influxes and effluxes of

* Corresponding author. Tel.: +65 6516 6617; fax: +65 6779 1936.
E-mail address: chegr@nus.edu.sg (R. Gunawan).

0168-1656/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2010.02.019

metabolites are combined into individual power-law terms, giving:

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}, \quad (2)$$

where the parameters α_i and β_i are the (non-negative) overall rate constants, and g_{ij} and h_{ij} are the effective kinetic order parameters. These and other canonical ODE models such as Lotka-Volterra models (Hernández-Bermejo and Fairén, 1997), linear-logarithmic approximation (Hatzimanikatis and Bailey, 1996), and saturable and cooperative (SC) formalism (Sorribas et al., 2007) permit the development of model identification procedures and quantitative/numerical analyses that can be tailored and highly optimized for each formalism, such as those within the Biochemical Systems Theory (BST) for power-law models (Marino and Voit, 2006; Voit and Almeida, 2004; Voit and Ferreira, 2000).

Power-law models have wide applications not only in metabolic networks (Voit and Ferreira, 2000), but also in other biological systems such gene networks and signal transduction pathways (Hlavacek and Savageau, 1996; Kimura et al., 2004; Vera et al., 2007), which motivates the focus on these models here. The benefits and limitations of power-law approximations have also been detailed elsewhere (see Chou and Voit, 2009 and references therein). Among the advantages of power-law models, there are two worth mentioning here. The first is the flexibility of power-law formalism in approximating a multitude of (nonlinear) cellular responses, which is the reason for its wide applicability in modeling general biochemical systems. The second advantage relates to the direct one-to-one relationship between parameter values and network structure (see Section 2 for further explanation). Hence, the structural and kinetic identification in biochemical model construction can be formulated as a combined parameter identification problem.

Some of the rate constants and kinetic orders in power-law models can be set to zero from the available knowledge of network structure. However, most parameters are usually estimated from experimental data, among which dynamical data are most informative and of interest in this paper. Although there exist a large body of works on the parameter estimation of power-law models, which have been discussed at length in a recent review (Chou and Voit, 2009), this problem remains unsolved and has become the bottleneck in model identification within the BST. In other applications of ODE modeling, parameter identifiability or more specifically the lack of it has been shown to be an issue, in which there can be many parameter combinations that reproduce the same dynamic data (Chu and Hahn, 2009; Hengl et al., 2007; Zak et al., 2003). The lack of parameter identifiability can lead to grossly inaccurate parameter estimates, rendering the model useless for downstream applications, such as process or strain optimization for the production of certain metabolites. This identifiability problem seems to plague the parameter estimation of power-law models (Vilela et al., 2009), but parameter identifiability has not been addressed in this context. This paper aims to fill this gap. In particular, existing and newly developed methods are applied to investigate the parameter identifiability in two previously published power-law models of metabolic networks: lactate production in *L. lactis* (GMA) (Voit et al., 2006) and *E. coli* metabolism (S-system) (Ko et al., 2006).

2. Methods

In this paper, model identifiability is defined as the ability to uniquely determine model structure and parameters from a given set of experimental data (Carson et al., 1983). Although this definition lacks mathematical rigor, we will show that it has a practical relevance in the development of identifiability analyses. Also, the "uniqueness" requirement can be relaxed when considering noisy measurements. In power-law formalism, model structure, i.e. the connectivity among states, can be inferred from the values of the kinetic order parameters, where a positive value indicates a substrate or a regulatory activation and a negative value implies a regulatory inhibition. Therefore, model identifiability is equivalent to parameter identifiability in this context.

The topic of parameter identifiability is well established in mathematical modeling within science and engineering, including biotechnology (Godfrey, 1986; Jimenez-Hornero et al., 2008; Nikerel et al., 2008). In general, there are three key factors that influence parameter identifiability from experimental data, namely the (1) degrees of freedom (DOF), (2) parameter correlation, and (3) data noise. The DOF, which is calculated as the difference between the number of parameters and data points (not counting replicates), refers to the number of model parameters that can be changed independently without affecting the outcome of data fitting. The second factor describes the similarity among model parameters, in which a high positive (negative) correlation between two parameters indicates that these parameters affect the model simulation in the same (opposite) manner. Thus, two highly correlated parameters have a compensatory effect in which their values can be varied together with little change in the simulation outcomes. Finally, the third factor relates to how random noise contaminates true measurement signals and thus decreases the information that is useful for parameter estimation.

Generally, there are two types of parameter identifiability; the first assumes noise-free data, which is referred to as *structural* or *a priori* identifiability, while the second considers data quality, referred to as *practical* identifiability. The first two of the aforementioned factors are considered in *a priori* identifiability, while practical identifiability takes into account all three factors. The following sections present methods to study these identifiability conditions. As power-law models are dynamic, the most useful information for parameter estimation is naturally dynamical data, i.e. time-series concentration data. The analysis of *a priori* identifiability is based on the sensitivity matrix following an existing method (Yao et al., 2003). In addition, three practical identifiability methods are developed and applied based on statistical analyses of the parameter estimation problem. Although we focus on power-law models, these analyses have applicability to general ODE models.

2.1. A priori parameter identifiability analysis

A priori or structural identifiability analysis addresses the uniqueness of the inverse modeling problem under the assumption that the data are noise-free. The lack of *a priori* identifiability typically arises due to model over-parametrization. The term "*a priori*" implies that the analysis can or should be done before an experiment is carried out, whereas the term structural can be taken to mean that the outcome should depend only on the model structure (Godfrey, 1986). Structural identifiability in metabolic pathways using the power-law formalism was previously studied by Sorribas and Cascante (1994) in the situation where steady state data are available for model structure determination. When considering general dynamical models and data, global *a priori* parameter identifiability methods have been developed based on Taylor series expansion, Lie derivatives or differential algebra, which have been mainly applied to linear systems and simple nonlinear systems (Jimenez-Hornero et al., 2008).

The method described below is based on the first order derivatives of the model (measured) output with respect to the parameters, also called the sensitivity matrix **S** (Ingalls, 2008; Turányi, 1990; Varma et al., 1999). The sensitivity matrix reflects how much changes in the parameter values will affect the output. If the outputs have zero sensitivity with respect to a parameter, then