

Research Article

The Union between Structural and Practical Identifiability Makes Strength in Reducing Oncological Model Complexity: A Case Study

Maria Pia Saccomani ¹ and **Karl Thomaseth**²

¹Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy

²IEIT-CNR, c/o Department of Information Engineering, University of Padova, Via Gradenigo 6/a, 35131 Padova, Italy

Correspondence should be addressed to Maria Pia Saccomani; pia@dei.unipd.it

Received 9 September 2017; Accepted 14 January 2018; Published 11 February 2018

Academic Editor: Peter Giesl

Copyright © 2018 Maria Pia Saccomani and Karl Thomaseth. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mathematical models are increasingly proposed to describe tumor's dynamic response to treatments with the aims of improving their efficacy. The most widely used are nonlinear ODE models, whose identification is often difficult due to experimental limitations. We focus on the issue of parameter estimation in model-based oncological studies. Given their complexity, many of these models are unidentifiable having an infinite number of parameter solutions. These equivalently describe experimental data but are associated with different dynamic evolution of unmeasurable variables. We propose a joint use of two different identifiability methodologies, *structural identifiability* and *practical identifiability*, which are traditionally regarded as disjoint. This new methodology provides the number of parameter solutions, the analytic relations between the unidentifiable parameters useful to reduce model complexity, a ranking between parameters revealing the most reliable estimates, and a way to disentangle the various causes of nonidentifiability. It is implementable by using available differential algebra software and statistical packages. This methodology can constitute a powerful tool for the oncologist to discover the behavior of inaccessible variables of clinical interest and to correctly address the experimental design. A complex model to study “in vivo” antitumor activity of interleukin-21 on tumor eradication in different cancers in mice is illustrated.

1. Introduction

Many mathematical models of tumor growth at different levels from gene expression to the macroscopic tumor development have been formulated, [1–11]. Recently, mathematical models of tumor-immune interactions have also been considered to evaluate the efficacy of immunotherapy in the context of tumor challenge. Models of cancer treatments, both chemotherapy or/and immunotherapy, have been also widely employed. Furthermore, pharmacokinetic-pharmacodynamic (PK-PD) models are developed to describe the interaction between tumor growth, drug absorption, and effect of the drug in terms of patients' response to therapies. Among different mathematical frameworks used to describe these models, the most widely used in *mathematical oncology* are based on nonlinear ordinary differential equations (ODE). These models are being intensely studied

to describe the complex processes of tumorigenesis and to produce an integrated mathematical view of tumorigenesis, cancer progression, and evaluation of anticancer agents under different oncological settings. Often, mathematical models are combined with optimal control techniques for quantitatively describing tumor progression and optimal treatment planning. Clinically validated mathematical models have been proposed for the development of the so-called “virtual patient” [12] to accurately predict efficacy and toxicity of various oncological therapeutic combinations in individuals and in populations. Many of these models have been successfully simulated and validated against clinical observations, promoting modeling and optimization control as a therapy planning tool in clinic.

Given the increased model complexity required to describe the more and more available data, many of the models employed in oncological studies are unidentifiable;

that is, they have an infinite number of parameter solutions. However, this problem is not always recognized. Typically, these solutions are equivalently describing experimental data, but they are associated with different dynamic evolution of the not directly measurable variables. Such a situation is undesirable and frustrates one of the most useful aspects of mathematical models, that is, that of providing a means to infer *unobservable* quantities and time-varying phenomena.

By starting from these observations, in this paper, we focus on unidentifiable models, in particular on the mathematical issue of parameter estimation [13] in model-based oncological studies. This issue will become crucial also to increase the precision of the recently proposed treatments personalization methods.

To guarantee goodness and reliability of the parameter estimation results, we propose a new joint use of two different identifiability methodologies, namely, *structural identifiability* and *practical identifiability*, which are traditionally regarded as disjoint because they are based, in turn, on differential algebraic manipulations or numerical simulation of systems equations. Nevertheless, also the structural analysis can provide useful practical information if applied in a particular point of the admissible parameter space of clinical interest.

We first propose an algorithmic method to count the number of parameter solutions of a model, technically speaking to check the *structural identifiability* of the model [14, 15]. In particular, if the model parameterization is unique (*global identifiable* model), the numerical estimate of the unknown parameter provided by whatever optimization algorithm is correct and allows arriving at reliable conclusions.

From a methodological point of view, testing structural identifiability before collecting experimental data is an essential prerequisite for assessing whether the experimental design is adequate for a hypothesized model and whether the parameter estimation problem is well posed. In the model-based oncological studies literature, however, the identifiability issue is still neglected, and collection of experimental data precedes the formulation of mathematical models, which is often carried out by trial and error by fitting different model structures to the acquired data. This approach is surely dictated by the fact that there are many software tools to perform parameter estimation, while checking identifiability in some cases can be prohibitively complex, for example, for large models containing many states and parameters. However, if the postulated model has an infinite number (*unidentifiable*) of parameter solutions, the parameter estimates that could still be obtained by some numerical optimization algorithms would be unreliable and vary randomly depending, for instance, on the initialization of the algorithm [15]. Vice versa, in case of nonidentifiability, the outcome of structural identifiability, that is, the Gröbner bases of the exhaustive summary, provides algebraic nonlinear equations which define the relations between unidentifiable parameters [16, 17]. These equations describe the equivalence class of parameters with respect to their ability to describe the output function. Just by inspection, one can know the degrees of freedom of the system and thus which are the redundant parameters of the model. In particular, the analytic expressions of the

dependencies among parameters can be included in the original model in order to reduce its number of parameters and to define an equivalent identifiable model of reduced complexity.

Although necessary, structural identifiability is not sufficient to guarantee an accurate identification of the model parameters from real, possibly noisy, input/output data. The parameter estimates obtained by standard algorithms, even for a structurally identifiable model, may be very sensitive to noise and a measure of this sensitivity can be important in applications. Thus, there is a need to perform, besides structural identifiability tests, a *practical identifiability* analysis. In many studies, to have an idea of how much the outputs of the model are biased by the parameter values, the sensitivity of the output functions with respect to each parameter is calculated. Usually the parameter correlations are also calculated to try to identify, by trial and error, which are the right parameters to fix in order to calculate the others. In this paper, we show how to analytically calculate the relations between the unidentifiable parameters. These relations remain completely hidden to the investigator when calculating the correlations. From these, it is immediate to know which parameter to fix in order to analytically calculate all the correlated parameters.

In principle, the whole proposed methodology can be checked by suitable mathematical procedures directly on the model, without the need for collecting experimental data. This may avoid waste of resources for doing uninformative experiments, given the high costs, not only in economic terms, of oncological experiments. Only structural identifiability can be tested without assuming prior knowledge on parameter values, whereas practical identifiability, based on sensitivity analysis, requires “nominal” parameter values for numerical simulation [18].

We choose to illustrate our procedure and to show our results in a simple linear model.

Finally, this paper aims to demonstrate that, in oncological studies, when a model is formulated and its parameters need to be estimated from available measurements, checking the uniqueness of the parameter solution is crucial. Since the conclusions of model-based oncological studies are generally founded on the numerical estimates of the unknown parameters from experimental data, by neglecting all the solutions of an unidentifiable parameter (except that estimated with an optimization algorithm), the investigator can arrive at totally erroneous conclusions. Furthermore, the calculation of the analytic relations between the unidentifiable parameters is an effective approach to discover the behavior of nonaccessible variables of clinical interest as well as for scheduling cancer therapy by guaranteeing the reliability of the results. In order to do this, we will apply our methodology to a relevant benchmark model. This is a mathematical model to study and evaluate the “in vivo” antitumor activity of interleukin-21 (IL-21) on tumor eradication in different cancers in mice [4, 8].

2. Mathematical Background

2.1. Definitions. This section provides the reader with the definitions that are necessary to set the notations used in

the paper. Consider a nonlinear dynamic system described in state space form as

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \\ \mathbf{y}(t) &= \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})\end{aligned}\quad (1)$$

with state $\mathbf{x}(t) \in \mathbb{R}^n$, input $\mathbf{u}(t) \in \mathbb{R}^q$ ranging on some vector space of piecewise smooth (infinitely differentiable) functions, output $\mathbf{y}(t) \in \mathbb{R}^m$, and constant unknown parameter vector $\boldsymbol{\theta}$ belonging to some open subset $\Theta \subseteq \mathbb{R}^p$. Whenever initial conditions are specified, the relevant equation $\mathbf{x}(0) = \mathbf{x}_0$ is added to the system. The essential assumption here is that the functions \mathbf{f} and \mathbf{g} are vectors of *rational functions* in \mathbf{x} . We also assume that there is no feedback, so that \mathbf{u} is a free variable not depending on \mathbf{y} .

In the following, we will assume that the input-output map of system (1) which started at the initial state \mathbf{x}_0 exists and we denote it with

$$\mathbf{y} = \psi_{\mathbf{x}_0}(\boldsymbol{\theta}, \mathbf{u}). \quad (2)$$

2.2. Structural Identifiability via Differential Algebra. In general, the assessment of parameter values of ODE models can only be approached indirectly as a parameter estimation problem starting from external, input-output measurements. A basic question is whether the parameters of the model can be determined (uniquely) from input-output measurements, at least for suitable input functions, assuming that all observable variables are error-free. This is a mathematical property called *a priori* or *structural identifiability* of the model. It is a *property of the model alone* and of course it depends on how it is parameterized. Structural identifiability can (and should) in principle be checked before collecting experimental data.

We adopt the definitions of structural identifiability used in Saccomani et al.'s work [16].

Definition 1. System (1) is *a priori globally (or uniquely) identifiable from input-output data* if, for at least a generic set of points $\boldsymbol{\theta}^* \in \Theta$, there exists (at least) one input function \mathbf{u} such that the equation

$$\psi_{\mathbf{x}_0}(\boldsymbol{\theta}, \mathbf{u}) = \psi_{\mathbf{x}_0}(\boldsymbol{\theta}^*, \mathbf{u}) \quad (3)$$

has only one solution $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ for almost all initial states $\mathbf{x}_0 \in X \subseteq \mathbb{R}^n$.

If (3) has generically an infinite number of solutions for all input functions \mathbf{u} , system (1) is *unidentifiable*.

We will apply structural identifiability based on differential algebra and on the free dedicated software DAISY (Differential Algebra for Identifiability of SYstems), [19]. The reader is referred to Audoly et al. [14] and Saccomani et al. [16] for a detailed explanation of the theory behind the software tool and to Bellu et al. [19] for the algorithm.

Briefly, this algorithm permits eliminating the unobserved state variables from system (1) and finding the *input-output relation*: a set of polynomial differential equations involving only the variables $(\mathbf{u}(t), \mathbf{y}(t))$ and their time derivatives describing all input-output pairs satisfying the original

dynamic system. The coefficients of the input-output relation provide a set of (nonlinear) algebraic functions of the unknown parameter $\boldsymbol{\theta}$ of the original model. These functions form the *exhaustive summary* of the model. They appear linearly in the input-output relation so that they can be easily extracted. Identifiability is tested by checking injectivity of the exhaustive summary function with respect to parameter $\boldsymbol{\theta}$. By applying Buchberger's computer algebra algorithm [20], it is possible to compute a Gröbner basis of the system. This algorithm represents a common generalization for nonlinear equations and for more variables of the Gaussian and the Euclidean algorithm, respectively. In particular, the Gröbner basis allows counting the number of solutions of the unknown parameter $\boldsymbol{\theta}$ and shows if parameters satisfy algebraic relations or have instead a one-to-one relation with the exhaustive summary, in which case the model is *globally identifiable*.

DAISY automatically ranks the input, output, state variables and their derivatives, starts the pseudodivision algorithm, and calculates the differential polynomials which form the *input-output relation* of the model. Buchberger's algorithm is then applied to the (nonlinear) algebraic equations obtained after equating the coefficients of the input-output relation to a set of pseudo-randomly chosen numerical points in their range set. DAISY calculates the Gröbner basis of this algebraic nonlinear system and provides the identifiability results holding in all the parameter space.

In general, a Gröbner basis can be represented as

$$\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \{\mathbf{G}_1(\boldsymbol{\theta}, \boldsymbol{\theta}^*), \dots, \mathbf{G}_r(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\} \in \mathbb{R}^r, \quad (4)$$

where $\mathbf{G}_i(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ are algebraic nonlinear polynomials.

The possibly finite or infinite multiple solutions of the system of r equations in the p unknowns $\boldsymbol{\theta}$,

$$\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \mathbf{0}, \quad (5)$$

provide a parametrization of $\boldsymbol{\theta}$, which satisfies (3).

In the case of unique identifiability, the Gröbner basis functions become simply

$$\mathbf{G}_i(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \boldsymbol{\theta}_i - \boldsymbol{\theta}_i^*, \quad \forall i = 1, \dots, p. \quad (6)$$

In case of a unidentifiable model, that is, at least one parameter is unidentifiable, the Gröbner basis (4) provides the analytic relations between the unidentifiable parameters which hold in the whole admissible parameter space, not only around a given parameter value [16]. The $r < p$ solutions of (5) provide a uniquely identifiable parametrization of $\boldsymbol{\theta}$, as function of the known parameters $\boldsymbol{\theta}^*$ and with $p - r$ unknowns that are unidentifiable. These unknowns are free parameters that can be assigned arbitrary values without affecting the input-output relationship (3).

2.3. Practical Identifiability via Sensitivity Analysis. In the literature, *practical identifiability* [21, 22] is generally understood as a study of the sensitivity of some criterion function, for example, the likelihood, with respect to the parameters to be estimated, in particular with the purpose of detecting sensitive or nonunique minima. This can be done on more realistic models which explicitly involve noise in the measurements

and may use actual measurement data subject to disturbances of various nature. Checking practical identifiability by data-based (or simulation-based) procedures cannot, however, provide a mathematically rigorous answer to the uniqueness problem.

However, since for a fixed input function the parameter estimates should minimize a criterion function which depends, besides the parameter vector, on the actual output function, the nonuniqueness of minima can also be studied by studying the sensitivity of the output with respect to parameter variations. It should be evident from the very problem setting that this sensitivity should play a key role in identifiability analysis: obviously a model whose output has zero sensitivity with respect to some parameter variations is clearly indicative of nonidentifiability. But the role goes much farther, since likelihood optimization generally requires the calculation of the gradient of the cost function which in turn depends on the output sensitivities.

The choice of a particular approach for testing practical identifiability can lead, however, to inconclusive or only qualitative results, particularly if random noise is added to numerical simulations in the attempt of making them more realistic. Large unpredictable errors can also occur if model output sensitivities are determined numerically by finite difference approximation, which are easily affected by roundoff errors if sensitivities with respect to parameters are very small. This latter kind of errors can be avoided by algorithmic differentiation. Still, random noise in sensitivity analysis may obfuscate deterministic relationships among parameters that can be assessed only through analytic mathematical approaches.

The practical parameter identification framework considered is based on (simulated) noisy measurements of the dynamic system (see (1)) output taken over a finite horizon at discrete time points $\{t_j, j = 1, \dots, N\}$; that is,

$$\mathbf{z}(t_j) = \mathbf{y}(t_j, \boldsymbol{\theta}^*) + \mathbf{e}(t_j). \quad (7)$$

For notation simplicity, it is assumed that $\mathbf{e}(t_j) \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_j^{-1})$. To check *practical identifiability* from a finite set of N input-output measurements, one can form the average weighted squared *prediction error*:

$$V_N(\boldsymbol{\theta}) := \frac{1}{N} \sum_{j=1}^N [\mathbf{z}(t_j) - \hat{\mathbf{y}}(t_j, \boldsymbol{\theta})]^T \mathbf{W}_j [\mathbf{z}(t_j) - \hat{\mathbf{y}}(t_j, \boldsymbol{\theta})], \quad (8)$$

where $\hat{\mathbf{y}}(t_j, \boldsymbol{\theta})$ is the output predictor based on a generic parameter value $\boldsymbol{\theta}$. Assume that $V_N(\boldsymbol{\theta})$ has only one absolute minimum,

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} V_N(\boldsymbol{\theta}), \quad (9)$$

compared to Ljung [13]. According to Raue et al. [23], the i th component θ_i of the parameter is *practically unidentifiable* if the one-dimensional confidence region about $\hat{\theta}_i$ extends to infinity. Naturally this statement cannot be checked exactly

with real data and needs to be interpreted as an asymptotic statement for sample sizes $N \rightarrow \infty$ when $V_N(\boldsymbol{\theta})$ has an asymptotic distribution of χ^2 type. Approximate confidence regions of parameter estimates can be, however, calculated a priori from the Fisher information matrix or simply from the rank of the sensitivity matrix formed as

$$\mathbf{S}(\boldsymbol{\theta})^T = [\mathbf{S}_1(\boldsymbol{\theta})^T, \mathbf{S}_2(\boldsymbol{\theta})^T, \dots, \mathbf{S}_m(\boldsymbol{\theta})^T], \quad (10)$$

where $\mathbf{S}_i(\boldsymbol{\theta})^T = [\nabla_{\boldsymbol{\theta}} \mathbf{y}_i(t_1), \dots, \nabla_{\boldsymbol{\theta}} \mathbf{y}_i(t_N)]$ and $\nabla_{\boldsymbol{\theta}} \mathbf{y}_i(t_j)$ are the sensitivities of model outputs at sampling times t_j of the i th output components with respect to the parameter vector $\boldsymbol{\theta}$. Without loss of generality, but not without side effects because sensitivity analysis is susceptible to parameter scaling, (10) can be thought of as normalized sensitivities according to various possible definitions. For instance, with heteroscedastic measurement noise, for example, $\mathbf{W}_j^{-1/2}$ in (8), to reduce the effect of parameter scaling and without assumptions on measurement noise, we consider the following sensitivities as derivatives of (unnormalized) model outputs with respect to *fractional* parameter variations or *logarithmic* derivatives; that is, $\mathbf{S} = \{\partial \mathbf{y}_i / \partial \log \theta_j\}_{ij} = \{(\partial \mathbf{y}_i / \partial \theta_j) \theta_j\}_{ij}$.

The above theory is implemented in almost all the statistical model fitting software usually based on the quadratic approximation of the likelihood function or also, for example, on Monte Carlo simulation. The reader is referred to AMIGO [24], PLE [23], and COPASI [25] for the biological and physiological models and to NONMEM [26] and ADAPT [27] for population pharmacokinetic and pharmacokinetic-pharmacodynamic modeling.

Practical identifiability is then tested here by the long-standing Principal Component Analysis, for example, Vajda et al. [28], for which we can finally give the following formal definition.

Definition 2. The system described by (1) and (7) identified by nonlinear least squares is *practically identifiable* if the sensitivity matrix $\mathbf{S}(\boldsymbol{\theta})$ has full rank.

This can be ascertained through Singular Value Decomposition (SVD) which provides the following factorization:

$$\mathbf{S}(\boldsymbol{\theta}) = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T, \quad (11)$$

where $\mathbf{U} \in \mathbb{R}^{m \times N \times m \times N}$ and $\mathbf{V} \in \mathbb{R}^{p \times p}$ are the orthonormal eigenvector matrices of $\mathbf{S}(\boldsymbol{\theta})\mathbf{S}(\boldsymbol{\theta})^T$ and $\mathbf{S}(\boldsymbol{\theta})^T\mathbf{S}(\boldsymbol{\theta})$, respectively, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times N \times p}$ is diagonal (referring to the top $p \times p$ submatrix) with sorted *singular values* $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$, which are also the square roots of the eigenvalues of the positive-semidefinite matrix $\mathbf{S}(\boldsymbol{\theta})^T\mathbf{S}(\boldsymbol{\theta})$. The theoretical (practical) rank of $\mathbf{S}(\boldsymbol{\theta})$ is defined as the smallest $r \leq p$ at which $\sigma_{r+1} = 0$ ($\bar{\sigma} > \sigma_{r+1}$, with $\bar{\sigma}$ being a user-defined threshold). A known application of SVD consists in representing estimated parameter vectors $\boldsymbol{\theta}$ as linear combinations of the first $r \leq p$ eigenvectors of \mathbf{V} , with *significance ranking* given by the singular values [18].

In contrast to this common application, in this paper, we aim to exploit the results of SVD on the lower end of singular

values to provide a ranking between unidentifiable parameters, possibly to distinguish between *strictly* structurally and *loosely* practically unidentifiable parameters. The working hypothesis is that structurally unidentifiable parameters should be associated exactly with zero singular values, whereas practical nonidentifiability should be more vaguely defined.

The SVD algorithm is implemented in all the general purpose software, as MATLAB (MathWorks, Inc., USA) or R [29], normally through standard linear algebra packages, for example, LAPACK [30].

3. New Perspectives on the Joint Use of Structural and Practical Identifiability Analysis

Structural identifiability analysis can provide, compared to practical identifiability analysis based on sensitivities, a much deeper insight into the properties of a given model and can indicate how to reparameterize a model that turns out to be unidentifiable. However, the analytic methods can be usefully employed also a posteriori, after having performed some preliminary model fitting to data. For this purpose, there is in fact no need of having all parameters identifiable, because tuning a model to experimental data is feasible even with overparameterized unidentifiable models [18]. In this context, the structural result may indicate which is the identifiable subset of the parameter vector.

The aim of the present paper is to exploit properties and results of structural identifiability to provide an analytic approach for interpreting parameter estimation results. For this purpose, we summarize already mentioned properties of Gröbner bases.

Practical identifiability techniques are essentially based on simulations and on the study of the level curves of a cost function, typically the likelihood function. Assuming that the minimization yields a *unique* parameter value, the level curves around the minimum define the confidence region. Nonidentifiability is defined in terms of diverging confidence regions in some direction above given thresholds. Instead, structural identifiability provides a dichotomous answer that does not depend on parameter values.

In case of global identifiability, sensitivity analysis proceeds in the classical way to show if for a given experiment design parameters still remain identifiable also in the practical situation around a nominal point (local identifiability). If the parameter turns out to be practically unidentifiable, only if structural identifiability of the model has been first tested is it possible to know whether there is a problem with experiment design or with the model structure, problems that must be solved differently in the two cases, to reach identifiability.

Here we show how practical identifiability analysis, based on model output sensitivities, can take advantage of information provided by structural identifiability analysis based on differential algebra by applying the following line of reasoning.

A practical numerical approach useful to assess (non)identifiability around a nominal point (locally) is to

consider the linear approximation of (5) and to evaluate whether small admissible perturbations in the parameter $\delta\theta$ exist. That is, if the expression

$$\mathbf{G}_i(\theta + \delta\theta, \theta^*) \approx \mathbf{G}_i(\theta, \theta^*) + \nabla \mathbf{G}_i(\theta, \theta^*)^T \delta\theta \quad (12)$$

is satisfied, locally, the perturbation $\delta\theta$ belongs to the null-space of the vector columns of $\nabla \mathbf{G}_i$ or, geometrically, $\delta\theta$ lies on the tangent plane to the constraint surface (5).

In particular, by exploiting the results of SVD, we will see that this joint use of structural and practical identifiability analysis allows the following:

- (1) Distinguishing between identifiable and unidentifiable parameters
- (2) Providing a uniquely identifiable parameterization to reduce the complexity of the model and to correctly proceed with optimization techniques
- (3) Exploiting the analytic relations among unidentifiable parameters described by Gröbner basis. The investigator can thus choose which parameter is convenient to fix, on the basis of its a priori knowledge, to analytically derive the related ones
- (4) Ordering the parameters with respect to their ability to influence the output function. This provides a useful suggestion to the oncologist by indicating which parameters are going to be estimated more precisely than others from the experimental data.

4. A Simple Example

The usefulness of the joint use of structural and practical identifiability analysis to analytically calculate the relation between correlated parameters is applied to a typical example of unidentifiable model: the two-compartment, single-input single-output model, with one accessible pool and elimination from both compartments. The model is linear in the input-output relationship but nonlinear in the parameters, which therefore does not lessen significance of the example. The model is described by the following equations:

$$\begin{aligned} \dot{x}_1 &= -(k_{01} + k_{21})x_1 + k_{12}x_2 + u(t) \\ \dot{x}_2 &= k_{21}x_1 - (k_{02} + k_{12})x_2 \\ y(t) &= x_1(t). \end{aligned} \quad (13)$$

To carry out numerical calculations, the following arbitrary nominal parameter values are assumed: $\theta^* = \{k_{01} = 0.005, k_{02} = 0.003, k_{12} = 0.01, k_{21} = 0.02\}$. The input is modeled as triangular-shaped profile with unit area.

We first perform the *Practical identifiability*, based on sensitivities of the model output trajectory with respect to the parameters, calculated at some nominal values θ^* . Results are shown in Figure 1, which does not evidence the fact that sensitivities are correlated. This information is provided by SVD, where the singular values computed numerically are $\text{diag}(\Sigma) = [318.8, 61.4, 11.2, 4.69 \times 10^{-14}]$ which clearly supports the conclusion that the sensitivity matrix has a reduced

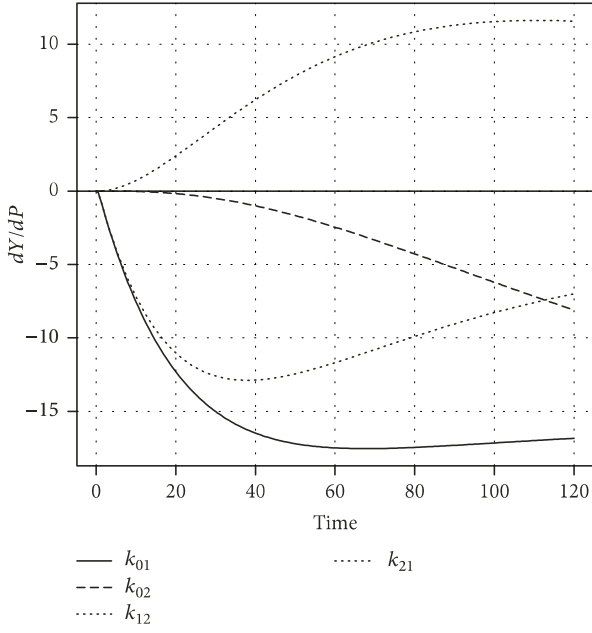


FIGURE 1: Time course of model output sensitivities with respect to parameters.

rank revealing that the model is overparameterized. Thus it should be simplified by reducing the degree of freedom.

Another bit of information that can be derived from the SVD is the right eigenvector associated with the (nearly) zero singular value, which was $\mathbf{v}_4 = [0.6325, -0.3162, 0.3162, -0.6325]$, where 0.6325 is the component corresponding to k_{01} , 0.3162 to k_{02} , 0.3162 to k_{12} , and 0.6325 to k_{21} . It expresses the sensitivity of the output trajectory with respect to *small* perturbations of the parameter vector in this direction. The sensitivity in this case is (nearly) zero and thus a small displacement along \mathbf{v}_4 does not modify the model output. In practice, \mathbf{v}_4 is a locally unfruitful search direction that changes all model parameters but without affecting the system output.

So far we know that the parameters are unidentifiable, but we do not know whether the problem stays in the model structure or in the experimental setting.

Structural identifiability is determined from the Gröbner basis, obtained with a given ranking. In principle, to check structural identifiability, the Gröbner basis could be calculated from a point randomly chosen in the parameter space. Here we apply, for the first time, this structural method to a nominal parameter value, to know not only the number of the parameter solutions and the exact functional relations among the unidentifiable parameters in the whole complex space, which is not so useful, but also the values of the parameter solutions and the above relations around the parameter value of clinical interest. Only the union of the two identifiability approaches allows arriving at these results of practical interest.

For the simple model of (13), one such Gröbner basis is

$$\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \{40 \cdot k_{01} + 40 \cdot k_{21} - 1, 1000 \cdot k_{02} + 1000 \cdot k_{12} - 13, 5000 \cdot k_{12}k_{21} - 1\}, \quad (14)$$

which vanishes, as expected, at the above assumed nominal parameter value; that is, $\mathbf{G}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) = \mathbf{0}$. Note that we represent numerical parameter values as rational numbers; thus all their calculations are without roundoff error.

The equation $\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \mathbf{0}$ does not provide a unique solution in $\boldsymbol{\theta}$, because there are only three equations in four unknowns. This means that, in this example, the investigator would obtain the same input-output behavior obtained with $\boldsymbol{\theta}^*$ by assigning arbitrarily values to one parameter and using (14) to calculate the remaining three. Thus we are now able to assess that nonidentifiability comes from the model structure. This is an important finding because it practically suggests how to solve the nonidentifiability problem: the oncologist should modify the model structure not the experimental setup. This allows avoiding wasting resources in uselessly modifying the experiment. The above Gröbner basis suggests also how to modify the model in order to make it uniquely identifiable. A possible solution, with k_{02} taken as free variable, yields

$$\begin{aligned} k_{21} &= \frac{-1}{5 \cdot (1000 \cdot k_{02} - 13)}, \\ k_{01} &= \frac{200 \cdot k_{02} - 1}{8 \cdot (1000 \cdot k_{02} - 13)}, \\ k_{12} &= \frac{13 - 1000 \cdot k_{02}}{1000}. \end{aligned} \quad (15)$$

These constraints, shown in Figure 2, exactly define a class of equivalence of the model parameters with respect to their ability to describe the output function. Practically speaking, all the parameter values satisfying these constraints are equivalent in describing the output function. However, they predict the unobservable variable x_2 in different ways. This means that if the investigator calculates the parameters from the experimental data and she/he is not aware that parameters are unidentifiable, she/he would ignore that the prediction of x_2 is totally random. Given that in most oncological models the aim of the studies is to infer unmeasured variables, this ambiguity has to be absolutely avoided. The correct way to proceed is thus to eliminate the degree of freedom present in the Gröbner basis (see (14)) to obtain a system having only one solution in the unknown parameters. In order to do this, the investigator has to fix one parameter to a sensible value, known from independent sources. In this case, the nominal value of $k_{02} = 0.003$ of Table S1 of Elishmereni et al. [8] is used and the remaining nominal parameter values are calculated from (15). The crucial point here is that this is only one of the infinite elements of the equivalence class (see (14)) of parameter solutions with respect to the output function. This means that, by fixing k_{02} to a different value, she/he can calculate a different corresponding solutions of k_{01} , k_{21} , and k_{12} . This second solution will predict the same time course of x_1 but a different time course of the unmeasured variable x_2 (Figure 3).

It is worth noting that, by constraining the parameters to be nonnegative, the above equations provide admissible values for a range of $k_{02} \in [0, 0.005]$, because of k_{01} becoming zero. Thus the definition of the equivalent classes can embed

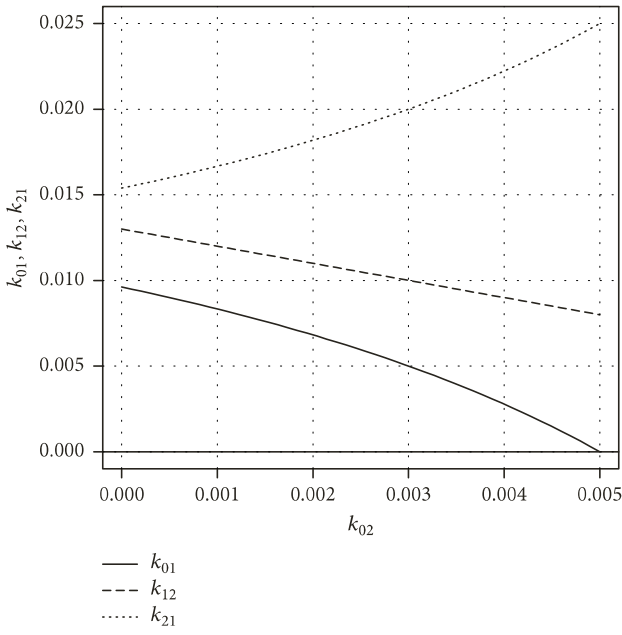


FIGURE 2: Equivalent parameterizations of k_{01} , k_{12} , and k_{21} as functions of k_{02} (see (15)).

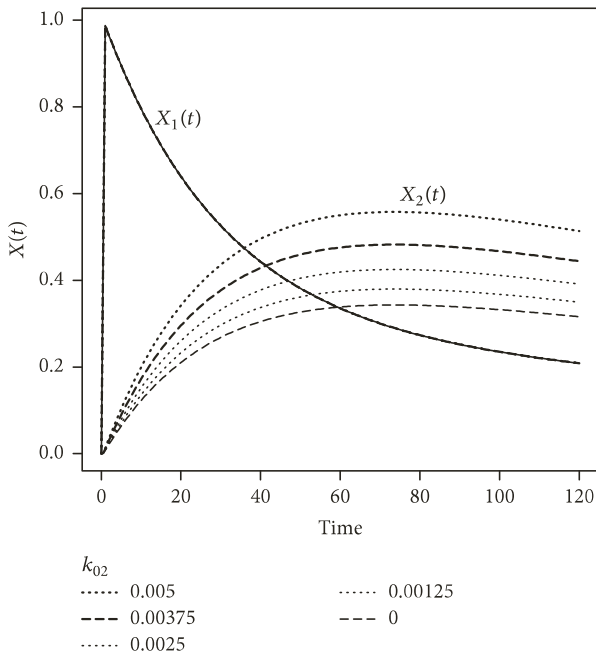


FIGURE 3: Time courses of model state trajectories $x_1(t)$ (measured) and $x_2(t)$.

constraints on parameters such as positivity and thus can allow, for the first time, to reject some inadmissible solutions. This is an important result which was not provided by the only structural identifiability analysis.

We can conclude that, for each (admissible) value of k_{02} and with the remaining parameters computed from the Gröbner basis as described above, the input-output behavior of the model is invariant; that is, measured outputs do not

change by moving the parameters on the surface $G(\theta, \theta^*) = 0$. This is shown in Figure 3, where the state trajectories for different admissible values of k_{02} are reported. As anticipated, the time course of the measured state variable, $x_1(t)$, does not change with parameterization but only that of the *hidden* state, $x_2(t)$.

5. A Mathematical Model for Evaluating Antitumor Activity of IL-21 and Potential Immunotherapy Treatments

We choose to show our results taking inspiration from a mathematical model for interleukin-21 (IL-21) immunotherapy based on the state-of-the-art biology of the system [8]. Originally the model was developed to study the antitumor effects of IL-21 on tumor eradication in the three different cancers of varying immunogenicity and growth dynamics [4]. We will use the model of the nonimmunogenic B16 melanoma as a benchmark model on which to apply our methodologies.

In particular, the model describes the underlying biological processes and its parameters were originally evaluated from experimental data in tumor-bearing mice treated with IL-21 via three different administration methods (the first drug application was associated with tumor mass; the other two were independent of tumor mass). This model has been successively modified and included in a more comprehensive PK-PD-disease model to predict relevant scenarios of IL-21 treatment following IC, SC, or IP administration in different cancer indications.

In this model, each parameter has a specific physical meaning and obviously different parameter numerical estimates from the experimental data lead to different conclusions. Given the clinical relevance of the study based on this model, it becomes of fundamental importance (1) to know if the estimated parameter value is unique or if there is a class of parameters which are equivalent with respect to the input/output and (2) to analytically calculate the class of equivalence.

Below we report the ODE equations of the model, where a logistic growth was assumed for the total tumor cells z :

$$\begin{aligned}
 \dot{n} &= r_1 n \left(1 - \frac{n(x_T(t) + q_1)}{(p_1 x_T(t) + p_2)} \right) \\
 \dot{c} &= r_2 c \left(1 - \frac{c}{(h_2(0) + \sigma m / (1 + m/D))} \right) \\
 \dot{m} &= a x_T(t) - \mu_2 m \\
 \dot{p} &= \frac{b_1 x_T(t)}{(b_2 + x_T(t))} - \mu_3 p \\
 \dot{z} &= r_3 z \left(1 - \frac{z}{K} \right) - k_1 p n z - k_2 p c z \\
 y_1(t) &= n(t) \\
 y_2(t) &= c(t) \\
 y_3(t) &= z(t),
 \end{aligned} \tag{16}$$

where $\{n, c, m, p, z\}$ are the five state variables; x_T is the known input; y_1, y_2 , and y_3 are the output equations; $\theta = \{a, b_1, b_2, D, h_2(0), K, k_1, k_2, \mu_2, \mu_3, p_1, p_2, q_1, r_1, r_2, r_3, \sigma\}$ is the unknown parameter vector.

A detailed derivation of the model can be found in [8].

Only for simplicity of presentation, we will assume the following:

(1) The input x_T is described by an exponential function simulating the PK model [8].

(2) A uniform sampling schedule is applied.

(3) The nominal parameter values are $\theta^* = \{0.57, 0.1, 0.1, 190, 0.0018, 1501.4, 0.376, 5.184, 0.014, 0.08, 0.01, 1.054, 0.54, 0.095, 0.26, 0.018, 0.0071\}$ given in Table S1 of Elishmereni et al. [8].

(4) Initial conditions of state variables are known.

The results of the study hold irrespective of the above hypotheses.

First, by applying the differential algebra algorithm described in Section 2.2, we calculate the Gröbner basis around the nominal parameter value θ^* and obtain the following results.

parameters $\{b_2, h_2(0), K, \mu_2, \mu_3, p_1, p_2, q_1, r_1, r_2, r_3\}$ are uniquely identifiable, while $\{a, b_1, D, k_1, k_2, \sigma\}$ are unidentifiable. The novelty of the result here is that the Gröbner basis provides the following analytic relations between unidentifiable parameters holding not in the whole complex space but around the parameter value of clinical interest:

$$\begin{aligned} b_1 &= \frac{47}{(1250k_1)} \\ k_2 &= \frac{648k_1}{47} \\ a &= \frac{3D}{1000} \\ \sigma &= \frac{1349}{(1000D)}. \end{aligned} \quad (17)$$

We know that all the values of parameters k_1, k_2, b_1 , satisfying the first two above equations, and a, D, σ , satisfying the last two, equivalently describe the output function of the model; thus, they cannot be distinguished by any sensitivity-based approach. Equations (17) indicate the number of degrees of freedom and provide the exact constraints to include into the model equations in order to reach global identifiability. In this case, it is easy to see from (17) that there are only two free parameters (k_1, D). By including (17) in the model equations (16), the four redundant parameters $\{b_1, k_2, a, \sigma\}$ are constrained as functions of the two free parameters.

Now we want to show that this result could be qualitatively determined by a sensitivity-based identifiability approach, possibly arriving to discover a correlation between the above six unidentifiable parameters.

As done in the previous simple example, see Section 4, in order to check practical identifiability of model of (16), we calculate the sensitivity matrix of the model and the SVD of the sensitivity matrix, as described in Section 2.3. The singular values, sorted in decreasing order, provide a ranking among

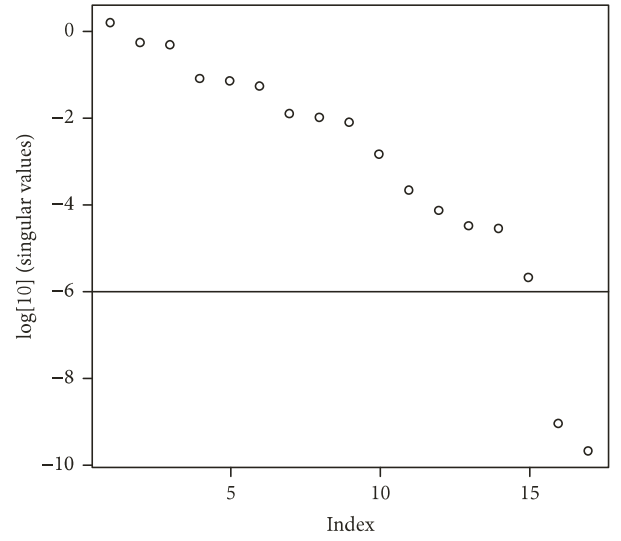


FIGURE 4: Singular values of the sensitivity matrix of model of (16) obtained with nominal parameter values and reported in decreasing order (the cutoff line at 10^{-6} was chosen arbitrarily).

the parameters with respect to their ability to affect the output function, which reveals the following:

- (i) On the upper side, the parameters that affect most the output function, thus providing most reliable estimates
- (ii) On the lower side, the presence of linear dependence among the output sensitivities, if the singular values of the sensitivity matrix are zero

In Figure 4, the singular values of the sensitivities matrix of model of (16) are reported in decreasing order. It is evident that the model is practically unidentifiable, since there are at least two singular values nearly zero. This means that there are at least two candidate solutions for markedly unidentifiable parameter combinations. We are interested in these smallest singular values and in their corresponding eigenvectors. We report in Table 1 the four last columns representing the eigenvectors corresponding to the smallest singular values (SV) in reverse order. The smallest SV reveals a relation between the three parameters a, D, σ , and the mathematical form of this relation is exactly given by (17) provided by the structural identifiability. Interesting is to observe that the values in the same column reveal dependence of these parameters also from $h_2(0)$, which was structurally identifiable. Thus we know that to reach its identifiability the investigator should modify the experimental setup, not the model structure.

The remaining eigenvectors reported in Table 1 show that there are parameters which are practically unidentifiable even if they were found to be structurally identifiable. In particular, by looking, for example, at the value corresponding to parameter “ K ” in the second eigenvector or to “ q_1 ” in the last eigenvector, it is evident that these two structurally identifiable parameters turn out to be practically unidentifiable. This is not surprising; in fact, the inability to practically estimate model parameters may be caused by a number of distinct reasons, such as (1) excessive noise in the measurements, (2) poor or very sparse sampling schedules, and (3) poorly designed

TABLE 1: The four smallest singular values of Figure 4 with their corresponding eigenvectors.

Par.	Singular values			
	$2.2 \cdot 10^{-10}$	$9.4 \cdot 10^{-10}$	$2.2 \cdot 10^{-6}$	$3 \cdot 10^{-5}$
	Right singular eigenvectors			
a^\dagger	0.525[‡]	−0.003	0	0
b_1	0	0	−0.085	−0.004
b_2	0	0.002	0.969	−0.003
D	0.526	−0.003	0	0
$h_2(0)$	0.414	−0.003	0	0
K	0.006	1.000	−0.002	0
k_1	0	0	−0.002	0
k_2	0	0	0.049	0.002
μ_2	0	0	0	0
μ_3	0	0	−0.068	−0.002
p_1	0	0	0	−0.04
p_2	0	0	0	−0.017
q_1	0	0	0	−0.998
r_1	0	0	0	−0.033
r_2	0	0	0	0
r_3	0	0	0.216	0.009
σ	−0.526	0.003	0	0

[†]Model parameters perturbations are expressed as linear combinations of rows elements. [‡]Table entries are in boldface if they are larger than about 5% and are 0 if they are smaller than 10^{-3} .

experiments, where measurement locations or inputs are insufficiently informative. However, if the model turns out to be practically unidentifiable, only by first checking structural identifiability is it possible to know for sure if the problem lays on an unwarranted model complexity or on the above reasons related to experimental data.

6. Conclusion

Mathematical ODE models employed in oncological studies are in general complex models. For this reason, very often they are unidentifiable; that is, they have an infinite number of parameter solutions. However, this problem is not always recognized and, by ignoring the fact that a parameter has more solutions, the investigator can arrive at totally erroneous conclusions. Model identifiability tests are essential to determine whether a model can be possibly inferred from the experimental data. In this paper, we propose a unified viewpoint of two different identifiability analysis techniques and motivate their joint use. The two methodologies, namely, *structural identifiability* and *practical identifiability*, are traditionally regarded as disjoint because they are based, in turn, on differential algebraic manipulations and on numerical simulation of systems equations.

Nevertheless, also the structural analysis can provide useful practical information if applied around a particular point of clinical interest belonging to the admissible parameter space.

In this paper, we propose utilizing first the structural identifiability test and successively, by taking advantage of

its analytic results, the sensitivity approach. For the first time, the joint implementation of these two identifiability methodologies allows the following:

- (1) Disentangling the various causes of nonidentifiability assessed with sensitivity-based approaches, actually providing some additional information helpful for experiment design (a priori) and for the interpretation of parameter estimation results (a posteriori)
- (2) Exactly knowing the analytic relations between the correlated parameters
- (3) Reducing the model's complexity by redefining an identifiable model equivalent to the original one but with a reduced number of unknown parameters
- (4) Making the parameter identification process from real data more rigorous and reliable

All the above findings would help the prognostic decision-making of the oncologist by simultaneously reducing the costs associated with clinical developments.

In principle, the whole proposed methodology does not require experimental data and thus it can be viewed as a tool for addressing the *experiment design* problem. Furthermore, it is implementable by using available differential algebra software together with statistical packages.

Finally, to show that the calculation of the analytic relations between unidentifiable parameters is an effective approach to discover the behavior of nonaccessible variables of clinical interest, we apply our methodology to a relevant benchmark model: a mathematical model to study the “in vivo” antitumor activity of interleukin-21 (IL-21) on tumor eradication in different cancers in mice.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] D. Jayachandran, A. E. Rundell, R. E. Hannemann, T. A. Vik, and D. Ramkrishna, “Optimal chemotherapy for leukemia: a model-based strategy for individualized treatment,” *PLoS ONE*, vol. 9, no. 10, Article ID e109623, 2014.
- [2] L. G. dePillis, A. Eladdadi, and A. E. Radunskaia, “Modeling cancer-immune responses to therapy,” *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 41, no. 5, pp. 461–478, 2014.
- [3] P. L. Bonate and A. B. Suttle, “Modeling tumor growth kinetics after treatment with pazopanib or placebo in patients with renal cell carcinoma,” *Cancer Chemotherapy and Pharmacology*, vol. 72, no. 1, pp. 231–240, 2013.
- [4] A. Cappuccio, M. Elishmereni, and Z. Agur, “Cancer immunotherapy by interleukin-21: potential treatment strategies evaluated in a mathematical model,” *Cancer Research*, vol. 66, no. 14, pp. 7293–7300, 2006.
- [5] J. A. Florian Jr., J. L. Eiseman, and R. S. Parker, “Nonlinear model predictive control for dosing daily anticancer agents using a novel saturating-rate cell-cycle model,” *Computers in Biology and Medicine*, vol. 38, no. 3, pp. 339–347, 2008.

- [6] A. Galante, K. Tamada, and D. Levy, "B7-H1 and a mathematical model for cytotoxic T cell and tumor cell interaction," *Bulletin of Mathematical Biology*, vol. 74, no. 1, pp. 91–102, 2012.
- [7] P. Gerlee, "The model muddle: In search of tumor growth laws," *Cancer Research*, vol. 73, no. 8, pp. 2407–2411, 2013.
- [8] M. Elishmereni, Y. Kheifetz, H. Søndergaard, R. V. Overgaard, and Z. Agur, "An integrated disease/pharmacokinetic/pharmacodynamic model suggests improved interleukin-21 regimens validated prospectively for mouse solid cancers," *PLoS Computational Biology*, vol. 7, no. 9, Article ID e1002206, 2011.
- [9] V. Raia, M. Schilling, M. Bohm et al., "Dynamic Mathematical Modeling of IL13-Induced Signaling in Hodgkin and Primary Mediastinal B-Cell Lymphoma Allows Prediction of Therapeutic Targets," *Cancer Research*, vol. 71, no. 3, pp. 693–704, 2011.
- [10] B. Ribba, G. Kaloshi, M. Peyre et al., "A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy," *Clinical Cancer Research*, vol. 18, no. 18, pp. 5071–5080, 2012.
- [11] N. Terranova, M. Germani, F. Del Bene, and P. Magni, "A predictive pharmacokinetic-pharmacodynamic model of tumor growth kinetics in xenograft mice after administration of anti-cancer agents given in combination," *Cancer Chemotherapy and Pharmacology*, vol. 72, no. 2, pp. 471–482, 2013.
- [12] Z. Agur, "From the evolution of toxin resistance to virtual clinical trials: the role of mathematical models in oncology," *Future Oncology*, vol. 6, no. 6, pp. 917–927, 2010.
- [13] L. Ljung, *System Identification - Theory For the User*, PTR, Prentice Hall, Upper Saddle River, N.J., USA, 2nd edition, 1999.
- [14] S. Audoly, G. Bellu, L. D'angiò, M. P. Saccomani, and C. Cobelli, "Global identifiability of nonlinear models of biological systems," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 55–65, 2001.
- [15] C. Cobelli and M. P. Saccomani, "Unappreciation of a priori identifiability in software packages causes ambiguities in numerical estimates," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 258, no. 6, pp. E1058–E1059, 1990.
- [16] M. P. Saccomani, S. Audoly, and L. D'Angiò, "Parameter identifiability of nonlinear systems: The role of initial conditions," *Automatica*, vol. 39, no. 4, pp. 619–632, 2003.
- [17] N. Meshkat, C. Er-zhen Kuo, and J. DiStefano, "On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: a novel web implementation," *PLoS ONE*, vol. 9, no. 10, Article ID e110261, 2014.
- [18] K. Thomaseth, J. J. M. Batzel, R. Bachar, and R. Furlan, "Parameter estimation of a model for baroreflex control of unstressed volume," in *Mathematical Modeling and Validation in Physiology*, J. J. Batzel, M. Bachar, and F. Kappel, Eds., pp. 215–246, Springer-Verlag, Berlin, Germany, 2013.
- [19] G. Bellu, M. P. Saccomani, S. Audoly, and L. D'Angiò, "DAISY: a new software tool to test global identifiability of biological and physiological systems," *Computer Methods and Programs in Biomedicine*, vol. 88, no. 1, pp. 52–61, 2007.
- [20] B. Buchberger, "An algorithmical criterion for the solvability of algebraic system of equation," *Aequationes Mathematicae*, vol. volume 4, 3, pp. 45–50, 1988.
- [21] M. Rodriguez-Fernandez, J. R. Banga, and I. Doyle, "Novel global sensitivity analysis methodology accounting for the crucial role of the distribution of input parameters: application to systems biology models," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 10, pp. 1082–1102, 2012.
- [22] M. Rodriguez-Fernandez, M. Rehberg, A. Kremling, and J. R. Banga, "Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems," *BMC Systems Biology*, vol. 7, article no. 76, 2013.
- [23] A. Raue, C. Kreutz, T. Maiwald et al., "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, 2009.
- [24] E. Balsa-Canto and J. R. Banga, "AMIGO, a toolbox for advanced model identification in systems biology using global optimization," *Bioinformatics*, vol. 27, no. 16, Article ID btr370, pp. 2311–2313, 2011.
- [25] S. Hoops, R. Gauges, C. Lee et al., "COPASI—a complex pathway simulator," *Bioinformatics*, vol. 22, no. 24, pp. 3067–3074, 2006.
- [26] S. Beal, L. B. Sheiner, A. Boeckmann, and R. J. Bauer, "NON-MEM's User's Guides," in *ICON Development Solutions*, Ellicott City, MD, USA, 2009.
- [27] D. Z. D'Argenio, A. Schumitzky, and X. Wang, "ADAPT 5 Users Guide: pharmacokinetic/pharmacodynamic systems analysis software," in *Biomedical Simulations Resource*, Los Angeles, 2009.
- [28] S. Vajda, P. Valko, and T. Turányi, "Principal component analysis of kinetic models," *International Journal of Chemical Kinetics*, vol. 17, no. 1, pp. 55–81, 1985.
- [29] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017, <http://www.R-project.org>.
- [30] E. Anderson, Z. Bai, C. Bischof et al., "LAPACK Users' Guide," in *Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, 3th edition, 1999.

