Mathematical Biosciences xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

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

Parameter uncertainty in biochemical models described by ordinary differential equations

J. Vanlier *, C.A. Tiemann, P.A.J. Hilbers, N.A.W. van Riel

Eindhoven University of Technology, Department of Biomedical Engineering, Eindhoven, The Netherlands
Netherlands Consortium for Systems Biology, University of Amsterdam, Amsterdam, 1098 XH, The Netherlands

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Inference
Ordinary differential equations
Dynamical systems
Parameter estimation
Uncertainty analysis
Bayesian

ABSTRACT

Improved mechanistic understanding of biochemical networks is one of the driving ambitions of Systems Biology. Computational modeling allows the integration of various sources of experimental data in order to put this conceptual understanding to the test in a quantitative manner. The aim of computational modeling is to obtain both predictive as well as explanatory models for complex phenomena, hereby providing useful approximations of reality with varying levels of detail. As the complexity required to describe different system increases, so does the need for determining how well such predictions can be made. Despite efforts to make tools for uncertainty analysis available to the field, these methods have not yet found widespread use in the field of Systems Biology. Additionally, the suitability of the different methods strongly depends on the problem and system under investigation. This review provides an introduction to some of the techniques available as well as gives an overview of the state-of-the-art methods for parameter uncertainty analysis.

1. Introduction

In the past decades, molecular biology has unraveled various pathways that play a role in biological phenomena. Many of the components and interactions involved have been identified by employing a reductionist approach, isolating parts of the system from the whole. Gradually, a picture of a complex network of various interacting subsystems is emerging. Investigation of the interactions between various components of a system is essential for understanding its emergent behavior. Consequently, there has been a shift in paradigm from reductionism to integration [1–4]. By formalizing hypotheses on how a pathway operates in the form of computational models our conceptual understanding can be put to the test quantitatively [5–7]. By repeatedly challenging such models with new data, we can iteratively obtain models that are decreasingly wrong. The aim of Systems Biology is to obtain both predictive as well as explanatory models for complex phenomena, hereby providing useful approximations of reality with varying levels of detail. In this review we focus mostly on dynamical models which consist of state variables (or states) which are quantities that change over time, and parameters (fixed with respect to time). These states are embedded in a system of equations which relate the different states of the model.

To simulate the model and make predictions, parameter values are required. These typically have to be estimated from data. Due to the fact that data is measured with finite accuracy and only a subset of the state variables is accessible experimentally, uncertainty analysis is a highly relevant topic. Despite efforts to make tools for uncertainty analysis available [8–11] these methods have not yet found widespread use in the field. Nevertheless, some examples of successful applications in computational biology are listed in Table 1. We provide an introduction to some of the techniques available as well as give an overview of the state-of-the-art methods for parameter uncertainty analysis [12–17]. Which methods are applicable to a specific problem strongly depends on the system under investigation and the assumptions one is willing to make. We hope to provide insight in the suitability of different techniques for addressing uncertainty in computational modeling of dynamical systems.

2. Computational modeling of biochemical systems

The use of dynamical models has a long history within several disciplines of science. In the realm of classical physics, models often comprise of physical laws with well established and invariant physical constants acting as parameters. In engineering, parameters are often application specific, but the individual components are usually well characterized and their interactions known. In computational biology, the challenges faced are different. Though methods for discovering interactions are well established [32–34], techniques for accurately determining biochemical parameters remain limited [12]. Additionally relying on measurements that were performed *in vitro* can lead to inaccuracies [35]

* Corresponding author at: Eindhoven University of Technology, Department of Biomedical Engineering, Eindhoven, The Netherlands. Tel.: +31 402475573.
E-mail address: j.vanlier@tue.nl (J. Vanlier).

0025-5564/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2013.03.006

Table 1
A list of the methods discussed in this paper and some relevant applications in the field of Computational Biology.

Name	Requirements	Result	Applied papers
Sensitivity analysis	None	Assessment of (local) sensitivity	[18–21]
Profile Likelihood	Likelihood function	Assessment of identifiability	[7,22]
Bootstrap	None	Distribution of parameters based on simulated replicates	[15,23,24]
Markov Chain Monte Carlo	Weak identifiability	Posterior distribution	[25–28]
Sequential Monte Carlo	Proper priors for all parameters	(Approximate) posterior distribution	[29–31]

due to differences in the biochemical environment and regulatory mechanisms that were not accounted for [16]. Despite the wealth of information that kinetic assays provide, such issues require attention and warrant future research. Since such measurements are both expensive as well as time consuming, the amount of data is often relatively scarce. Moreover, considering the complexity of models required to describe biological pathways, this leads to large uncertainties in the inferred values of these biochemical parameters. Consequently, this results in several poorly constrained and therefore uncertain model predictions [10,24,36–38]. To deal with such uncertainties and to ascertain their implications on scientific conclusions, several methods have been developed. Some of these are useful for probing model properties, others for designing informative experiments.

3. Parameter estimation

The scope of this review is restricted to dynamical systems that can be described by ordinary differential equations (ODEs) (1) [39–42]. Such a description is appropriate when the number of particles involved in the biochemical network is large enough to be able to consider concentrations and when spatial effects are negligible [43–48]. Typically, such models take the following form:

$$\dot{\vec{x}}(t) = f(\vec{x}(t), \vec{u}(t), \vec{p}) \tag{1}$$

$$\vec{y}(t) = g(\vec{x}(t), \vec{q}) + \vec{\xi}(t) \tag{2}$$

$$\vec{x}(0) = h(\vec{r}) \tag{3}$$

These models are described by equations which contain parameters $\vec{p}$ (which are fixed constants with respect to time) and inputs $\vec{u}(t)$ which depend on time and state variables $\vec{x}(t)$. Given a set of parameters, inputs and initial conditions $\vec{x}(0)$, these equations can subsequently be simulated, and time courses of the model state variables can be obtained. Such systems are typically partially observed, which means that measurements $\vec{y}(t)$ are performed on a subset or a combination of the total number of states N in the model. Additionally, these measurements are hampered by measurement noise $\vec{\xi}$. Furthermore, many of the employed measurement techniques necessitate the use of scaling and offset parameters $\vec{q}$ [49]. For ease of notation, $\vec{\theta}$ is defined as $\vec{\theta} = \{\vec{p}, \vec{q}, \vec{r}\}$, which lists all the parameters that should be defined in order to simulate the model.

After the model is postulated and data is acquired, parameter estimation can be performed. To do this, consider the probability density of observing data $\mathbf{y}$ given parameter values $\vec{\theta}$. For the sake of clarity of exposition, we assume independent additive Gaussian noise with constant variance for each measurement, which results in a probability density function defined as (4). Here p refers to a probability density.

$$p(\mathbf{y}|\vec{\theta}) = \prod_{i=1}^{M}\prod_{j=1}^{N_i} p(y_i(t_j), \vec{\theta}) \tag{4}$$

$$= \frac{1}{\prod_{i=1}^{M}\left(\sqrt{2\pi}\sigma_i\right)^{N_i}} e^{-\sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(\frac{y_{i,j}-y_i(t_j,\vec{\theta})}{\sqrt{2}\sigma_i}\right)^2} \tag{5}$$

It is at this point where the so-called frequentist and Bayesian methodologies begin to diverge. The former opts for a purely data-based approach, aiming to find all the parameter sets which describe the data to an acceptable degree $p(\mathbf{y}^D|\vec{\theta}) > p_{lim}$. Here the threshold p_{lim} is determined by choosing a significance level and computing the associated critical values of the uncertainty distribution of the data. In a Bayesian approach, inference is performed probabilistically, as parameters are treated as random variables and Bayes' theorem (6) is applied to incorporate prior knowledge on these parameter values (7) [25,50]. Priors are specified in the form of probability densities and are defined with respect to the parameters. They usually represent either current belief [51] or attempt to be non-informative. The latter is usually accomplished either by choosing wide priors (such as gamma priors for parameters with only positive support [38]) or priors that exhibit invariance to parameter transformation [52]. In brief, the aim in Bayesian inference is to sample from the posterior parameter probability density and determine bounds enclosing $(100-\alpha)\%$ of the probability density.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{6}$$

$$P(\vec{\theta}|\mathbf{y}^D) = \frac{p(\mathbf{y}^D|\vec{\theta})p(\vec{\theta})}{p(\mathbf{y}^D)} = \frac{p(\mathbf{y}^D|\vec{\theta})p(\vec{\theta})}{\int_\Omega p(\mathbf{y}^D|\vec{\theta})p(\vec{\theta})d\vec{\theta}} \tag{7}$$

Here the probability density of the parameter values is given by $p(\vec{\theta}|\mathbf{y}^D)$ which can be computed from the probability density $p(\mathbf{y}^D|\vec{\theta})$ of the data given parameters $\vec{\theta}$, the prior probability density of the parameters $p(\vec{\theta})$ and the marginal likelihood or model evidence $p(\mathbf{y}^D)$. Since the marginal likelihood does not depend on the parameters, this merely acts as a normalizing constant.

In the frequentist paradigm, one usually proceeds by maximizing the likelihood function $L(\vec{\theta})$ (Maximum Likelihood Estimation or MLE) whose right hand side is the same as (4). Since the logarithm does not change the location of the optimum with respect to the parameters, one often minimizes the negative log-likelihood in practice as it allows for efficient optimization due to its quadratic nature. When the data variances are known, the normalization constant is independent of the parameters and minimizing the Residual Sum of Squares (RSS) becomes equivalent to maximizing the likelihood (8).

$$\chi^2 = \sum_{i=1}^{M}\sum_{j=1}^{N_i}\left(\frac{y_{i,j}^D - y_i(t_j,\vec{\theta})}{\sigma_i}\right)^2 \tag{8}$$

When the variances σ_i have to be estimated from the data however, they should preferably be treated as additional parameters. Since σ_i also appears in the normalizing constant, such an approach also leads to an additional term in the χ^2 function namely $\sum_{i=1}^{M} N_i log(2\pi\sigma_i^2)$. Additionally, one can incorporate a prior distribution on the parameters and perform estimation based on the combined quantity. Maximizing this quantity results in the Maximum A Posteriori (MAP) estimator, which can also be interpreted as a regularized form of MLE.

Determining which parameters actually maximize these quantities is a non-trivial problem due to the existence of locally

Please cite this article in press as: J. Vanlier et al., Parameter uncertainty in biochemical models described by ordinary differential equations, Math. Biosci. (2013), http://dx.doi.org/10.1016/j.mbs.2013.03.006