Local Identifiability Analysis of NonLinear ODE Models: How to Determine All Candidate Solutions

Karl Thomaseth [*] Maria Pia Saccomani [**]

[*] National Research Council (IEIIT-CNR), c/o DEI, University of Padova, Italy (e-mail: karl.thomaseth@cnr.it).
[**] Department of Information Engineering, University of Padova, Italy (e-mail: pia@dei.unipd.it).

Abstract: Identifiability analysis aims at answering the theoretical question whether the inverse problem is solved, uniquely, by a particular value of the free parameters, or if there is a finite or infinite number of parameter vectors that generate identical output trajectories. Multiple solutions of locally identifiable parameters imply different time courses of unmeasured variables, and arbitrarily chosen solutions can lead to misinterpretations and to erroneous conclusions. We present theoretical background and applications to locally identifiable ODE models described by rational functions, showing that structural identifiability analysis reinforces the practical identifiability approach. In a first example using a three compartment model, we discuss the algorithm that allows to find all the equivalent parameter solutions. In the second example on HIV dynamics, we show how two solutions can provide two major different scenarios regarding the prediction of unobservable variables, which may lead to different treatment strategies. In conclusion, for locally identifiable models we propose an algorithmic approach which, for the first time, allows the calculation of *all* numerical model solutions, the possible rejection of non admissible parameters, and the simulation of the trajectories of unobservable variables.

1. INTRODUCTION

Nonlinear ordinary differential equations are widely employed for the interpretation of biological systems dynamics observed during controlled input/output (I/O) perturbation experiments. In this paper we address the problem of determining whether the model has a unique, or a finite number, or infinite many parameter solutions that generate identical output trajectories.

The most desirable situation is that of *unique* structural identifiability, which ensures that the inverse problem is theoretically well-posed. On the opposite side, structural *unidentifiabiliy* is usually unacceptable because it does not provide useful information on the system under study. Finally, the "intermediate" case of multiple solutions is not easily recognizable in practice because the inverse problem is well-defined locally, in the neighborhood of isolated parameters taken from a finite pool of candidate solutions.

Multiple solutions of non-uniquely identifiable parameters are equivalent from an I/O point of view but are associated with a 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, i.e., that of providing a means to infer on *unobservable* quantities and time-varying phenomena.

Given the relevance of this issue in biological modeling and the feasibility through analytic tools, this study deals with the generally understated situation of *local* identifiability. As will be shown, local identifiability with multiple solutions can arise even with simple model structures, not

being caused by symmetries due to permutations of the state variables. Here, we propose a unified viewpoint of different identifiability analysis techniques and motivate their joint use for studying local identifiability.

The two methodologies, namely *structural identifiability* and *practical identifiability* (Chis *et al.*, 2011; Ljung and Glad, 1994; Saccomani *et al.*, 2004; Stigter and Molenaar, 2015; Raue *et al.*, 2014; Rodriguez-Fernandez *et al.*, 2013; Thomaseth *et al.*, 2013; Saccomani and Thomaseth, 2016; Janzén *et al.*, 2016), are traditionally regarded as disjoint, as being applicable either a priori or a posteriori and based on analytic calculations (we will consider the differential algebraic based approach) and on numerical simulation of systems equations, respectively. Essentially both approaches do not require experimental data, but only the former can be tested without assuming prior knowledge on parameter values, while the method based on sensitivity analysis requires nominal parameter values for numerical simulation. Unfortunately, structural identifiability analysis is applicable only to particular model classes and may fail to provide answers with overly complex structures.

In the biomedical literature the problem of multiple local solutions (equivalently describing the I/O behavior of the system), has been generally understated. Almost all the traditional identification methods, based on minimization of a cost function, allow to study only one solution by neglecting the others that remain *hidden*. Methods such as the multi-start optimization as well as profile likelihood allow for identification of more than one parameter solution.

They partially solve the local identifiability problem but they still do not guarantee to calculate *all* the equivalent model solutions.

In this paper we show, for the first time, how results based on structural identifiability analysis can provide practically useful information for dissecting the issue of multiple local solutions. We also show that only one solution, estimated with an optimization algorithm, is necessary to numerically calculate all the others. Thus in the following we suggest how to:

(1) numerically calculate *all* the solutions of a locally identifiable model;
(2) simulate all the possible different behaviours of the unobservable variables corresponding to each of the previous solutions;
(3) possibly reject some of them based on existing constraints on parameters. In some cases, by rejecting non admissible estimates, one can arrive at a globally identifiable model.
(4) provide reliable initial conditions for statistical methods such as multi-start searches and profile likelihood.

Thus we show that knowledge of the numerical values of multiple solutions allows to solve the identification problem in a more rigorous way. The relevance of applying this identifiability methodology is illustrated by an example with a three compartment model.

Being based on analytic calculations, this approach is limited by the model dimension. However, we successfully applied it to several models published in the recent biological literature. A case study describing a classical HIV model is also reported. The model has two different solutions describing equivalently the I/O experimental data, but showing two completely different scenarios with regard to the unobservable variables.

2. MATHEMATICAL BACKGROUND

This section provides the definitions that are necessary to set the notations used in the paper. Consider a nonlinear dynamic system described in state space form as

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \tag{1}$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \tag{2}$$

with state $\mathbf{x}(t) \in \mathbb{R}^n$; input $\mathbf{u}(t) \in \mathbb{R}^q$ ranging on some vector space, $\mathcal{U}$, of piecewise smooth (infinitely differentiable) functions; output $\mathbf{y}(t) \in \mathbb{R}^m$; and constant unknown parameter vector $\boldsymbol{\theta}$ belonging to some admissible subset $\Theta \subseteq \mathbb{C}^p$. $\mathbf{f}$ and $\mathbf{g}$ are rational functions. Yet nonlinearities as exponential or logarithmic can be dealt with as illustrated in (DAngiò *et al.* (2009)). The class of admissible parameters is assumed to be $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^p | \boldsymbol{\theta} > \mathbf{0}\}$. It is further assumed that at least one admissible parameter exists, which will be indicated by $\boldsymbol{\theta}^*$, which does not represent a *true*, yet unknown value to be estimated, but rather a feasible reference value for model parameters obtained, e.g., from fitting the model to pilot experimental data.

Whenever initial conditions are specified, the equation $\mathbf{x}(0) = \mathbf{x}_0$ is added to the system. We also assume that there is no feedback, so that $\mathbf{u}$ is a free variable not depending on $\mathbf{y}$. If equality constraints on the parameters are present, these can be considered in the model by adding the polynomial equation:

$$h(\boldsymbol{\theta}) = 0 . \tag{3}$$

The I/O map of system (1, 2) with initial state $\mathbf{x}_0$ will be denoted with

$$\mathbf{y} = \psi_{\mathbf{x}_0}(\boldsymbol{\theta}, \mathbf{u}) . \tag{4}$$

This equation has at least one solution if evaluated in $\boldsymbol{\theta}^*$.

In the following we focus on *local identifiability* (the reader is referred to (Saccomani *et al.*, 2004) for other definitions of identifiability).

Definition 1. The system (1, 2) is *locally identifiable at* $\boldsymbol{\theta}^* \in \Theta$ *from I/O data*, if there exists (at least) one input function $\mathbf{u}$ and a neighborhood $U_{\boldsymbol{\theta}^*}$ of $\boldsymbol{\theta}^*$, such

$$\psi_{\mathbf{x}_0}(\boldsymbol{\theta}, \mathbf{u}) \equiv \psi_{\mathbf{x}_0}(\boldsymbol{\theta}^*, \mathbf{u}) \tag{5}$$

has a unique solution $\boldsymbol{\theta} \in U_{\boldsymbol{\theta}^*}$ for all $\mathbf{x}_0 \in X \subseteq \mathbb{R}^n$.

This definition deals thus with multiple solutions of Equation (5), provided that they are isolated points in $\mathbb{C}$. It is important to stress that multiple solutions are not rare, even in unsophisticated models, as will be shown in the case study in Section 4.

The analysis of identifiability for a particular system (1, 2), can be viewed as the study of the properties of the equivalence classes induced by the equivalence relation (5). This defines the indistinguishability between the elements of the algebraic variety

$$\mathcal{V}(\boldsymbol{\theta}^*) = \{\boldsymbol{\theta}^* \sim \boldsymbol{\theta}_i, i = 1, \ldots, n_{\boldsymbol{\theta}} \,|\, \psi_{\mathbf{x}_0}(\boldsymbol{\theta}, \mathbf{u}) \equiv \psi_{\mathbf{x}_0}(\boldsymbol{\theta}^*, \mathbf{u})\} \tag{6}$$

where $\sim$ means equivalent, and $n_{\boldsymbol{\theta}}$ is the number of parameter solutions of system (1, 2) in the whole complex space. In particular, $n_{\boldsymbol{\theta}} = 1$ in the globally identifiable; $1 < n_{\boldsymbol{\theta}} < \infty$ in the locally identifiable; or $n_{\boldsymbol{\theta}} = \infty$ in the unidentifiable system case.

We will apply a structural identifiability method implemented in the software DAISY (Differential Algebra Identifiability of Systems) (Bellu *et al.*, 2007). The reader is referred to (Ritt, 1950) for a formal treatment of differential algebra, and to (Bellu *et al.*, 2007; Saccomani *et al.*, 2004) for a detailed explanation of the theory behind the software tool.

This differential algebra method provides a final algebraic nonlinear system in the unknown $\boldsymbol{\theta}$. This is solved by applying the Buchberger algorithm which calculates its Gröbner basis. The Gröbner basis provides exactly the algebraic variety represented by Equation (6). It allows to calculate the $n_{\boldsymbol{\theta}}$ solutions of system (1, 2) in the complex space. In particular, if Equation (5) has $n_{\boldsymbol{\theta}}$ finitely multiple solutions, the Gröbner basis takes the form:

$$\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \{\{\boldsymbol{\theta} - \boldsymbol{\theta}_1\}, \ldots, \{\boldsymbol{\theta} - \boldsymbol{\theta}_{n_{\boldsymbol{\theta}}}\}\} \tag{7}$$

If $n_{\boldsymbol{\theta}} = 1$, model (1, 2) is globally identifiable.

Note that if the k-th components of $\boldsymbol{\theta}_i$ are equal for all $i = 1, ..., n_{\boldsymbol{\theta}}$, the k-th parameter component is globally identifiable (see, for example, parameter k_{21} of the locally identifiable model (8)).

In this case of local identifiability, the basis provides the following relevant information:

– the cardinality $n_{\boldsymbol{\theta}}$ of the equivalence class of the parameter solutions, i.e. the finite number of solutions equivalently describing the experimental I/O data,

– a means for analytically calculating all other solutions, that would otherwise remain "hidden" (see Section 3).

3. A COMBINED APPROACH TO CALCULATE ALL SOLUTIONS OF LOCALLY IDENTIFIABLE MODELS

Prior knowledge of global or local uniqueness of parameter estimates is useful for assessing whether the experi-

mental design is adequate for a hypothesized model and whether the parameter estimation problem is well-posed.

It is worth noting that the solutions of structural identifiability analysis, being roots of Equation (6) are sought in the complex plane $\mathbb{C}$, can be either real positive; real negative; or complex conjugate. Consequently, when a model is globally identifiable, the unique solution corresponds to a real one, but when the model is locally identifiable, the solutions belong to the complex space, and not only to the admissible parameter space Θ.

In the following it is shown how practical identifiability tests based on model output sensitivities, can be integrated with the information provided by structural identifiability analysis. In particular, we want to show how the specific value simulated or estimated with the practical approach, together with the class of equivalence of Equation (6), allows to calculate all the other candidate solutions and, possibly, by checking the known physical constraints on these solutions, to accept or reject some of them. In general, one can arrive at diminishing the number of plausible solutions but still having more than one solution left. For this purpose, the methodological steps are:

(1) estimate a numerical value $\boldsymbol{\theta}^*$ with a practical identification method;
(2) go back to Equation (6) and substitute the just estimated value in order to provide its known coefficient;
(3) by solving the obtained equations, calculate all the remaining solutions;
(4) check the known physical constraints on these solutions, i.e. reality and positivity, in order to reject some of them, possibly arriving at a unique solution.

The identifiability analysis is extended here by introducing constraints on parameters to define admissible regions, typically positivity or interval constraints. In particular, it is assumed that there exists one reference (admissible) parameter, $\boldsymbol{\theta}^*$ in Θ. In this set, the *a priori* locally identifiable model may well be reconsidered as uniquely identifiable *a posteriori* if all the alternative solutions, equivalently describing the I-O data, were to be discarded because not admissible.

To appreciate the relevance of using both the structural and practical approaches in local analysis, we consider a simple locally identifiable model, that provides structurally three solutions, i.e. identical model predictions for three different values of parameter vectors. In this case numerical optimisation algorithms could lead to multiple different minima in the whole parameter space, having all the same I/O prediction and thus producing the same value of the cost function. The above methods cannot, in principle, guarantee to find all of the equivalent solutions. To do that, in fact, they should start from an infinite number of points in the admissible parameter space. Usually the investigator considers one or at most a subset of all the possible solutions, not even being aware of the existence of the other equivalent ones. By neglecting these, one may arrive at erroneous model interpretations providing unreliable results. This can be crucial in physiological models, where a pathological state can be revealed by a specific parameter value threshold.

Conversely, if an optimization algorithm is preceded by a structural identifiability test, one becomes aware of the presence of multiple solutions that can be found by using the Gröbner basis, Eq.(7). In particular, by starting from

an estimated global minimum, the Gröbner basis allows to analytically calculate all the equivalent model solutions. It is worth noting that the differential algebra method is placed in the analytic mathematical setting, not in the statistical one. Thus its results can provide reliable initial conditions for the statistical methods, such as multi-start searches and profile likelihood, where issues related to random noise are considered.

Note that if a model is structurally identifiable, it may still turn out to be practically non-identifiable. In this case the inability to unequivocally estimate model parameters may be caused by a number of distinct reasons, among which: 1) excessive noise in the measurements, 2) very sparse sampling schedules, 3) poorly designed 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 it is possible to know for sure if the problem lays on an unwarranted model complexity or on the above reasons related to experimental data.

To show the practical consequences of the joint use of structural and practical identifiability for local identifiable models, a simple example is illustrated in the following.

3.1 Calculation of all three solutions of a simple locally identifiable model

We consider as a benchmark example the single-input single-output three compartment model depicted in Figure 1. It is described by the following equations:

$$
\begin{aligned}
\dot{x}_1 &= -(k_{01} + k_{21})\,x_1 + k_{13}\,x_3 + u(t) \\
\dot{x}_2 &= k_{21}\,x_1 - k_{32}\,x_2 + k_{23}\,x_3 \\
\dot{x}_3 &= k_{32}\,x_2 - (k_{13} + k_{23})\,x_3 \\
y(t) &= x_2(t)
\end{aligned}
\tag{8}
$$

where $\boldsymbol{\theta} = [k_{01}, k_{13}, k_{21}, k_{23}, k_{32}]$ is the unknown parameter vector, $\mathbf{x} = [x_1, x_2, x_3]$ the states vector, u the input and y the measured output. The initial conditions are supposed to be unknown. First we count the parameter
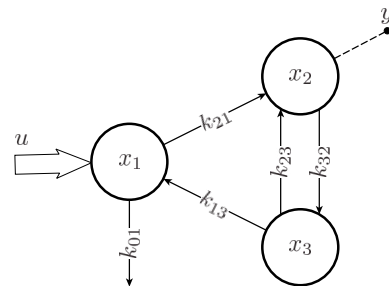

Fig. 1. A locally identifiable model. (Norton, 1982)

solutions by applying the differential algebra method implemented by DAISY and we discover that the model has three solutions (the only globally identifiable parameter is k_{21}) all being equivalently able to describe the output function of the model. More important the fact that in correspondence of each of these three solutions, the behavior of all the unobservable variables is different.

This will be the starting point to provide a method able to numerically calculate *all* the model solutions describing the I/O experimental data. To find all the possible solutions, by starting from the estimated one, is the relevant finding of this work.

To show this we move now to the practical context and provide the I/O simulated data by assuming two different nominal values $\boldsymbol{\theta}^*$ in order to discuss two possible different scenarios. We enter first $\boldsymbol{\theta}^* = [0.02374, 0.00181, 0.01331,$ $0.03089, 0.01729]$ in the differential algorithm. In particular, after having calculated the I-O map (4) of the model with the differential algebra algorithm, its coefficients, are extracted. These are known functions of the unknown parameter vector $\boldsymbol{\theta}$. We calculate the so-called *exhaustive summary* (Saccomani *et al.* (2004)) of the model by equating these functions of $\boldsymbol{\theta}$ to a known coefficient obtained by evaluating the same functions in $\boldsymbol{\theta}^*$. The obtained system of algebraic equations is:

$$\begin{cases} 10^5 \cdot k_{01} + 10^5 \cdot k_{32} = 4103 \\ 5244385 \cdot 10^3 \cdot k_{13} - 154 \cdot 10^9 \cdot k_{32}^2 + 4406795 \cdot 10^3 \cdot k_{32} = 39648431, \\ 10^5 \cdot k_{21} = 1331, \\ 1048877 \cdot 10^4 \cdot k_{23} + 308 \cdot 10^9 \cdot k_{32}^2 - 881359 \cdot 10^4 \cdot k_{32} = 263685917, \\ 5 \cdot 10^{14} \cdot k_{32}^3 - 31335 \cdot 10^9 \cdot k_{32}^2 + 4971978 \cdot 10^5 \cdot k_{32} = 1813508333 \end{cases} \tag{9}$$

Identifiability is then checked by solving, for the unknown parameters $\boldsymbol{\theta}$, this system. To do that we apply the Buchberger algorithm which calculates the Groebner basis of the system and provides the other two equivalent numerical solutions, reported in Table 1. In the following we will denote $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}_1$ and the two remaining solutions as $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$. This shows that the numerical solution $\boldsymbol{\theta}^*$ estimated by

Table 1. Admissible solutions for the first randomized parameter vector

	$\boldsymbol{\theta}_1$	$\boldsymbol{\theta}_2$	$\boldsymbol{\theta}_3$
k_{01}	0.02374	0.03581	0.0008737
k_{13}	0.00181	0.003971	0.02117
$k_{21}{}^a$	0.01331	0.01331	0.01331
k_{23}	0.03089	0.02873	0.01153
k_{32}	0.01729	0.005225	0.04016

^{a} globally identifiable parameter.

using an optimization algorithm (the plausible solution) has nothing more or special with respect to those that remain hidden. The relevant fact is that $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ predict the output function of the model exactly as the selected $\boldsymbol{\theta}^*$ does, as shown in Figure 2, and, in this case, they are both belonging to the admissible parameter space Θ (suppose the real and positive space). In addition one can appreciate from Figure 2 that, while the evolution of the measured variable x_2 is invariant, the trajectories of the unobservable variables x_1 and x_3 markedly change for each solution, but maintain a similar shape, e.g. $x_1(t)$ decays exponentially, and $x_3(t)$ increases slowly with time. Thus, the same measured curve can be equivalently described by the three different parameter sets, and all of them are possible candidates to be the true one.

Since in biological and biomedical studies, the goal of model identification is usually to estimate variables that are not directly measurable but whose distinct values can discriminate a pathological from a normal state, neglecting the hidden solutions could lead to completely erroneous conclusions. Note that, in this case, the only way to reject some solutions is to check if the behaviours of the unobserved variables are physically unacceptable. If not, the corresponding parameter value can be rejected.

Table 1 shows that all three solutions belong to the admissible parameter space Θ and allow to conclude that the three parameterisations are equally plausible. However, in general, we could obtain complex and/or negative
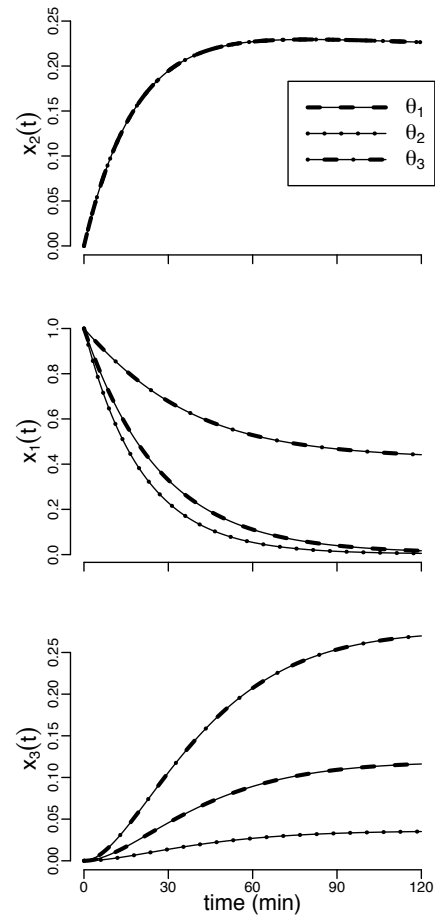

Fig. 2. State trajectories of the three compartment model of Figure 1 determined for the three locally identifiable parameterization (see Table 1). Top panel: the identical model output $y(t) = x_2(t)$. Middle panel: the different time-courses of the unobserved variable $x_1(t)$. Bottom panel: the different time-courses of the unobserved variable $x_3(t)$.

solutions. If this is the case, only at this practical stage we are allowed to reject the non admissible solutions possibly arriving at a globally identifiable model.

An example of this situation can be viewed by using the following second randomized nominal parameter vector $\boldsymbol{\theta}^* = [0.02324, 0.001834, 0.1202, 0.03072, 0.01632]$. By following the same line of reasoning as before, we analytically calculate the other two equivalent solutions, reported in Table 2, with the same notation adopted in Table 1. It is

Table 2. Non admissible solutions for the second randomized parameter vector

	$\boldsymbol{\theta}_1$	$\boldsymbol{\theta}_2$	$\boldsymbol{\theta}_3$
k_{01}	0.02324	−0.000422	−0.07089
k_{13}	0.001834	−0.04121	−8.88·10^{-5}
$k_{21}{}^a$	0.1202	0.1202	0.1202
k_{23}	0.03072	0.07377	0.03264
k_{32}	0.01632	0.03998	0.1104

^{a} globally identifiable parameter.

easy to see that this additional two solutions do not belong to the admissible parameter space. This is a favourable situation in which additional solutions can be rejected, showing, in practice, global identifiability of the model.

4. CASE STUDY. A LOCALLY IDENTIFIABLE HIV MODEL WITH TWO SOLUTIONS

We analyze a nonlinear model proposed for the study of HIV virus dynamics (Perelson *et al.*, 1993) and depicted in Figure 3. The model examines the interaction of HIV with $CD4^+$ T cells and is described by the following polynomial nonlinear ODEs:

$$\begin{cases} \dot{T}_c(t) = s - \mu_T T_c + rT_c(1 - (T_c + T_1 + T_2)/T_{\max}) - \\ \qquad\quad k_1 V T_c \\ \dot{T}_1(t) = k_1 V T_c - \mu_T T_1 - k_2 T_1 \\ \dot{T}_2(t) = k_2 T_1 - \mu_b T_2 \\ \dot{V}(t) = N\mu_b T_2 - k_1 V T_c - \mu_v V \\ y_1(t) = T_c(t) \\ y_2(t) = V(t) \end{cases}$$

(10)

where T_c is the population size of uninfected $CD4^+$ cells; T_1 of latently infected cells; T_2 of productively (or actively) infected cells; and V of free HIV virus particles. $\boldsymbol{\theta} = [s, r, T_{\max}, \mu_T, \mu_b, \mu_v, k_1, k_2, N]$ is the unknown parameter vector, and y_1 and y_2 are measured outputs in blood. One of the relevant goals of the model is that of predicting, from the I/O experimental data, the time-course of the unmeasured state variables, T_1 and T_2. By applying the
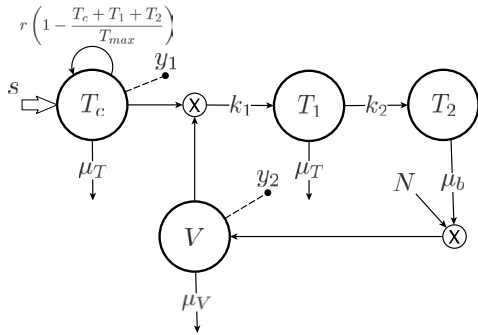

Fig. 3. A model of the HIV virus dynamics.

above described differential algebra method we show that parameters k_1, μ_T, μ_v, r, s and T_{max} are globally identifiable, while the remaining ones have two different solutions. Note that by applying a standard optimization technique to numerically estimate the model parameter vector, only one of these two existing solutions (indicated with $\boldsymbol{\theta}^*$) is obtained without any hint on the existence of the second one.

By joining the two structural and practical identifiability approaches we are able to analytically calculate the second solution, say $\boldsymbol{\theta}_2$. In fact, as illustrated in the previous example, we can calculate the known coefficients of Equation (6) as functions of the estimated $\boldsymbol{\theta}^*$ and solve the provided algebraic nonlinear system in the unknowns $s, r, T_{\max}, \mu_T, \mu_b, \mu_v, k_1, k_2, N$. The two solutions $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}_2$ of the calculated Gröbner basis are reported in Table 3. Note that only μ_b, k_2 and N have two solutions while the remaining parameters are globally identifiable. The relevant fact is that these two solutions are equivalent in the sense that they equivalently describe the I/O data but they are different even by an order of magnitude. Not surprisingly, they lead to two different predictions of the two unmeasured state variables T_1 and T_2. In particular

Table 3. The two solutions of the HIV model

Parameter	Units	$\boldsymbol{\theta}^*$	$\boldsymbol{\theta}_2$
s	$(\text{day}^{-1}\text{mm}^{-3})$	10	10
r	(day^{-1})	0.03	0.03
T_{max}	(mm^{-3})	1500	1500
μ_T	(day^{-1})	0.02	0.02
μ_b	(day^{-1})	0.24	0.023
μ_v	(day^{-1})	2.4	2.4
k_1	$(\text{mm}^3\ \text{day}^{-1})$	$2.4\cdot10^{-5}$	$2.4\cdot10^{-5}$
k_2	(day^{-1})	0.003	0.22
N		1400	199.21

we refer to parameters μ_b, k_2 and N which have a central role in the possible interpretations of the model and of its results. In fact, by considering $\boldsymbol{\theta}^*$, one can see that the model predicts a very high viral cell production N and a very low conversion rate k_2 of latent T_1 cells to infected T_2 cells, and a high rate constant μ_b. Conversely, $\boldsymbol{\theta}_2$ provides a much smaller value for N together with a k_2 two orders of magnitude larger than before, and μ_b one order of magnitude smaller than before. This means that the same time-course of V and T_c are due to both a high value of k_2 and a small value of μ_b and of N, or viceversa. Furthermore, by looking at Figure 4, one can easily realize that not only these predictions are quite different, but the difference appears only after two years. It is interesting
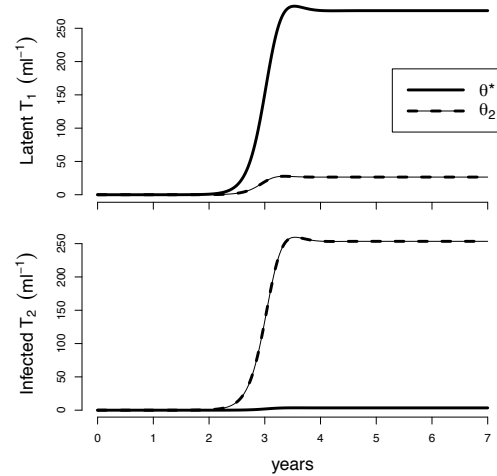

Fig. 4. Predicted trajectories of the unobservable variables representing cell concentrations of latent T_1 (upper panel) and infected T_2 (bottom panel) cells.

to observe that the product of all rate constants that influence the closed loop gain, namely $k_2 \cdot \mu_b \cdot N$, must be invariant for both solutions $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}_2$. In this case the above product is equal to 1.008 corresponding to a fractional daily growth rate of 0.8% (day^{-1}).

Incidentally, in the original paper Perelson et al. (1993), the authors state that some features did not match some observations. Very interesting is the observation "*the number of latently infected cells grows to unrealistically high levels, ...*". This may induce to think that the "hidden" solution $\boldsymbol{\theta}_2$ could possibly be the more realistic one. From our joint identifiability method, we know exactly the numerical value of both the "candidates" $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}_2$. This knowledge completes the results obtained in (Perelson *et al.*, 1993) and would allow a correct discussion on them. In principle, in fact, the two solutions are equivalent with respect to the

description of the I/O experimental data, thus the model predictions provided by $\boldsymbol{\theta}_2$ should be discussed on the same terms as done for those provided by $\boldsymbol{\theta}^*$. Maybe $\boldsymbol{\theta}_2$ should be rejected in favour of $\boldsymbol{\theta}^*$, maybe it could better match the model independent known clinical data, providing the correct solution and avoiding the introduction of assumptions made only to correct the model predictions.

In this case study, it is interesting to observe that the "hidden" solution $\boldsymbol{\theta}_2$ predicts: 1. a number of latently infected cells, appearing only after two years, very different from that predicted by using $\boldsymbol{\theta}^*$, 2. a number of free virus produced by lysing a $CD4^+$ T cell about seven times smaller than that predicted by using $\boldsymbol{\theta}^*$, which could be a more plausible result. Ignoring this possible solution can lead the physician to possibly erroneous decisions given that depletion values are used in a clinical setting as indicators of the disease stage.

5. CONCLUSIONS

In this paper we dealt with local identifiability analysis of biological models, with the aim of determining whether given I/O experimental data can be equally well predicted by a particular model but with different sets of parameter values. This issue becomes a major problem whenever model-based predictions are used as surrogate measurements of unaccessible/unobservable model quantities. The presence of multiple parameter solutions can in fact be easily overlooked, as parameter estimation algorithms normally converge to one of the possible equivalent solutions.

We propose here to jointly use two different identifiability analysis techniques, that are traditionally regarded as independent, namely *structural identifiability* and *practical identifiability*, being based, in turn, on differential algebraic manipulations and on numerical simulation of models and sensitivity equations. We use first the structural identifiability tool, not only to count the parameter solutions but also to determine the analytic expressions of their equivalence classes with respect to the description of the I/O experimental data. Successively, by starting from a numerical parameter solution estimated with a global optimization algorithm, e.g. multi-start searches, these equivalence classes allow to analytically calculate all the equivalent numerical parameter solutions, i.e. that equivalently describe the I/O experimental data. It is worth noting that these solutions, being calculated in a noise-free hypothesis, can provide good initial conditions for statistical methods, e.g. multi-start searches and profile likelihood.

Finally, we are able to reject the numerical solutions not belonging to the admissible parameter space, possibly arriving *a posteriori* at a globally identifiable model.

The above methodology opens new perspectives in theoretical and practical identifiability analysis with important practical implications, as illustrated by applying this joint methodology to a benchmark biological model.

In order to show the relevance of these ideas in biological/biomedical modeling, we applied them to a model of HIV virus dynamics by showing that, without calculating all the equivalent parameter solutions, there is a risk of arriving to ambiguous conclusions.
We can conclude that the knowledge of each of the finite number of parameter solutions of a locally identifiable model is essential to provide a complete picture for a correct interpretation of results.

REFERENCES

Bellu, G. Saccomani, M.P. Audoly, S. D'Angiò, L. (2007) DAISY: A new software tool to test global identifiability of biological and physiological systems. *Comp. Meth. Prog. Biom.*, **88**, 52-61.

Chis, O.T, Banga, J.R., Balso-Canto, E. (2011) Structural identiability of systems biology models: a critical comparison of methods. *PlosOne*, **6** (11), e27755. doi:10.1371/journal.pone.0027755

DAngiò, L ., Saccomani, M.P., Audoly, S., Bellu, G. (2009) Identifiability of nonaccessible nonlinear systems. In Bru, R., Romero-Vivò, S., (eds),*Positive Systems, Lecture Notes in Control and Information Sciences*, Springer-Verlag, Berlin, 389:269-277.

Janzén, D.L.I., Bergenholm, L., Jirstrand, M., Parkinson, J., Yates, J., Evan, N.D., Chappell, M.J. (2016) Parameter identifiability of fundamental pharmacodynamic models. *Front. Physiol.*, **7**, 590. doi:10.3389/fphys.2016.00590

Ljung, L., Glad, S.T. (1994). On global identifiability for arbitrary model parameterizations *Automatica*, **30**, 2, 265-276.

Miao, H., Xia, X., Perelson, A.S. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM review*, **53**, 1, 3-39.

Norton, J.P. (1982) An Investigation of the Sources of Nonuniqueness in Deterministic Identifiability. *Math. Biosci.*, **60**, 89-108.

Perelson, A.S., Kirschner, D.E., DeBoer, R. (1993) Dynamics of HIV Infection of CD4$^+$ T cells. *Math. Biosci.*, **114**, 81-125.

Raue, A., Karlsson, J., Saccomani, M.P., Jirstrand, M., Timmer, J. (2014) Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, **30**, 10, 1440-1448.

Ritt, J.F. (1950) *Differential Algebra*, Providence, RI: American Mathematical Society.

Rodriguez-Fernandez, M., Rehberg, M., Kremling, A., Banga, J.R. (2013) Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems. *BMC System Biology*, **7** (76), 1-14.

Saccomani, M.P., Audoly, S., D'Angiò, L. (2004) Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, **39**, 619-632.

Saccomani, M.P., Thomaseth, K. (2016) Structural vs Practical Identifiability of Nonlinear Differential Equation Models in Systems Biology. In Rogato, A., Zazzu, V., Guarracino, M.R. (eds.), *Dynamics of Mathematical Models in Biology*, Springer International Publishing, Switzerland, 31-42. doi: 10.1007/978-3-319-45723-9_3

Stigter, J.D., Molenaar, J. (2015) A fast algorithm to assess local structural identifiability. *Automatica*, **58**, 118-124.

Thomaseth, K., Batzel, J.J., Bachar, M., Furlan R. (2013) Parameter estimation of a model for baroreflex control of unstressed volume. In Batzel, J.J. Bachar, M., Kappel, F. (eds), *Mathematical Modeling and Validation in Physiology*, Springer-Verlag, Berlin, 215-246.

Villaverde, A.F., Barreiro, A., Papachristodoulou, A. (2016) Structural identifiability of dynamic systems biology models. *PLoS Comput. Biol.*, **12** (10), e1005153. doi: 10.1371/journal.pcbi.1005153