Global sensitivity and identifiability implications in systems biology

Simona Dobre, Thierry Bastogne, Alain Richard

Global sensitivity and identifiability implications in systems biology

Simona Dobre * **Thierry Bastogne** * **Alain Richard** *

* *Centre de recherche en Automatique de Nancy (CRAN), Nancy-Universite, CNRS UMR 7039, Campus Science, BP 70239, 54506, Vandoeuvre-les-Nancy Cedex, France*

Abstract: In systems biology, a common approach to model biological processes is to use large systems of differential equations.The associated parameter estimation problem requires to prior handle identifiability and sensitivity issues in a practical biological framework. The lack of method to assess global practical identifiability has leaded us to analyze and establish bridges between global sensitivity and identifiability measures. Specifically, we are interested in deriving conditions of global practical non-identifiability in association with global sensitivity results. Two cases are considered: i) insensitive (or non-observable) parameters ; ii) two (or more) correlated sensitivity measures of the model output with respect to model parameters. Propositions of relationships between sensitivity and identifiability, and their proofs are developped herein. Academic examples are also treated in order to illustrate contents of these propositions.

Keywords: identifiability, sensitivity analysis, nonlinear systems, systems biology.

1. INTRODUCTION

Systems biology involves the combination of two objectives: the comprehensively information gather from each of the distinct levels of individual biological systems and the integration of these data in order to generate predictive mathematical models of the system. These mathematical models are often described by nonlinear differential equations [1] and a central problem is to test the theoretical and practical identifiabilities of the biological parameters Dochain et al. (1995).

Theoretical and structural identifiability of model parameters examine the question of existence and uniqueness of solution to the parameter estimation problem. While the theoretical identifiability is studied in an idealized framework - where the system and the model have identical structures, the data are noise-free and where the input signals and measurement times are chosen at will - the practical identifiability takes quality of experimental data into account [Vanrolleghem et al. (1995)]. These two properties are also called *a priori* and *a posteriori* identifiability [Ghidaoui and Prasad (2000)]. Sensitivity analysis of the model output with respect to changes in model parameters is another tool used in system modeling to discriminate influent and non influent parameters [Saltelli et al. (2008)].

As an indicator of the *status quo* of the sensitivity and identifiability practices used in systems biology in the modeling community, we have classified in Fig. 1 all papers of the FOSBE [2] 2007 Proceedings that investigated

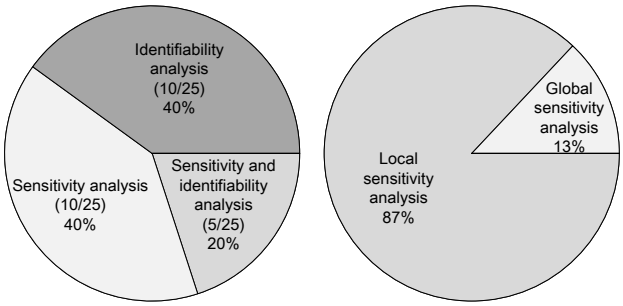

Fig. 1. In FOSBE'2007 about 29% (25/86) of papers dealt with sensitivity or identifiability. The latter can be decomposed as a) Sensitivity versus identifiability analysis; b) Local and global sensitivity

identifiability and/or sensitivity analysis. 25 out of 86 papers treated identifiability and/or sensitivity parameter analysis. It appears that the need of sensitivity analysis to validate model inference is universally acknowledged. The Monte Carlo approach allowing for the simultaneous propagation of the entire parameter distribution is often used for uncertainty analysis purposes (5/15), while for sensitivity analysis, the methods applied are mainly the local derivatives or the one-at-a-time approach (8/15) and only rarely (2/15) global methods. The link between identifiability and sensitivity was encountered in five papers, and only from a local point of view. One main application area in systems biology is the reconstruction of biological networks, at different levels: genetic, proteomic or metabolic scale.

If the equivalence between local sensitivity and identifiability analysis seems to be generally acknowledged, the link between the global studies is less obvious. As a matter of

[1] In Ideker et al. (2001) a sampling of systems biology approaches is presented and it can be noted that differential equations hold an important place in model analysis and simulation.

[2] FOSBE - Foundations of Systems Biology in Engineering (www.fosbe.org)

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)]. Nevertheless, sensitive parameters could also be non-identifiable.

The parameter non-identifiability can arise from structural properties of the model (theoretical identifiability study) or from extreme experimental constraints on the input signals and measurement time instants (practical identifiability study). Methods exist to test the global theoretical identifiability (like, for example, the Taylor series approach). However, in our knowledge no such global practical method exist. Our goal herein is to gain more insight on global practical identifiability by the means of global sensitivity analysis. Thus, the case of linear dependence of global sensitivity measures is considered, and its consequence on the non-identifiability is discussed.

This paper is structured as follows: theoretical and practical identifiabilities are firstly tackled and the Taylor series approach is presented. Section 3 is devoted to the global sensitivity analysis. Links between the sensitivity and identifiability global analyses are presented on section 4, focusing on the implication of colinear sensitivity measures and the non-identifiability of model parameters. Finally, results of global sensitivity and identifiability analysis for two simple examples are presented.

2. IDENTIFIABILITY ANALYSIS

Let us consider a nonlinear state-space model defined as follows:

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

where $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{y}$ denote the state, input and output vectors respectively. $\mathbf{x}_0$ is the initial value of the state vector and t is the time variable. $f\left(\cdot\right)$ and $h\left(\cdot\right)$ contain the state and output equations respectively.

Structural, theoretical or *a priori* identifiability [Walter and Pronzato (1997)] deals with the possibility to give a unique value to each parameter of a mathematical model structure. The uniqueness of this solution is assessed in an idealized framework where the process and the model have identical structures, the data are noise-free, and where the input signals and the measurement times can be chosen at will. However, in practice, experimental conditions are often subject to economical and/or technical constraints which can sometimes prevent input design from being applied to the process. Moreover, the number of observations is often limited to a few data points collected at time instants $\{t_j\}, j = 0, \ldots, N-1$. These are two crucial points in systems biology. In such a case, even if a parameter is globally or locally structurally identifiable, it may not be so in practice, due to a lack of information in the available observations. For that reason, D. Dochain and P. Vanrolleghem, in [Vanrolleghem et al. (1995); Dochain et al. (1995)], have introduced the notion of practical or *a posteriori* identifiability, which also includes the quality of the data. If we exclude the noise factor from the study, the practical identifiability is a particular case of the output distinguishability [Grewal and Glover (1976)] for a finite collection of observations $\{t_j\}$ and a given experiment $(\mathbf{x}_0, \mathbf{u})$. Then the practical 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 is practically identifiable, if for almost all $\mathbf{p}^* \in \mathbb{P}$,

$$\mathbf{y}\left(t_j, \mathbf{p}, \mathbf{x}_0, \mathbf{u}\right) = \mathbf{y}\left(t_j, \mathbf{p}^*, \mathbf{x}_0, \mathbf{u}\right) \implies p_i = p_i^*, \tag{2}$$

$\forall t_j \in \mathbb{T}$ and $\forall \mathbf{p} \in \mathbb{P}$ (generic identifiability).

To our knowledge, no such global identifiability method exists, and only local conditions can be tested. The local practical identifiability, corresponding to $\mathbf{p}^* \in \mathbb{V}(\mathbf{p})$ where $\mathbb{V}(\mathbf{p})$ denotes a neighbourhood of $\mathbf{p}$, is not considered herein since not relevant to biological models for which the initial guesses of parameters are either rarely available or largely uncertain.

In terms of global theoretical identifiability, several methods exist, based on state isomorphism's [Peeters and Hanzon (2005)], differential algebra [Ljung and Glad (1994); Audoly et al. (2001); Saccomani et al. (2003); Saccomani (2004)] or power series expansion [Pohjanpalo (1978); Walter and Pronzato (1997)]. This last class of methods supposes two types of expansions of the model output (as a function of input and time): Taylor series and generating series. Hereafter, we will apply this Taylor series approach.

2.1 Taylor series approach

In the case of Taylor series, the output vector and its time derivatives are typically developed around the initial time. Successive time derivatives, starting with the zeroth order term and going up to the n_pth order, (where n_p is the number of parameters) are used to form an algebraic equation system. Then identifiability is assessed by investigating whether the algebraic equation system is symbolically solvable, by determining the number of solutions for the parameter set under investigation [Walter and Pronzato (1997)].

Consider the model structure defined in eq. (1), with f and h infinitely continuously differentiable and let

$$\mathbf{a}_k\left(\mathbf{p}\right) = \lim_{t \to t_0} \frac{d^k}{dt^k}\mathbf{y}\left(t, \mathbf{p}\right) \tag{3}$$

The condition $\mathbf{y}\left(t, \mathbf{p}\right) = \mathbf{y}\left(t, \mathbf{p}^*\right)$ implies

$$\mathbf{a}_k\left(\mathbf{p}\right) = \mathbf{a}_k\left(\mathbf{p}^*\right), \quad k = 0, 1, \ldots$$

A sufficient condition for $\mathcal{M}$ to be structurally globally identifiable is therefore [Pohjanpalo (1978); Walter and Pronzato (1997)]:

$$\mathbf{a}_k\left(\mathbf{p}\right) = \mathbf{a}_k\left(\mathbf{p}^*\right), \quad k = 0, 1, \ldots, k_{max}, \implies \mathbf{p} = \mathbf{p}^* \tag{4}$$

where k_{max} is some positive integer.

3. GLOBAL SENSITIVITY ANALYSIS

Several categories of sensitivity analysis methods are presented in [Saltelli et al. (2008)]. In this study we will only use global methods providing quantitative results while incorporating the entire uncertainty range of parameters. This is of particular importance in biological modeling, since the model parameters can vary within large intervals (two or three decades) depending on their physiological

meaning. An additional advantage of global sensitivity analysis is that the sensitivity estimates of individual parameters are evaluated while varying all other parameters as well. In this way, the relative variability of each parameter is taken into account, thus revealing any existing interactions.

Hereafter, we have adapted the Sobol' sensitivity method [Sobol' (2001)] to dynamic systems, and we adjust the terminology from "sensitivity indice" for static systems to "sensitivity functions" for dynamic systems. In order to analyze the global sensitivity of the output variable with respect to the model parameters, all other computational factors which could affect the output have to be kept constant (as simulation method, sampling time, input signals etc.).

Hypothesis of the method: (H_1) the parameters are considered as i.i.d random variables uniformly distributed over $[0,1]^{n_p}$, with n_p the number of model parameters; (H_2) $\mathbf{y}(t,\mathbf{p})$ is continuous, derivative and square integrable.

The Sobol' method is a variance-based method that allows the computation of both the sensitivity functions of individual parameters and those of interactions between these parameters. This approach is based on a high dimensional representation of the output variable,

$$\mathbf{y}(t,\mathbf{p}) = \mathbf{y}_0(t) + \sum_{i=1}^{n_p} \mathbf{y}_i(t,p_i) + \sum_{i=1}^{n_p-1}\sum_{j>i}^{n_p} \mathbf{y}_{i,j}(t,p_i,p_j) + \ldots + \mathbf{y}_{1,\ldots,n_p}(t,p_1,\ldots,p_{n_p}) \quad (5)$$

where the terms of the decomposition have the following properties:

$$\mathbf{y}_0(t) = \int_{[0,1]^{n_p}} \mathbf{y}(t,\mathbf{p})\mathbf{dp} \quad (6)$$

$$\int_0^1 \mathbf{y}_{i_1,\ldots,i_r}(t,p_{i_1},\ldots,p_{i_r})dp_k = 0 \quad (7)$$

with $k \in \{i_1,\ldots,i_r\}$ and $1 \le i_1 < \ldots < i_r \le n_p$. Equation (7) also implies

$$\int_0^1 \mathbf{y}_{i_1,\ldots,i_r}(t,p_{i_1},\ldots,p_{i_r}) \cdot \mathbf{y}_{j_1,\ldots,i_s}(t,p_{j_1},\ldots,p_{j_s})dp_k = 0$$

i.e the terms of the Sobol' decomposition are orthogonal, with $\{i_1,\ldots,i_r\} \neq \{j_1,\ldots,j_s\}$ and $1 \le j_1 < \ldots < j_s \le n_p$. Moreover, if $\mathbf{y}(t,\mathbf{p})$ is square integrable, then it's variance, denoted $V(t)$, is given by

$$\begin{aligned} V(t) &= \int_{[0,1]^{n_p}} \left(\mathbf{y}^2(t,\mathbf{p}) - \mathbf{y}_0^2(t)\right)\mathbf{dp} \\ &= \sum_{i=1}^{n_p} \int_{[0,1]^{n_p}} \mathbf{y}_i^2(t,p_i)\mathbf{dp} \\ &\quad + \sum_{i=1}^{n_p-1}\sum_{j>i}^{n_p} \int_{[0,1]^{n_p}} \mathbf{y}_{i,j}^2(t,p_i,p_j)\mathbf{dp} \\ &\quad \vdots \\ &\quad + \int_{[0,1]^{n_p}} \mathbf{y}_{1,\ldots,n_p}^2(t,p_1,\ldots,p_{n_p})\mathbf{dp} \end{aligned}$$

due to the orthogonality of the terms in eq. (5). Since the parameters are i.i.d. and uniformily random variables distributed over $[0,1]^2$, and according to the properties (6) and (7), it can be deduced that

$$V(t) = \sum_{i=1}^{n_p} V_i(t) + \sum_{i=1}^{n_p-1}\sum_{j>i}^{n_p} V_{i,j}(t) + \ldots + V_{1,\ldots,n_p}(t) \quad (8)$$

where $V_{i_1,\ldots,i_r}(t)$ represents the variance of the model output by varying simultaneously the parameters p_{i_1}, …, p_{i_r}. Sensitivity functions are defined as

$$S_{i_1,\ldots,i_r} = \frac{V_{i_1,\ldots,i_r}(t)}{V(t)} \quad (9)$$

First-order sensitivity functions, $S_i(t)$ represent the direct sensitivity w.r.t. the parameter p_i, while $S_{i_1,\ldots,i_r}(t)$ represents the sensibility w.r.t. a group of parameters. The total sensitivity functions, $S_{T_i}(t)$, regroup the sensitivity of the model output w.r.t. the influence of a parameter in the different forms (direct and interactions with other parameters), and will be defined as

$$S_{T_i}(t) = S_i(t) + \sum_{j=1}^{n_p} S_{i,j}(t) + \ldots + S_{1,\ldots,n_p}(t) \quad (10)$$

The sensitivity functions (namely first-order and total sensitivity functions) are either calculated by computer algebra if the explicit form of $\mathbf{y}(t,\mathbf{p})$ is known or estimated by Monte Carlo methods otherwise. In the latter case, an estimation algorithm is given in [Saltelli et al. (2008)].

4. IMPLICATIONS OF GLOBAL SENSITIVITY AND IDENTIFIABILITY ANALYSIS

4.1 Definition of Ψ_i, Ω_i and $\Omega_{i,j}$ functions

Let $\Psi_i(t,\mathbf{p})$ be the total effect on the model output $\mathbf{y}$ of a parameter p_i, based on equation (5), and defined as

$$\Psi_i(t,\mathbf{p}) = \mathbf{y}_i(t,p_i) + \sum_{j\neq i}^{n_p} \mathbf{y}_{i,j}(t,p_i,p_j) + \ldots + \mathbf{y}_{1,\ldots,n_p}(t,p_1,\ldots,p_{n_p}) \quad (11)$$

which can be further decomposed as

$$\Psi_i(t,\mathbf{p}) = \Omega_i(t,\mathbf{p}_{\sim j}) + \Omega_{i,j}(t,\mathbf{p}) \quad (12)$$

whereas $\Omega_i(t,\mathbf{p}_{\sim j})$ represents the influence on the output $\mathbf{y}$ of the parameter p_i, independently of p_j and $\Omega_{i,j}(t,\mathbf{p})$ its complementary effect on $\mathbf{y}$ (corresponding to the combined action on $\mathbf{y}$ of p_i and p_j). $\mathbf{p}_{\sim j}$ correspond to the set of all parameters except p_j. See section 4.5 for examples of the definition of Ω-functions. Therefore, the total variance function of the output $\mathbf{y}$ w.r.t the influence of the parameter p_i is defined by

$$V_{T_i}(t) = \int_{[0,1]^{n_p}} \Omega_i^2(t,\mathbf{p}_{\sim j})\mathrm{d}\mathbf{p} + \int_{[0,1]^{n_p}} \Omega_{i,j}^2(t,\mathbf{p})d\mathbf{p} \quad (13)$$

as all the terms of the Sobol' decomposition (5) are orthogonal.

4.2 The link between the nullity of a total sensitivity function and non-identifiability

In [Sobol' (2001)], it was proven that a null total sensitivity function, $S_{T_i}(t) = 0$, leads the nullity of all functions depending on the parameter in p_i, i.e.

$$\mathbf{y}_i(t,p_i) = \mathbf{y}_{i,k}(t,p_i,p_k) = \ldots = \mathbf{y}_{1,\ldots,n_p}(t,\mathbf{p}) = 0,$$

and therefore $\Psi_i(t,\mathbf{p}) = \Omega_i(t,\mathbf{p}_{\sim k}) + \Omega_{i,k}(t,\mathbf{p}) = 0$, with $k \neq i$. Consequently, we can write $\mathbf{y}(t,\mathbf{p})$ as a function of $n_p - 1$ parameters, $\mathbf{p}_{\sim i}$. In other terms, p_i has no influence on the output, and is thus non-identifiable.

4.3 Linear dependency of Ω_i, Ω_j and $\Omega_{i,j}$ functions and their link with the p_i or p_j non-identifiability

If the case of null sensitivity functions is generally acknowledged as corresponding to non-identifiable parameters, the case of correlated sensitivity functions and their link with the parameter non-identifiability is less clear.

One possible cause for the time correlation of the sensitivity functions is the linear dependence, with respect to time, of two distinct $\Psi_i(t,\mathbf{p})$ and $\Psi_j(t,\mathbf{p})$ functions decomposed as:

$$\begin{aligned}\Psi_i(t,\mathbf{p}) &= \Omega_i(t,\mathbf{p}_{\sim j}) + \Omega_{i,j}(t,\mathbf{p})\\ \Psi_j(t,\mathbf{p}) &= \Omega_j(t,\mathbf{p}_{\sim i}) + \Omega_{i,j}(t,\mathbf{p})\end{aligned}$$

and which is itself caused by the linear dependence, two by two, of the functions Ω_i, $\Omega_{i,j}$ and Ω_j (see appendix A for further details).

Let us consider now the following propositions, which describes the link between dependent functions Ω_i, $\Omega_{i,j}$,Ω_j and the non-identifiability of p_i or p_j.

Proposition 4.1. The linear dependence, with respect to time, of the functions $\Omega_i(t,\mathbf{p}_{\sim j})$, $\Omega_{i,j}(t,\mathbf{p})$ and $\Omega_j(t,\mathbf{p}_{\sim i})$, implies that they fulfill the following properties:

$$\begin{aligned}\Omega_i(t,\mathbf{p}_{\sim j}) &= h_i(p_i)\cdot g(t,\mathbf{p}_{\sim i,j})\\ \Omega_j(t,\mathbf{p}_{\sim i}) &= h_j(p_j)\cdot g(t,\mathbf{p}_{\sim i,j})\\ \Omega_{i,j}(t,\mathbf{p}) &= h_{i,j}(p_i,p_j)\cdot g(t,\mathbf{p}_{\sim i,j})\end{aligned}$$

Proof To simplify reading, we will consider a two-parameter model $\mathbf{y}(t,\mathbf{p})$ with $\mathbf{p} = [p_1,p_2]^T \in [0,1]^2$. According to the Sobol' decomposition, its output variable, $\mathbf{y}(t,\mathbf{p})$, can be represented by

$$\mathbf{y}(t,\mathbf{p}) = \mathbf{y}_0(t) + \mathbf{y}_1(t,p_1) + \mathbf{y}_2(t,p_2) + \ + \mathbf{y}_{1,2}(t,p_1,p_2)$$

Herein, $\Omega_1(t,\mathbf{p}_{\sim 2}) = \mathbf{y}_1(t,p_1)$, $\Omega_2(t,\mathbf{p}_{\sim 1}) = \mathbf{y}_2(t,p_2)$ and $\Omega_{1,2}(t,\mathbf{p}) = \mathbf{y}_{1,2}(t,p_1,p_2)$ according to the definition of the Ω-functions in section 4.1.

Let us consider the linear dependence, w.r.t. time, between Ω_1 and $\Omega_{1,2}$, Ω_2 and $\Omega_{1,2}$ and respectively Ω_1 and Ω_2:

$$\alpha_1(\mathbf{p})\,\Omega_1(t,\mathbf{p}_{\sim 2}) + \alpha_{1,2}(\mathbf{p})\,\Omega_{1,2}(t,\mathbf{p}) = \mathbf{0} \qquad (14)$$

$$\beta_2(\mathbf{p})\,\Omega_2(t,\mathbf{p}_{\sim 1}) + \beta_{1,2}(\mathbf{p})\,\Omega_{1,2}(t,\mathbf{p}) = \mathbf{0} \qquad (15)$$

$$\gamma_1(\mathbf{p})\,\Omega_1(t,\mathbf{p}_{\sim 2}) + \gamma_2(\mathbf{p})\,\Omega_2(t,\mathbf{p}_{\sim 1}) = \mathbf{0} \qquad (16)$$

whereas α_1, $\alpha_{1,2}$, β_2, $\beta_{1,2}$, γ_1 and γ_2 are some non-null coefficients (parameter functions) and $\alpha_1(\mathbf{p}) \neq \alpha_{1,2}(\mathbf{p})$, $\beta_2(\mathbf{p}) \neq \beta_{1,2}(\mathbf{p})$ (assuming that Ψ_1 and Ψ_2 are not null). Since Ω_1 and Ω_2 depend on time and each on a distinct parameter, we can conclude that in order to have the linear dependence w.r.t. time, these functions must be factorized as follows

$$\begin{aligned}\Omega_1(t,\mathbf{p}_{\sim 2}) &= h_1(p_1)\cdot g(t)\\ \Omega_2(t,\mathbf{p}_{\sim 1}) &= h_2(p_2)\cdot g(t).\end{aligned}$$

This being said, eq. (15) becomes

$$\alpha_1(\mathbf{p})\,h_1(p_1)\cdot g(t) + \alpha_{1,2}(\mathbf{p})\,\Omega_{1,2}(t,\mathbf{p}) = \mathbf{0} \qquad (17)$$

allowing us to affirm that the linear dependency between Ω_1 and $\Omega_{1,2}$ w.r.t. time is assured iff $\Omega_{1,2}(t,\mathbf{p})$ can be factorized as $h_{1,2}(p_1,p_2)\cdot g(t)$.

The consequence of this linear dependence on the non-identifiability of parameters is stated below:

Proposition 4.2. The linear dependence, w.r.t. time, of the functions $\Omega_i(t,\mathbf{p}_{\sim j})$, $\Omega_{i,j}(t,\mathbf{p})$ and $\Omega_j(t,\mathbf{p}_{\sim i})$ leads to the non-identifiability of the parameter p_i or p_j.

Proof Based on the proposition 4.1, we can write $\mathbf{y}(t,\mathbf{p})$ as

$$\mathbf{y}(t,\mathbf{p}) = \mathbf{y}_0(t) + h(\mathbf{p})\cdot g(t)$$

with $h(\mathbf{p}) = h_1(p_1) + h_2(p_2) + h_{1,2}(\mathbf{p})$.

The global identifiability analysis amounts to test whether or not the system of equations described in (4) has a unique solution $\mathbf{p} = \mathbf{p}^*$. Applying the Taylor series approach in our case, reveals that

$$\mathbf{a}_k(\mathbf{p}) = \lim_{t\to t_0}\mathbf{y}_0(t) + h(\mathbf{p})\cdot\lim_{t\to t_0}\frac{d^k}{dt^k}g(t)$$

and hence

$$\mathbf{a}_k(\mathbf{p}) = \mathbf{a}_k(\mathbf{p}^*) \Longrightarrow h(\mathbf{p}) = h(\mathbf{p}^*) \quad \forall k \qquad (18)$$

In other terms, the system of equations (18) is only composed of one equation with two unknown parameters. Since Ψ_1 et Ψ_2 are assumed to be not-null, it is then impossible to identify both p_1 and p_2.

4.4 Linear dependency of Ω_i, Ω_j and $\Omega_{i,j}$ functions and their relation with the colinearity of the total sensitivity functions S_{T_i} and S_{T_j}

Proposition 4.3. The linear dependence of the functions $\Omega_i(t,\mathbf{p}_{\sim j})$, $\Omega_{i,j}(t,\mathbf{p})$ and $\Omega_i(t,\mathbf{p}_{\sim j})$, with respect to time, implies the colinearity of the total sensitivity functions $S_{T_i}(t)$ and $S_{T_j}(t)$.

Proof Taking into account the proposition 4.1 and the definition of the functions Ψ_1 and Ψ_2 for a two-parameter model, we can write:

$$\begin{aligned}\Psi_1(t,\mathbf{p}) &= (h_1(p_1) + h_{1,2}(\mathbf{p}))\cdot g(t)\\ \Psi_2(t,\mathbf{p}) &= (h_2(p_2) + h_{1,2}(\mathbf{p}))\cdot g(t)\end{aligned}$$

Furthermore, the total variance functions may be written as

$$V_{T_1}(t) = \left(\int_0^1 h_1^2(p_1)\,dp_1 + \int_{[0,1]^2} h_{1,2}^2(\mathbf{p})\,d\mathbf{p}\right)\cdot g^2(t)$$

$$V_{T_2}(t) = \left(\int_0^1 h_2^2(p_2)\,dp_2 + \int_{[0,1]^2} h_{1,2}^2(\mathbf{p})\,d\mathbf{p}\right)\cdot g^2(t)$$

were we can observe the linear dependency of $V_{T_1}(t)$ and $V_{T_2}(t)$. As $V(t)$ is a not-null function, then the linear dependency of $V_{T_1}(t)$ and $V_{T_2}(t)$ will imply also the colinearity of the total sensitivity functions $S_{T_1}(t)$ and $S_{T_2}(t)$.

4.5 Academic examples

Example illustrating the theoretical analysis

Let us consider a four-parameter model defined by

$$\begin{aligned}\mathbf{y}(t,\mathbf{p}) = {} & p_2p_3(1-\exp(-t)) +\\ & p_1p_4(1-\exp(-tp_2))\end{aligned} \qquad (19)$$

employed to illustrate the propositions 4.1, 4.2 and 4.3. It clearly appears that p_1 and p_4 are not both identifiable.

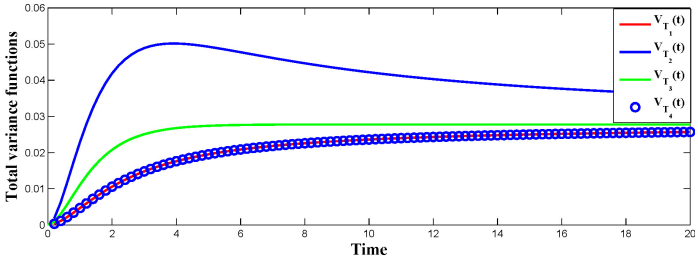

Fig. 2. Total variance functions for the example expressed in eq. (19)

We computed by computer algebra the $2^{n_p} = 16$ functions that compose the Sobol' high dimensional representation in (5), followed by the functions Ω_1, Ω_4 and $\Omega_{1,4}$:

$$\begin{aligned}
\Omega_1(t, \mathbf{p}_{\sim 4}) &= y_1(t, p_1) + y_{1,2}(t, p_1, p_2) + y_{1,3}(t, p_1, p_3) \\
&\quad + y_{1,2,3}(t, p_1, p_2, p_3) \\
&= 0.25(2p_1 - 1)(1 - \exp(-t \cdot p_2)) \\
\Omega_4(t, \mathbf{p}_{\sim 1}) &= y_4(t, p_4) + y_{2,4}(t, p_2, p_4) + y_{3,4}(t, p_3, p_4) \\
&\quad + y_{2,3,4}(t, p_2, p_3, p_4) \\
&= 0.25(2p_4 - 1)(1 - \exp(-t \cdot p_2)) \\
\Omega_{1,4}(t, \mathbf{p}_{\sim 1}) &= y_{1,4}(t, p_1, p_4) + y_{1,2,4}(t, p_1, p_2, p_4) \\
&\quad + y_{1,3,4}(t, p_1, p_3, p_4) \\
&\quad + y_{1,2,3,4}(t, p_1, p_2, p_3, p_4) \\
&= 0.25(2p_4 - 1)(2p_1 - 1)(1 - \exp(-t \cdot p_2))
\end{aligned}$$

This result is conform to the form of the Ω-functions in Prop. 4.1, with $g(t, \mathbf{p}_{\sim 1,4}) = 0.25(1 - \exp(-t \cdot p_2))$. Moreover, we can remark that Ω_1, Ω_4 and $\Omega_{1,4}$ are - two by two - linear dependent w.r.t. time. Finally, the total variance functions associated with p_1 and p_4 are

$$\begin{aligned}
V_{T_1}(t) &= \int_{[0,1]^4} \left(\Omega_1^2(t, \mathbf{p}_{\sim 4}) + \Omega_{1,4}^2(t, \mathbf{p})\right) \mathrm{d}\mathbf{p} \\
&= \frac{2t + 4\exp(-t) - 3 - \exp(-2t)}{72t} \\
V_{T_4}(t) &= \int_{[0,1]^4} \left(\Omega_4^2(t, \mathbf{p}_{\sim 1}) + \Omega_{1,4}^2(t, \mathbf{p})\right) \mathrm{d}\mathbf{p} \\
&= \frac{2t + 4\exp(-t) - 3 - \exp(-2t)}{72t}
\end{aligned}$$

This result shows the colinearity of V_{T_1} and V_{T_4} and thus the one of sensitivity functions S_{T_1} and S_{T_4}, as shown also in Fig. 2.

Example illustrating the practical analysis

Let us consider another simple example with

$$\mathbf{y}(t, \boldsymbol{\theta}) = (\theta_1 + \theta_2 \sin(2\pi t)) \cdot \exp(-0.1t) + \exp\left(-\theta_3 t^2\right)$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]^T$ the model parameters, and $\theta_1 \in [0, 2]$, $\theta_2 \in [0, 4]$ and $\theta_3 \in [0, 1]$. This example is meant to illustrate the difference between global theoretical identifiability and global practical identifiability (insights from a global sensitivity analysis point of view).

In order to decompose $\mathbf{y}$, the hypotheses H_1 and H_2 must be fulfilled, i.e we have to consider a set of normalized parameters gathered in $\mathbf{p}$ defined as

$$p_i = \frac{\theta_i - \theta_i^{min}}{\theta_i^{max} - \theta_i^{min}}$$

with $i = 1, 2, 3$. The new expression of $\mathbf{y}$ is

$$\mathbf{y}(t, \mathbf{p}) = (2p_1 + 4p_2 \sin(2\pi t)) \cdot \exp(-0.1t) + \exp\left(-p_3 t^2\right)$$

Its HDMR components are:

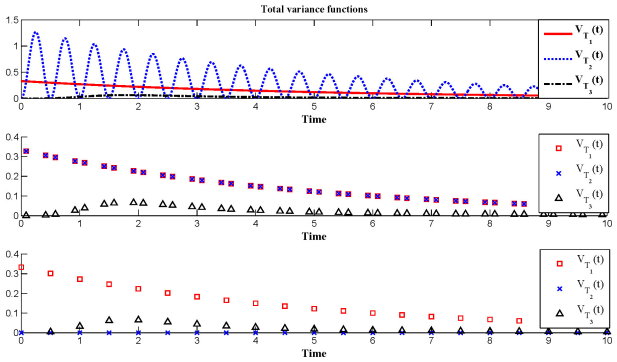

Fig. 3. Total variance functions of the 2nd example, for different time-sampling strategies

$$\begin{aligned}
\mathbf{y}_1(t, p_1) &= (2p_1 - 1) \cdot \exp(-0.1t) \\
\mathbf{y}_2(t, p_2) &= 2(2p_2 - 1) \cdot \sin(2\pi t) \cdot \exp(-0.1t) \\
\mathbf{y}_3(t, p_3) &= \exp\left(-p_3 t^2\right) + \frac{\exp(-t^2) - 1}{t^2}
\end{aligned}$$

In this case, all other terms of the Sobol' decomposition (see eq. (5)) are null. An *a priori* identifiability analysis (through a Taylor series approach) indicates three global identifiable parameters. The associated total variance functions, obtained by computer algebra, will be defined by

$$\begin{aligned}
V_{T_1}(t) &= 0.33 \cdot \exp(-0.2t) \\
V_{T_2}(t) &= 1.33 \cdot \sin(2\pi t)^2 \cdot \exp(-0.2t) \\
V_{T_3}(t) &= 0.5\left(1 - \exp(-2t^2)\right) - \\
&\quad - \frac{1 - 2\exp(-t^2) + \exp(-2t^2)}{t^4}
\end{aligned}$$

These functions are plotted in Fig. 3 with three different sampling strategies for $t_k \in [0, 10]$. In the upper figure, we propose a good sampling rate, $t_k = 0.01 \cdot k$. A first conclusion that can be drawn from this figure is that even in a best case scenario p_3 is a poor-sensitive parameter as $V_{T_3}(t) < 0.1$ and $S_{T_3}(t) < 0.2$. In the middle figure, we chose measurement time instants as to assure colinear total variance functions, $V_{T_1}(t)$ and $V_{T_2}(t)$. For such sampling instants $\sin(2 \cdot \pi \cdot t_k)^2 = 1/4$, allowing as to write $\mathbf{y}$ as

$$\mathbf{y}(t, \mathbf{p}) = (2p_1 \pm 2p_2) \cdot \exp(-0.1t) + \exp\left(-p_3 t^2\right) \quad (20)$$

indicating also the non-identifiablity of parameters. The bottom figure exploits the total variance function for $t_k = 0.5 \cdot k$. In this case, for all t_k, the total variance $V_{T_2}(t_k)$ is null, leaving only one sensitive parameter, p_1.

5. DISCUSSION

We have studied herein two links between the non-identifiability of parameters and the properties of global sensitivity functions. One corresponds to a generally acknowledged association: insensitive parameters mean non-identifiable parameters and a null total sensitivity function. The second is a less corroborated link: the consequences of the linear dependence between the Ω - functions on the identifiability of parameters and global sensitivity measures. Let us analyse the Fig. 4 representing this last bridge between identifiability and sensitivity.

The correspondence between linear dependent Ω functions (w.r.t. time) and the parameter non-identifiability was tackled through a Taylor series approach (this was only a

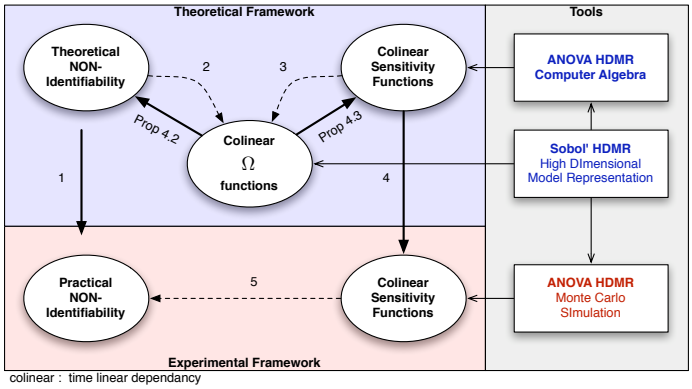

Fig. 4. Schematic representation of relationship between identifiability and sensitivity

subjective choice, any other identifiability method could be applied) indicating a theoretical global non-identifiability (Prop. 4.2). But theoretical identifiability is only a necessary identifiability condition, designating also a global practical non-identifiability (arrow **1** in Fig. 4).

The collinearity w.r.t. time of the Ω-functions also implies the collinearity of the total sensitivity functions (Prop. 4.3). In practice, the latter sensitivity functions are rarely explicitly determined. A more realistic approach is to estimate them by a Monte Carlo simulation approach (arrow 4 in Fig. 4) that takes experimental conditions into account.

Dashed arrows 2, 3 and 5 in Fig.4 indicate ongoing works. The conclusions on global practical identifiability from the analysis of the empirical sensitivity functions (arrow 5) mainly depends now on the reciprocal condition described by arrow 3.

6. CONCLUSION

Due to the lack of methods to test the global practical identifiability, we focused on global sensitivity methods. Global sensitivity measures are the result of an *a posteriori* study, in which the time instants and input signals are fixed. It is natural to think of it also as an *a posteriori* identifiability measure. Nevertheless, the links between these two domains are not yet fully explored.

This present work provides new insights into the relationships between these two concepts. In perspective, these results could lead to the development of new approaches to test the non-identifiability of parameters in an experimental framework. Systems biology is a particular application area of such solutions.

The present results are based exclusively on the output trajectories in explicit algebraic form. A more realistic case study is presented in Dobre et al. (2010).

REFERENCES

Audoly, S., Bellu, G., D'Angio, L., Saccomani, M.P., and Cobelli, C. (2001). Global identifiability of nonlinear models of biological systems. *IEEE Transactions on Biomedical Engineering*, 48(1), 55–65.

Dobre, S., Bastogne, T., and Richard, A. (2010). Global sensitivity and identifiability implications in systems biology. In *Proc of the 11th IFAC Symposium on Computer Applications in Biotechnology*.

Dochain, D., Vanrolleghem, P.A., and Daele, M.V. (1995). Structural identifiability of biokinetic models of activated sludge respiration. *Water Research*, 29(11), 2571–2578.

Ghidaoui, M.S. and Prasad, K.H. (2000). A priori identifiability of unsaturated soil parameters. *Journal of irrigation and drainage engineering*, 126, 163–171.

Grewal, M.S. and Glover, K. (1976). Identifiability of linear and nonlinear dynamical systems. *IEEE Trans. on Automatic Control*, 21(6), 833–837.

Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics*, 2, 343–372.

Ljung, L. and Glad, T. (1994). On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2), 265–276.

Peeters, R.L. and Hanzon, B. (2005). Identifiability of homogeneous systems using the state isomorphism approach. *Automatica*, 41, 513–529.

Pohjanpalo, H. (1978). System identifiability based on power series expansion of the solution. *Mathematical Biosciences*, 41, 21–33.

Saccomani, M.P. (2004). Some results on parameter identification of nonlinear systems. *Cardiovascular Engineering: An international journal*, 4(1), 95–102.

Saccomani, M.P., Audoly, S., and D'Angio, L. (2003). Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, 39, 619–632.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis : The primer*.

Sobol', I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3), 271–280. doi:http://dx.doi.org/10.1016/S0378-4754(00)00270-6.

Vanrolleghem, P.A., Daele, M.V., and Dochain, D. (1995). Practical identifiability of a biokinetic model of activated sludge respiration. *Water Research*, 29(11), 2561–1570.

Walter, E. and Pronzato, L. (1997). *Identification of Parametric Models from experimental data.*

Appendix A. DEPENDENCE OF Ψ_I AND Ψ_J

In the literature, the non-estimability of parameters is explained sometimes by the compensation on the ouput $\mathbf{y}$ of the effects of two (or more) parameters. But, the function Ψ_i is supposed to reflect the total effect of the parameter p_i, and V_{T_i} depends only on Ψ_i.

How can the parameter p_j compensate the effect of p_i ?

This compensation must take place within Ψ_i, meaning through Ω_i and $\Omega_{i,j}$ (that's why we decomposed Ψ_i in a part which is independent of p_i, Ω_i, and its complement $\Omega_{i,j}$ which characterize the joined action between p_i and p_j). Therefore, a cause of compensated effect is the linear dependence of the functions Ω_i and $\Omega_{i,j}$, respectively Ω_j and $\Omega_{i,j}$. This will also imply the dependence of Ψ_i and Ψ_j.