Mathematical Biosciences 000 (2016) 1–17

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Nonparametric dynamic modeling

Mojdeh Faraji, Eberhard O. Voit*

Department of Biomedical Engineering, Georgia Institute of Technology, 950 Atlantic Drive, Suite 2115, Atlanta, GA 30332-2000, USA

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Dynamic Flux Estimation (DFE)
Metabolic Pathway Analysis
Nonlinear Compartment Model
Pathway Structure Identification
Systems Biology

ABSTRACT

Challenging as it typically is, the estimation of parameter values seems to be an unavoidable step in the design and implementation of any dynamic model. Here, we demonstrate that it is possible to set up, diagnose, and simulate dynamic models without the need to estimate parameter values, if the situation is favorable. Specifically, it is possible to establish nonparametric models for nonlinear compartment models, including metabolic pathway models, if sufficiently many high-quality time series data are available that describe the biological phenomenon under investigation in an appropriate and representative manner. The proposed nonparametric strategy is a variant of the method of Dynamic Flux Estimation (DFE), which permits the estimation of numerical flux profiles from metabolic time series data. However, instead of attempting to formulate these numerical profiles as explicit functions and to optimize their parameter values, as it is done in DFE, the metabolite and flux profiles are used here directly as a scaffold for a library from which values are interpolated and retrieved for the simulation of the differential equations describing the model. Beyond simulations, the proposed methods render it possible to determine steady states from non-steady state data, perform sensitivity analyses, and estimate the Jacobian of the system at a steady state.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Ever since the digital revolution drove analog computing to the brink of extinction, the design of computational models for complex systems has become an effort in choosing optimal mathematical representations and their parameter values. For most physical and engineering systems, the choice of model functions is directly guided by our rather solid understanding of basic physical concepts, such as mechanical or electrical forces, dilution and dispersion phenomena, optical processes, and the features of electric circuits. Biological systems are, of course, objects of the physical world and must therefore obey the laws of physics, but most processes that govern even moderately sized biological systems are so convoluted that they cannot be dissected into elementary physical representations [1]. As an example, the transmission of a neuronal signal at a dopamine synapse requires electrical activation, the prior biochemical production of dopamine and its packaging into membrane vesicles, the move of these vesicles through the crowded cytoplasm toward the synapse, the merging of vesicle and cell membranes, the opening of this membrane toward the synapse, the release of dopamine out of the vesicle and through the synaptic cleft to a receptor on the postsynaptic neuron, possible interactions with other neurotransmitters, and binding to the receptor. This binding in turn triggers a slew of additional mechanisms inside the signal receiving cell, including the complex process of signal interpretation which in the case of dopamine is often accomplished through multiple phosphorylation of the specific protein DARPP32, and the possible long-term adaptation to repeated stimuli [2–5]. Thus, a very coarse model could easily capture the fact that a signal moved from one neuron to another, but a detailed mechanistic model becomes quickly bogged down in the minutiae of the numerous intertwined biophysical processes that are involved in signal transduction.

Because elementary physical descriptions are often infeasible, the biological systems modeler is forced to resort to "higher-order" process representations, *ad hoc* models, suitable approximations, or combinations thereof. A good example is the Michaelis–Menten function of enzyme kinetics [6]. Its underlying concept is a process that postulates the reversible formation of a biochemical complex between an enzyme and its substrate and the subsequent release of the product of the reaction and of the enzyme, which is used over and over again. Under idealistic conditions *in vitro*, this concept is believed to be quite realistic. However, within living cells, the prerequisites for the involved mass-action functions are clearly not satisfied, and the so-called quasi-steady-state assumption, which is needed to formulate the process with a simple,

* Corresponding author.
E-mail addresses: mojdeh@gatech.edu (M. Faraji), eberhard.voit@bme.gatech.edu (E.O. Voit).

http://dx.doi.org/10.1016/j.mbs.2016.08.004
0025-5564/© 2016 Elsevier Inc. All rights reserved.

explicit function, only holds under certain conditions. Thus, an idealized concept, formulated with the help of somewhat doubtful approximations, becomes a higher-order process representation for enzyme catalyzed reactions. Indeed, the Michaelis–Menten function performs well *in vitro* and, in an approximate sense, presumably *in vivo*, although this is not really known. For simulations of large pathway systems, this function is often used as well, but its mathematical features become rather cumbersome, even for standard model assessments such as sensitivity analyses [7].

Notwithstanding these mathematical issues, it is common for the biological modeling community to base simulation studies in a variety of fields on a rather small set of functions, which are used time and again and prominently include mass-action, Michaelis–Menten and Hill functions, which often include regulatory terms [8]. The users of these functions rely on the argument that these functions suit their purposes—quasi as black boxes—and are sufficiently accurate if one considers the typical noise encountered in biological data. Furthermore, these particular functions at least have some foundation and rationale in biology, whereas the use of a function like a shifted arctangent has very little justification, except that its graph is s-shaped and therefore might resemble some saturation processes in biology.

True alternatives to these *ad hoc* approaches are generic approximations. Linearization, the simplest of these, has been enjoying enormous successes in engineering applications for many decades. For the representation of biological phenomena, by contrast, linear models tend to run into conflicts with the genuine nonlinearities that characterize living systems. For instance, common features like saturation, stable oscillations, threshold phenomena, synergisms, or chaos cannot directly be modeled with linear equations. A logical solution might seem to be the expansion of linear models to second-order Taylor approximations, but these become so awkward for larger systems [9] that very few modelers have resorted to this option. Instead, many biological modeling groups have been using power-law approximations, which are nonlinear, but have linear characteristics in logarithmic space. Biochemical Systems Theory (BST; [10–13]) and Metabolic Control Analysis (MCA; [14–17]), which directly or indirectly utilize power-law representations [18,19], respectively, have had success with analyses of a wide variety of complex biological systems (for a review, see [20]). Notwithstanding their successes, power-law representations are local approximations and therefore genuinely limited in their accuracy of capturing phenomena over large ranges of variation in the involved variables. As a case in point, univariate power-law functions in BST do not saturate for large substrate concentrations, and lin-log models, which are associated with MCA, become negative for small substrate concentrations and tend toward $-\infty$ for substrate concentrations approaching 0 [21–23]. As an alternative to these canonical power-law models, one could use sigmoidal basis functions, but for realistic models this option requires correspondingly larger numbers of parameters that need to be estimated [24,25].

Even if reasonable guideposts could be found to justify the choice of appropriate model representations, the second step of model identification is still to be performed, namely the estimation of parameter values. For moderately sized or large models, this estimation is always challenging [26–28], due to noise in the data, non-convergence of the search algorithm and other problems, or because the wrong model was chosen after all. To make matters worse, even an excellent fit is not necessarily optimal, and the parameterized model may perform poorly in extrapolations, because the original fit was obscuring the compensation of errors among some terms within the model (*e.g.*, see [29,30]). Furthermore, an excellent fit may be the result of overfitting with a model containing too many parameters.

These challenges and compromises lead to the obvious question of whether it might be possible to glean appropriate functions directly from experimental biological data, without presupposing potentially unjustified mathematical formats. The method of Dynamic Flux Estimation (DFE), which permits a relatively unbiased estimation of fluxes within a system and which will be reviewed later, took a first step toward answering this question affirmatively, at least for metabolic systems under ideal conditions [31]. Still, DFE requires some choices of model frameworks when the task is setting up a model from scratch.

In this paper, we describe a novel variant of DFE that makes such choices unnecessary, at least under favorable conditions. Given such conditions, the overall result of the proposed strategy is that it is possible to develop dynamic models in a nonparametric manner. Intriguingly, the resulting nonparametric models, which make no assumptions regarding parameter values or even mathematical formats, beyond the topology of the system, permit most of the typical diagnoses and analyses that are possible with a fully parametric model, which may be considered the gold standard in the field. As a consequence, simulations and other analyses can be performed without the complicated and often biased step of choosing models and parameterizing them, if suitable data are available. The data needed for this purpose consist of sets of time series that representatively capture the dynamics of a system under relevant inputs.

Both DFE and the nonparametric variant proposed here are particularly well suited for nonlinear, dynamic, regulated compartment models, because these possess the property of mass conservation, which imposes strong, unbiased constraints that greatly aid the formulation of appropriate models. As an illustration, and for ease of discussion, we will focus here on metabolic pathway systems, but it appears that other nonlinear compartment systems, such as SIR models of epidemiology and pharmacokinetic models, can be treated in the same manner.

2. Methods

2.1. Dynamic flux estimation (DFE)

The stoichiometric equation

$$\dot{\mathbf{X}} = \mathbf{S} \cdot \mathbf{V} \tag{1}$$

provides a generic description of the dynamics of a metabolic pathway system. This well-known equation collectively formulates dynamic changes in each metabolite of the system, $\frac{dX}{dt} = \dot{X}$, as a product between the stoichiometric matrix **S** and a vector of reactions or fluxes, **V**. This product formulation is remarkable, as it naturally separates the linear aspects of the system from its nonlinear features. Specifically, consider the situation where the slopes of all metabolites on the left-hand sides are known for some given time point. If so, Eq. (1) is a system of linear algebraic equations, where each variable V_j represents the state of a flux at this time point, rather than a metabolite. The nonlinear features enter the system secondarily, by virtue of the fact that each component of the flux vector is a possibly complicated function of metabolites and regulators, and therefore of time. Dynamic Flux Estimation (DFE) makes maximal use of this separation of the model into linear and nonlinear components.

In typical analyses, such as Flux Balance Analysis, the stoichiometric Eq. (1) is studied at a steady state of the system [32–34], where the vector on the left-hand side contains zeros. DFE reaches beyond the steady state, by addressing the system at many time points of a system's trajectory, where the vector of derivatives is different from zero. In its first phase, DFE uses time series measurements of metabolite concentrations, X_1, ..., X_n, along with estimates of the slopes of these time courses. Thus, DFE evaluates

Please cite this article as: M. Faraji, E.O. Voit, Nonparametric dynamic modeling, Mathematical Biosciences (2016), http://dx.doi.org/10.1016/j.mbs.2016.08.004

Download English Version:

https://daneshyari.com/en/article/5760409

Download Persian Version:

https://daneshyari.com/article/5760409

Daneshyari.com