Journal of Theoretical Biology 264 (2010) 528–537

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Cellular signaling identifiability analysis: A case study

Ryan T. Roper [a,*], Maria Pia Saccomani [b], Paolo Vicini [a,1]

[a] Resource Facility for Population Kinetics, Department of Bioengineering Box 355061, University of Washington, Seattle, WA 98195-5061, United States
[b] Department of Information Engineering, University of Padova, Padova 35131, Italy

ARTICLE INFO

Article history:
Received 23 June 2009
Received in revised form
9 February 2010
Accepted 18 February 2010
Available online 24 February 2010

Keywords:
Global identifiability
Mathematical modeling
Parameter estimation
Mitogen-activated protein kinase pathway

ABSTRACT

Two primary purposes for mathematical modeling in cell biology are (1) simulation for making predictions of experimental outcomes and (2) parameter estimation for drawing inferences from experimental data about unobserved aspects of biological systems. While the former purpose has become common in the biological sciences, the latter is less common, particularly when studying cellular and subcellular phenomena such as signaling—the focus of the current study. Data are difficult to obtain at this level. Therefore, even models of only modest complexity can contain parameters for which the available data are insufficient for estimation. In the present study, we use a set of published cellular signaling models to address issues related to global parameter identifiability. That is, we address the following question: assuming known time courses for some model variables, which parameters is it theoretically impossible to estimate, even with continuous, noise-free data? Following an introduction to this problem and its relevance, we perform a full identifiability analysis on a set of cellular signaling models using DAISY (Differential Algebra for the Identifiability of SYstems). We use our analysis to bring to light important issues related to parameter identifiability in ordinary differential equation (ODE) models. We contend that this is, as of yet, an under-appreciated issue in biological modeling and, more particularly, cell biology.

1. Introduction to global identifiability

Mathematical models that purport to describe biological systems are often highly parameterized. The complexity of such systems warrants more modeling detail and this is ultimately reflected in model complexity. This tends to be the case, for example, with models of membrane electrodynamics such as those describing the sinoatrial node action potential and ventricular contraction (Dokos et al., 1996; Kurata et al., 2002; Luo and Rudy, 1994a, b). Other examples include metabolic (Lambeth and Kushmerick, 2002; Mulquiney and Kuchel, 1999) and signaling systems (Huang and Ferrell, 1996; Kholodenko et al., 1999; Orton et al., 2005; Sedaghat et al., 2002). Often, these models are used for simulation and hypothesis generation, but their parameters need to be determined and tested for reliability before the model can be used for predictive purposes. Determining the values (or ranges thereof) of unknown parameters is not a simple task. Generally, when the system can be controlled and probed in detail, parameter values can be obtained by independent means from simple experimental systems. However, the need is quickly increasing for modeling formalisms that are suitable for parameter estimation from partial observations, for example in a clinical setting. Most often, parameters of interest (e.g. insulin sensitivity in diabetic subjects or drug effectiveness and clearance rates in clinical trials) vary considerably between individuals or even within individuals (between-occasion variability). Such models can provide a powerful resource both for the pharmaceuticals industry (model-based biomarkers and surrogate endpoints for drug development Vicini et al., 2002; http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html) and for clinical applications in disease diagnosis and prognosis.

For quantitative model-based strategies to fulfill current hopes and expectations, feasibility of parameter estimation, the "inverse problem", plays a fundamental role. The topic is very broad, as it includes practical questions related to experimental design and data gathering, in addition to theoretical considerations about the structural properties of the model such as identifiability, controllability and complexity. These, as we will see later, are only partially related to specific experimental considerations and are best addressed in a theoretical context.

The focus of this paper is that of model complexity and sufficiency of experimental data for estimating parameters of interest. This is a central issue associated with the selection of models and the quantification of parameters. It is usually not possible to measure or probe the dynamics of every part of the system, particularly in biological systems and especially in human

* Corresponding author. Tel.: +1 206 616 9424; fax: +1 206 685 3300.
E-mail address: rtroper@u.washington.edu (R.T. Roper).
[1] Current address: Pfizer Global Research and Development, San Diego, CA 92121, United States.

0022-5193/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2010.02.029

subjects. In such cases, the information requirement of the mathematical model (i.e. the information necessary to specify the model structure or parameters) may exceed the information content of the available data. In other words, the model may be over parameterized, having too many degrees of freedom. Thus, there may be multiple or infinite sets of parameter values for which the model predictions or dynamics are precisely identical in the measured state variables.

A general definition of this problem, with reference to compartmental models, is described in Cobelli and DiStefano (1980) and Godfrey (1983), and early numerical considerations are described in Jacquez and Greif (1985). More recent investigations of model and parameter identifiability are often accomplished by means of numerical simulation and include, for example, mechanistically realistic biological models (Fink and Noble, 2009). There have been applications to signaling systems (Vilela et al., 2009) and drug action models (Evans et al., 2004). Others have emphasized the identifiability of parameter combinations (Feng and DiStefano, 1995), also in the context of compartmental models. Methodological suggestions on the various facets of the problem span computer and differential algebra (Audoly et al., 1998; Denis-Vidal and Joly-Blanchard, 2004; Ljung and Glad, 1994; Saccomani et al., 2003) and statistical methods (Chu and Hahn, 2009) have been proposed. This is by no means an exhaustive set of references, but it serves to attest to the interest the topic has received and is receiving in the data modeling community, in several areas of applications.

The issue of structural, or *a priori*, parameter identifiability of mathematical models is well-known and rigorous strategies exist for addressing this problem (Audoly et al., 1998, 2001; Chapman et al., 2003; Cobelli and DiStefano, 1980; Chappell and Godfrey, 1992; Walter and Lecourtier 1982; Ljung and Glad, 1994; Saccomani et al. 2003; Denis-Vidal and Joly-Blanchard, 2004). Nevertheless, there is no unambiguously appreciated and widely applied methodological solution as there is, for example, for the problem of parameter estimation (i.e. by nonlinear regression).

We will briefly describe a powerful approach previously developed to check identifiability of linear and nonlinear dynamic models described by differential equations involving polynomial or rational functions, based on differential algebra and symbolic computation (Audoly et al., 2001; Saccomani et al., 2003). We will also illustrate its application by using it to analyze, in a systematic way, the *a priori* identifiability properties of recently reported cell biology models, in addition to simpler compartmental models. In this work, we simply aim to bring to light and discuss general principles and concepts associated with *a priori* parameter identifiability. The intent, for now, is to introduce the topic in a way that is accessible to a wide audience. In doing so, we wish to emphasize its importance and relevance for the proper application of quantitative models of molecular biology.

1.1. Definition of a priori global identifiability

A priori global, or unique, parameter identifiability of mathematical models is a theoretical problem aimed at resolving the following question:

> Given (1) a set of model equations, (2) the known perturbing input(s) or forcing function(s) to the modeled system and (3) one or more measurement equations, what model parameters is it theoretically possible to globally (uniquely) estimate for an idealized, best-case situation where measurements are taken continuously without measurement error?

The significance of this theoretical question is that the answer tells us, before the experimentally measured time-course(s) are actually gathered, what parameters would be impossible to uniquely estimate, even with perfect, noise-free data acquired continuously. This is valuable information for two primary reasons:

(1) It can guide important modeling decisions such as what derived parameters might be defined to simplify the model structure (i.e. model reparameterization) or what parameters can or must the investigators assume values for (by means of independent biological information) to obtain unique estimates on the remaining parameters.
(2) It can provide information on portions of the modeled system for which it is necessary (but not automatically sufficient, depending on sampling rate and error in the data) to collect experimental data to uniquely estimate model parameters of interest. Note that the assumption of no measurement error is not a limitation of the approach. *A priori* identifiability is carried out in the (entirely theoretical or structural) best possible circumstances of noiseless, unlimited data. However, this is not a drawback, as an *a priori* not identifiable model cannot, under any circumstances – except changes in structure and parameter value assumptions – be identifiable *a posteriori*, from actual data. Thus, *a priori* identifiability is a necessary, but not sufficient condition for *a posteriori* identifiability, which depends on experimental design considerations

The first question is particularly relevant when experimental data have already been collected before assessing *a priori* identifiability, while the second is an experimental design question where the data are collected only after *a priori* identifiability has been considered.

1.2. Introduction of basic concepts using a two-compartment model

To introduce to the journal readership some of the basics of *a priori* identifiability, we will focus on a simple linear, two-compartmental model (see e.g. Carson et al., 1983; Godfrey, 1983 for a review of compartmental models) with (1) an irreversible loss from each compartment, (2) a perturbing input in compartment 1 and (3) measurements from compartment 1. This modeling construct is schematically illustrated in Fig. 1a. We will also consider a related model having a nonlinear loss from compartment 1. The corresponding identifiability results are interesting and bring to light some important points in relation to a *priori* identifiability.

We will not provide a mathematically rigorous treatment of the subject of *a priori* identifiability, since this has been done previously, as we have mentioned (Audoly et al., 1998, 2001; Cobelli and DiStefano, 1980; Walter and Lecourtier, 1982; Chappell and Godfrey, 1992; Ljung and Glad, 1994; Saccomani et al., 2003; Denis-Vidal and Joly-Blanchard, 2004). We present here only sufficient detail to practically illustrate the key concepts introduced in Section 1.1. The identifiability method we will discuss is based on differential algebra theory (Ritt, 1950) and has been described in some detail previously (Audoly et al., 2001; Saccomani et al., 2003). In particular, it has been recently implemented in a freely available software (Bellu et al., 2007) which does not require expertise on mathematical modeling by the medical/biological investigator.

The two-compartment model illustrated in Fig. 1a will be used as a case study to describe the general steps in determining *a priori* identifiability of a differential equation model. The equations for this two-compartment model are given in Eqs. (1). The state variables, $x_1(t)$ and $x_2(t)$, are the amounts in compartments 1