Automatica 48 (2012) 2740–2749

Contents lists available at SciVerse ScienceDirect

Automatica

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

# Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models☆

Simona Dobre [a,1], Thierry Bastogne [a,b], Christophe Profeta [c], Muriel Barberi-Heyob [d], Alain Richard [a]

[a] Centre de recherche en Automatique de Nancy (CRAN), Nancy-Université, CNRS UMR 7039, Campus Science, BP 70239, 54506, Vandœuvre-lès-Nancy Cedex, France
[b] BIGS (INRIA Nancy Grand-Est), France
[c] Institut Elie Cartan, Nancy-Université, Campus Science, BP 70239, 54506, Vandœuvre-lès-Nancy Cedex, France
[d] Centre de recherche en Automatique de Nancy (CRAN), Nancy-Université, CNRS UMR 7039, Centre Alexis Vautrin, 54511, Vandœuvre-lès-Nancy Cedex, France

ARTICLE INFO

*Article history:*
Received 16 October 2010
Received in revised form
15 October 2011
Accepted 5 December 2011
Available online 30 August 2012

*Keywords:*
Identifiability
Sensitivity analysis
Nonlinear systems
Dynamical systems
Systems biology

ABSTRACT

In systems biology, a common approach to model biological processes is to use large systems of nonlinear differential equations. The associated parameter estimation problem then requires a prior handling of the global identifiability question in a realistic experimental framework. The lack of a method able to solve this issue has indirectly encouraged the use of global sensitivity analysis to select the subset of parameters to estimate. Nevertheless, the links between these two global analyses are not yet fully explored.

The present work reveals new bridges between sensitivity analyses and global non-identifiability, through the use of functions derived from the Sobol' high dimensional representation of the model output. We particularly specify limits of variance-based sensitivity tools to completely conclude on global non-identifiability of parameters in a given experimental context.



## 1. Introduction

In systems biology, the inference of biological networks from quantitative properties of their elementary constituents is a major area of research (Cassman et al., 2007; Kitano, 2001). This raises particular challenges such as the identification of high-dimensional nonlinear dynamical systems and more precisely the analysis of their parameter identifiability (Mirsky et al., 2009; Raue et al., 2009; Stelling & Gilles, 2001).

Different classifications of parameter identifiability definitions exist. We refer herein to three classes (see Fig. 1): *a priori* identifiability, *a posteriori* identifiability and practical identifiability. The first class, known also as the theoretical or structural identifiability of model parameters, examines the question of existence and uniqueness of a solution to the parameter estimation problem (Walter & Pronzato, 1997), in an idealized framework where (1) the system and the model have an identical structure (no characterization error); (2) the data are noise-free, and (3) the input signals and measurement times can be chosen at will. However, this is only a necessary condition which cannot guarantee successful parameter estimation from real data. The second class, namely the *a posteriori* identifiability only considers the first two working assumptions and is a particular case of the output distinguishability (Grewal & Glover, 1976) for a finite collection of noise-free observations and a given input signal. The last class, practical identifiability, only relies on the first hypothesis and accounts for the noise factor but is generally established for a given estimation criterion (Dochain, Vanrolleghem, & Van Daele, 1995; Vanrolleghem, Van Daele, & Dochain, 1995). For that reason, this class of identifiability is often linked to the theory of optimization in mathematics.

Sensitivity analysis of the model output with respect to changes in model parameters is another technique widely used in system modeling to discriminate influential and non influential parameters (Saltelli et al., 2008; Streif, Findeisen, Waldherr, & Allgöwer, 2009). Dynamic sensitivity analysis has already been applied to biological networks for various purposes such as experimental design (Schlosser, 1994), parameter estimation (Miller & Frenklach, 2004) or the analysis of oscillatory systems (Rand, 2008; Zak, Stelling, & Doyle, 2005).

Several investigations on the connections between dynamic sensitivity and parameter identifiability analyses have been carried

☆ The material in this paper was partially presented at the 11th IFAC Symposium on Computer Applications in Biotechnology (CAB-2010), July 7–9, 2010, Leuven, Belgium. This paper was recommended for publication in revised form by Associate Editor Michael A. Henson under the direction of Editor Frank Allgöwer.

*E-mail addresses:* simona.dobre@cran.uhp-nancy.fr (S. Dobre), thierry.bastogne@cran.uhp-nancy.fr (T. Bastogne), christophe.profeta@iecn.u-nancy.fr (C. Profeta), m.barberi@nancy.fnclcc.fr (M. Barberi-Heyob), Alain.Richard@cran.uhp-nancy.fr (A. Richard).

[1] Tel.: +33 6 08 05 53 48; fax: +33 3 83 68 44 62.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.05.004

out (Brun, Kühni, Siegrist, Gujer, & Reichert, 2002; Stelling & Gilles, 2001; Yue et al., 2006), but the latter were only focused on local analysis.

In genomics, proteomics or metabolomics, biological parameters may vary widely within different ranges. As a consequence, global *a posteriori* identifiability needs to be addressed. Unfortunately, there is no technique able to assess the global identifiability condition in a given experimental context, i.e. when the input signal and the sampling conditions are imposed by the experimental context. Consequently, authors generally prefer to apply global sensitivity analysis techniques without solid justifications related to identifiability. Indeed, while the relationship between local sensitivity and identifiability analysis, through the Fisher information matrix, is clearly established (Brun, Reichert, & Künsch, 2001; Yao, Shaw, Kou, McAuley, & Bacon, 2003; Zak, Gonye, Schwaber, & Doyle, 2003), the link between global studies is less obvious. As a matter of fact, only insensitive parameters are generally considered as being non-identifiable. This is not surprising since global sensitivity measures usually serve as model reduction principles (before parameter estimation) or in tandem with uncertainty analysis for model robustness analysis (Saltelli et al., 2008). However, sensitive parameters could also be non-identifiable.

The objective of this paper is to present new results on the connections between global *a posteriori* identifiability and global dynamic sensitivity analysis. This study is structured around the Sobol' decomposition[2] method (Sobol', 2001). Specific functions, entitled $\Psi$ and $\Omega$-functions, derived from the Sobol' high dimensional model representation, are introduced. Their linear time-dependence and injectivity are examined, and their consequences on the non-identifiability of parameters are discussed. We show that variance-based sensitivity analysis can be used to test only one out of three causes of non-identifiability in a given experimental context, where the input signal and the measurement times are imposed. We also point out that the conclusions on parameter non-identifiability, in the case of colinear sensitivity measures, must be treated with caution.

This paper is structured as follows: *a priori* and *a posteriori* identifiabilities are first defined. The global sensitivity analysis based on the Sobol' high dimensional model representation is then briefly introduced in Section 3. Finally, the main contributions of this study are presented in Section 4 in which the links between sensitivity and global identifiability analyses are decomposed, in a theoretical framework. For simplicity reasons, the nomenclature used throughout this paper is detailed in Table 1.

## 2. Identifiability analysis

Let us consider a dynamic system described by a nonlinear state-space model defined as follows:

$$\begin{cases} \dfrac{d}{dt}\mathbf{x}(t) = f(\mathbf{x}(t), \mathbf{u}(t), t, \mathbf{p}); \mathbf{x}(0) = \mathbf{x}_0(\mathbf{p}) \\ \mathbf{y}(t, \mathbf{p}) = h(\mathbf{x}(t), \mathbf{p}) \end{cases} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$, $\mathbf{u} \in \mathbb{R}^{n_u}$ and $\mathbf{y} \in \mathbb{R}^{n_y}$ denote the state, input and output vectors respectively. The variable $\mathbf{x}_0$ is the initial value of the state vector, $\mathbf{p} \in \mathbb{R}^n$ is the vector of model parameters and $t$ is the time variable. $f(\cdot)$ and $h(\cdot)$ contain the state and output equations respectively.

In *a priori* identifiability, the solution uniqueness of the parameter estimation problem is assessed in an idealized framework (Walter & Pronzato, 1997). However, in experimental biology, the input design is often subject to economical and/or technical constraints and the number of observations is frequently limited to a few data points collected at time instants $\{t_k\} \in \mathbb{T}$, $k = 0, \ldots, N-1$. In such restrictive experimental frameworks, even if a parameter is *a priori* identifiable, it may not be so in practice, due to a lack of information in the available observations. The *a posteriori* identifiability condition can be stated as follows: given a parametric model structure with given input signals $\mathbf{u}$ and initial conditions $\mathbf{x}_0$, a parameter $p_i$, with $i \in 1, \ldots, n$ is *a posteriori* identifiable, if for almost all $\mathbf{p}^* \in \mathbb{P} \subset \mathbb{R}^n$ with $n$ the number of model parameters, the following condition is satisfied

$$\mathbf{y}(t_k, \mathbf{p}) = \mathbf{y}(t_k, \mathbf{p}^*) \quad \forall t_k \in \mathbb{T} \Longrightarrow p_i = p_i^*. \tag{2}$$

The local *a posteriori* identifiability, corresponding to $\mathbf{p}^* \in \mathbb{V}(\mathbf{p})$ where $\mathbb{V}(\mathbf{p})$ denotes the neighborhood of $\mathbf{p}$, is not considered here since it is not relevant to biological models, for which parameters may vary over wide ranges. Several methods exist to analyze global *a priori* identifiability, based on state isomorphisms (Peeters & Hanzon, 2005), differential algebra (Audoly, Bellu, D'Angio, Saccomani, & Cobelli, 2001; Ljung & Glad, 1994; Saccomani, 2004; Saccomani, Audoly, & D'Angio, 2003) or power series expansions (Pohjanpalo, 1978; Walter & Pronzato, 1997). Unfortunately, there is no available approach to assess *a posteriori* global identifiability. This lack of a practical solution has encouraged researchers in systems biology to use techniques of global sensitivity analysis (Kontoravdi, Asprey, Pistikopoulos, & Mantalaris, 2005; Peters Norton, 2009).

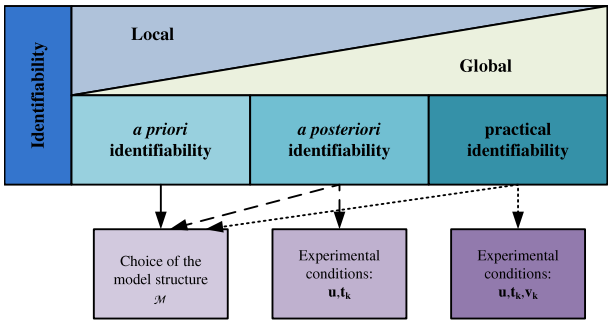


**Fig. 1.** Classification of identifiability definitions, where $\mathcal{M}$ denotes the model structure, $\mathbf{u}$ the input signals, $t_k$ the time measurements and $v_k$ the output noise.

## 3. Global sensitivity analysis

Several categories of sensitivity analysis methods already exist in previously published studies (Saltelli et al., 2008). Herein, we only focus on variance-based global methods and more precisely on the Sobol' sensitivity method (Sobol', 2001). This method allows the computation of the output's sensitivity with respect to the variation of model parameters over the entire parametric domain. The sensitivity measure related to a certain parameter is evaluated while varying all other parameters as well, revealing thus any existing interaction. Hereafter, we apply this method to dynamic systems, adjusting the terminology from *sensitivity indices* for static systems to *sensitivity functions* for dynamic ones.

*Hypotheses of the method*:

- ($\mathcal{H}_1$) the $n$ parameters are considered as i.i.d. random variables uniformly distributed over the $n$-dimensional unit cube $\mathbf{I}^n = [0, 1]^n$, i.e., $p_i \sim \mathcal{U}[0, 1], \forall i \in \{1, \ldots, n\}$;
- ($\mathcal{H}_2$) $y(t, \mathbf{p})$ is continuously differentiable and square integrable;
- ($\mathcal{H}_3$) when analyzing the global sensitivity analysis w.r.t. parameters, all other computational factors which could affect the model output, such as the simulation method, the sampling time, the input signal, etc., are not modified during the analysis.

[2] Decomposition known as the Hoeffding decomposition (Hoeffding, 1948), HDMR (high dimensional model representation) expansion (Rabitz & Aliş, 1999) or more recently, as the Sobol' decomposition (Sobol', 1993).