Mathematical Biosciences 260 (2015) 11–15

Contents lists available at ScienceDirect

Mathematical Biosciences

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

The inverse problem in mathematical biology

Gilles Clermont[a,*], Sven Zenker[b]

[a] Crisma Center, Departments of Critical Care Medicine, Mathematics, and Chemical Engineering, University of Pittsburgh, 200 Lothrop St, Pittsburgh, PA 16123, USA
[b] Department of Anesthesiology and Intensive Care Medicine, University of Bonn Medical Center, Sigmund-Freud-Str. 25, Bonn, 53105, Germany

ARTICLE INFO

Article history:
Received 11 August 2014
Accepted 3 September 2014
Available online 18 October 2014

Keywords:
Biological systems
Inverse problem
Parameter estimation
Ensemble models

ABSTRACT

Biological systems present particular challengers to model for the purposes of formulating predictions of generating biological insight. These systems are typically multi-scale, complex, and empirical observations are often sparse and subject to variability and uncertainty. This manuscript will review some of these specific challenges and introduce current methods used by modelers to construct meaningful solutions, in the context of preserving biological relevance. Opportunities to expand these methods are also discussed.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Biological systems at all levels of organizations are modeled for the purposes of gaining insight, testing hypotheses, or formulating predictions on future states of these systems following their normal evolution, or consequent to some outside perturbation or experiment. Depending on the intent of the modeler, models are formulated with varying levels of complexity. Simulation refers to applying a model in order to extract predictions. Depending on the nature of the model, these predictions are associated with varying levels of certainty, which may be quantifiable with appropriate methodology. This is the forward problem [1]. Before a model can be used for offering predictions, it must be appropriately parameterized. The activity of parameterizing a model from empirical observations, which may involve various tasks depending on the precise mathematical formulation of this model, constitutes the inverse problem [2,3]. Depending on the particular biological system and its precise model representation, solving the inverse problem ranges from manageable, to impractical, to impossible. Solutions to the inverse problems present significant challenges for all but the most simple of biological systems. While most of the effort in systems biology has focused on the interface of complex biological networks and high-throughput data [4,5], biological problems at other scales of description and triangulated by much sparser empirical datasets have received relatively little attention [1].

2. Posing the problem

2.1. Mathematical formulation

This manuscript will discuss solutions of biological problems that are mathematically represented by systems of difference or differential equations [6], or even more generally by sets of rules defining interactions such as present in rule-based or agent-based models [7,8]. With little loss of generality, let us assume that a biological system is modeled using a system of ordinary differential equations:

$$\dot{\underline{X}} = f(\underline{X}, \underline{p}, t) + \underline{U}(\underline{X}, t) + \underline{N}(t)$$
$$\underline{O} = g(\underline{X})$$

where $\dot{\underline{X}}$ is the first time derivative a vector of model variables, $\underline{O}$ is a vector of observables on the biological system being modeled, $\underline{U}(\underline{X}, t)$ a vector of external controls or influences on the system, $\underline{p}$ is a vector of parameters, $\underline{N}(t)$ are noise terms, the function f embodies the biology governing the interaction between the different variables on the system, and g is a map between system observables and model variables. For the model to be usable for prediction, additional data, typically in the form of a vector of initial conditions $\underline{X}_{t=0}$, turning the forward problem into an initial value problem, need to be specified. In addition, a number of observations are made on the system across several experimental conditions E which systematically or randomly vary initial conditions, boundary conditions, or manipulate the temporal evolution of the system in some other way. In its simplest form, solving the inverse problem therefore involves the definition of a function $\Psi(\underline{O}, \underline{p}, f, \underline{U})$ quantifying the error between model predictions and observations across all experiments, and identifying $\underline{p}$, but occasionally also f or $\underline{U}$ that will minimize this function, resulting in a model that best matches all available data across experiments.

* Corresponding author at: Scaife 602A, 3550 Terrace St, Pittsburgh, PA 16123, USA. Tel.: +1 412 647 9073.
E-mail addresses: clermontg@upmc.edu (G. Clermont), cler@pitt.edu (S. Zenker).

http://dx.doi.org/10.1016/j.mbs.2014.09.001
0025-5564/© 2014 Elsevier Inc. All rights reserved.

The careful construction of $\underline{\Psi}$ and how its various components are combined and weighted when multiple objectives are pursued is of paramount importance in biological systems, involving layers of considerations beyond identifying local minima of the error.

2.2. Sources of uncertainty

Any quantitative scientist collaborating with biological or social scientists is awed by the character of the empirical data derived from apparently perfectly reproducible experiments. Biological systems are complex and noisy. Noise may appear to govern a lot of the observed dynamics, either because of actual random dynamics of the data generating processes or because a significant portion of what drives the dynamical evolution of the system is either not represented or implicit in the model. In addition, there is generally an observability problem, that is, the knowledge obtained from the available observation insufficiently constrains the inverse problem of identifying states and/or parameters. Sources of variability in the data can be grouped under three major categories: (1) the experiment is difficult to reproduce because a number of known factors cannot be easily controlled, (2) the measurement is difficult to reproduce because instruments and tests have intrinsic variability, and (3) there is residual variability stemming from a large number of unknown and possibly uncontrollable factors, including the highly variable makeup of biological organisms, which can oftentimes turn out to be the dominant source of variability overall. There are also additional sources of variability that could be present in longitudinal data. Small animal experiments often require animal sacrifice at preset time points since the measurements cannot be performed while leaving the animal intact, so a given animal only provides data at a single time point. Many experiments in live animals may also result in death, caused by an experimental manipulation or not. Data from late time points are therefore populated by animals that survived to that time point, and data may therefore be biased in favor of animals naturally more resistant. A separate source of uncertainty resides in lack of a full understanding of the exact relationships between observables and model variables. In biological systems, an identity relationship is often assumed, thereby further simplifying biology. Data sparsity is also a common source of uncertainty for the modelers. It is therefore centrally important that experimental data be obtained over the dynamic range of the biological response. This problem is particularly pernicious when dealing with observational studies or clinical data. An example of this is related to efforts to model the molecular response to severe infections in humans: most humans with severe infections are first encountered well into the disease process, and at different, and unknown, times along that process [9].

Some of these shortcomings in data can be alleviated with careful experimental design or more sophisticated statistical treatment. Yet most cannot be addressed effectively, especially in situations where the experimental and modeling groups do not work in close and bilateral collaboration, particularly when data has already been collected. Any attempt at solving inverse problems must address these shortcomings in such a way as to map data uncertainty into model uncertainty with minimal bias.

3. A satisfactory solution to the inverse problem

A more precise formulation of the inverse problem in the context of biological systems may thus be more appropriately stated along the following lines: given the data at hand and its limitation, and given prior knowledge of the system one is trying to model, what is the least biased model representation of this system that can be offered. Whether this representation is useful in the broader sense remains to be investigated and will generally depend on the nature of the mechanistic insight or predictions being sought. A vast class of models offers a linear representation of biological systems. Given sufficient data, the optimization problems arising from the corresponding inverse problems are convex, that is admitting a single, global, optimal parameterization . One must realize that such representations are in fact a choice made by the modeler and that the actual system might only be approximately linear under very restricted biological operating conditions, a specific instance of the general fact that any mathematical representation of reality will involve simplifying assumptions. Such local linearization, performed explicitly or implicitly in the modeling process, is a useful tool as long as it is recognized that as soon as the system is perturbed away from the validity of this approximation, the model may lose any potential validity and thus usefulness, which will almost always be the case when studying biological systems under stress caused by environmental alterations, disease, etc.

More broadly, the modeler may also mandate or make the implicit assumption that, despite the existence of a large number of local optima, there exists among them a best one that nature has somehow adopted. An example of such a situation is instantiated in the protein folding problem where generally, the best local optimum (lowest energy) is promoted to global optimum within a range of conformations of interest and predicted to be the three-dimensional conformation of a sequence of amino acids. In this particular situation, folding is a process which is itself highly optimized requiring the presence of chaperon proteins at different steps of a growing primary sequence of amino-acids and thus the final product is “guided” to a local energy minimum. Therefore, approaching folding as a local optimization problem is sensible. Despite the knowledge of the existence of a large number of other minima, they are not favored by nature where function dictates structure. Such a conformation, although stable upon small perturbations representative of routine biological function, is not upon larger, biologically irrelevant perturbations such as heating beyond temperatures compatible with life.

Therefore, although methods for estimating convex or “almost” convex systems are plentiful, well understood and described and implemented in major software packages, they are of limited use and interest in estimating complex biological systems. Rather, we are interested in a class of problems where there are many, potentially infinitely many, solutions to the inverse problem, that is ill-posed problems, and where there is no clear biological indication as to which of those solutions are biologically implausible. Techniques have been developed to construct ensemble models, where each member of the ensemble is a parameter vector for a given a model structure [1,10–13]. Ensemble modeling, although computationally demanding, represents the most satisfactory solution to ill-posed inverse problems to date [14]. This suite of methods is undergoing active development in the context of complex biological systems and presents a number of open problems still to be addressed effectively (see below).

4. Parameter uncertainty and model identification

There exist good reviews on the topic of model identifiability [15]. Model identifiability is generally defined along an axis ranging from formal, or a priori global identifiability, to practical identifiability. A priori global identifiability or structural identifiability stipulates whether, given a model structure and a set of parameters and initial conditions, model parameters can always be uniquely identified assuming sufficient data. Several formal methods exist to resolve global and local structural identifiability, each having strength and weaknesses depending on the mathematical structure of the problem [16]. Generally, the more complex the system (e.g. presence of Hill or Michaelis–Menten kinetic terms) and the higher the number of parameters compared to observables, the more difficult it will be to resolve structural identifiability of a system. Formal methods based on differential algebra [17] (e.g.) have become popular recently. Software implementing differential algebra [18,19] (http://www.dei.unipd.it/~pia/) , generating series [20] , or a combination of approaches [21] have been developed to investigate a

Download English Version:

https://daneshyari.com/en/article/4500015

Download Persian Version:

https://daneshyari.com/article/4500015

Daneshyari.com