Journal of Theoretical Biology 254 (2008) 178–196

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Review

A methodology for performing global uncertainty and sensitivity analysis in systems biology

Simeone Marino, Ian B. Hogue, Christian J. Ray, Denise E. Kirschner *

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-0620, USA

ARTICLE INFO

Article history:
Received 6 December 2007
Received in revised form
14 March 2008
Accepted 12 April 2008
Available online 20 April 2008

Keywords:
Latin hypercube sampling (LHS)
Partial rank correlation coefficient (PRCC)
Extended Fourier amplitude sensitivity test (eFAST)
Agent-based model (ABM)
Sensitivity index
Monte Carlo methods
Aleatory uncertainty
Epistemic uncertainty

ABSTRACT

Accuracy of results from mathematical and computer models of biological systems is often complicated by the presence of uncertainties in experimental data that are used to estimate parameter values. Current mathematical modeling approaches typically use either single-parameter or local sensitivity analyses. However, these methods do not accurately assess uncertainty and sensitivity in the system as, by default, they hold all other parameters fixed at baseline values. Using techniques described within we demonstrate how a multi-dimensional parameter space can be studied globally so all uncertainties can be identified. Further, uncertainty and sensitivity analysis techniques can help to identify and ultimately control uncertainties. In this work we develop methods for applying existing analytical tools to perform analyses on a variety of mathematical and computer models. We compare two specific types of global sensitivity analysis indexes that have proven to be among the most robust and efficient. Through familiar and new examples of mathematical and computer models, we provide a complete methodology for performing these analyses, in both deterministic and stochastic settings, and propose novel techniques to handle problems encountered during these types of analyses.

Contents

* Corresponding author. Tel.: +1 734 647 7722; fax: +1 734 647 7723.
E-mail address: kirschne@umich.edu (D.E. Kirschner).

0022-5193/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2008.04.011

1. Introduction

Systems biology is the study of the interactions between the components of a biological system, and how these interactions give rise to the function and behavior of the system as a whole. The systems biology approach often involves the development of mathematical or computer models, based on reconstruction of a dynamic biological system from the quantitative properties of its elementary building blocks. Building mathematical and computational models is necessary to help decipher the massive amount of data experimentalists are uncovering today. The goal of the systems biologist or modeler is to represent, abstract, and ultimately understand the biological world using these mathematical and computational tools. Experimental data that are available for each system should guide, support, and shape the model building process. This can be a daunting task, especially when the components of a system form a very complex and intricate network.

Paraphrasing Albert Einstein, models should be as simple as possible, but not simpler. A parsimonious approach must be followed. Otherwise, if every mechanism and interaction is included, the resulting mathematical model will be comprised of a large number of variables, parameters, and constraints, most of them uncertain because they are difficult to measure experimentally, or are even completely unknown in many cases. Even when a parsimonious approach is followed during model building, available knowledge of phenomena is often incomplete, and experimental measures are lacking, ambiguous, or contradictory. So the question of how to address uncertainties naturally arises as part of the process. Uncertainty and sensitivity (US) analysis techniques help to assess and control these uncertainties.

Uncertainty analysis (UA) is performed to investigate the uncertainty in the model output that is generated from uncertainty in parameter inputs. Sensitivity analysis (SA) naturally follows UA as it assesses how variations in model outputs can be apportioned, qualitatively or quantitatively, to different input sources (Saltelli et al., 2000). In this work we review US analysis techniques in the context of deterministic dynamical models in biology, and propose a novel procedure to deal with a particular stochastic, discrete type of dynamical model (i.e. an agent-based model—ABM[1]).

By deterministic model, we mean that the output of the model is completely determined by the input parameters and structure of the model. The same input will produce the same output if the model were simulated multiple times. Therefore, the only uncertainty affecting the output is generated by input variation. This type of uncertainty is termed *epistemic* (or subjective, reducible, type B uncertainty; see Helton et al., 2006). Epistemic uncertainty derives from a lack of knowledge about the adequate value for a parameter/input/quantity that is assumed to be constant throughout model analysis. In contrast, a stochastic model will not produce the same output when repeated with the same inputs because of inherent randomness in the behavior of the system. This type of uncertainty is termed *aleatory* (or stochastic, irreducible, type A; see Helton et al., 2006). This distinction has been and still is an area of interest and study in the engineering and risk assessment community (see Apostolakis, 1990; Helton, 1997; Helton et al., 2007; Parry and Winter, 1981; Pate'-Cornell, 1996).

Many techniques have been developed to address US analysis: differential analysis, response surface methodology, Monte Carlo (MC) analysis, and variance decomposition methods. See Iman and Helton, (1988) and Saltelli et al. (2000) for details on each of these approaches and Cacuci and Ionescu-Bujor (2004), Draper (1995), Helton (1993) and Saltelli et al. (2005) for more general reviews on US analysis. Here we briefly illustrate the most popular, reliable, and efficient UA techniques and SA indexes. In Section 2, we describe two UA techniques: a MC approach and Latin hypercube sampling (LHS). In Section 3, we describe two SA indexes: partial rank correlation coefficient (PRCC) and extended Fourier amplitude sensitivity test (eFAST): PRCC is a sampling-based method, while eFAST is a variance-based method. In Section 4, we perform US analysis on both new and familiar deterministic dynamical models (quantifying epistemic uncertainty) from epidemiology and immunology, and discuss results. Section 5 presents an ABM, where we suggest a method to deal with the aleatory uncertainty that results from the stochasticity embedded in the model structure, to facilitate the use of PRCC and eFAST techniques. We use Matlab (Copyright 1984–2006 The MathWorks, Inc., Version 7.3.0.298 R2006b) to solve all the differential equation systems of Section 4 and to implement most of the US analysis functions described throughout the manuscript (available on our website, http://malthus.micro.med.umich.edu/lab/usanalysis.html).

2. Uncertainty analysis

Input factors for most mathematical models consist of parameters and initial conditions for independent and dependent model variables. As mentioned, these are not always known with a sufficient degree of certainty because of natural variation, error in measurements, or simply a lack of current techniques to measure them. The purpose of UA is to quantify the degree of confidence in the existing experimental data and parameter estimates. In this section we describe the most popular sampling-based approaches used to perform UA, MC methods, and their most efficient implementation, namely the LHS technique.

2.1. Monte Carlo simulation

MC methods are popular algorithms for solving various kinds of computational problems. They include any technique of statistical sampling employed to approximate solutions to quantitative problems. A MC simulation is based on performing multiple model evaluations using random or pseudo-random numbers to sample from probability distributions of model inputs. The results of these evaluations can be used to both determine the uncertainty in model output and perform SA. A large body of literature exists on the use of expert review processes to characterize epistemic uncertainty associated with poorly known model parameters (see for example Cooke, 1991; Evans et al., 1994; Hora and Iman, 1989; McKay and Meyer, 2000).

[1] IBM = Individual Based Modeling in fields like ecology.