Computers in Biology and Medicine 40 (2010) 402–407

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Examples of testing global identifiability of biological and biomedical models with the DAISY software ☆

Maria Pia Saccomani [a,*], Stefania Audoly [b], Giuseppina Bellu [b], Leontina D'Angiò [b]

[a] Department of Information Engineering, University of Padova, Padova, Italy
[b] Department of Mathematics, University of Cagliari, Cagliari, Italy

ARTICLE INFO

Article history:
Received 29 September 2009
Accepted 8 February 2010

Keywords:
Biological models
Nonlinear dynamic systems
A priori identifiability
Parameter estimation
Software tool

ABSTRACT

DAISY (Differential Algebra for Identifiability of SYstems) is a recently developed computer algebra software tool which can be used to automatically check global identifiability of (linear and) nonlinear dynamic models described by differential equations involving polynomial or rational functions. Global identifiability is a fundamental prerequisite for model identification which is important not only for biological or medical systems but also for many physical and engineering systems derived from first principles. Lack of identifiability implies that the parameter estimation techniques may not fail but any obtained numerical estimates will be meaningless.

The software does not require understanding of the underlying mathematical principles and can be used by researchers in applied fields with a minimum of mathematical background.

We illustrate the DAISY software by checking the a priori global identifiability of two benchmark nonlinear models taken from the literature. The analysis of these two examples includes comparison with other methods and demonstrates how identifiability analysis is simplified by this tool. Thus we illustrate the identifiability analysis of other two examples, by including discussion of some specific aspects related to the role of observability and knowledge of initial conditions in testing identifiability and to the computational complexity of the software. The main focus of this paper is not on the description of the mathematical background of the algorithm, which has been presented elsewhere, but on illustrating its use and on some of its more interesting features.

DAISY is available on the web site http://www.dei.unipd.it/~pia/.

1. Introduction

Mathematical models used to described biological dynamical systems are often complex, nonlinear and depending on unknown parameters. For example the Michaelis–Menten equation is often used to describe the internal structure of the biochemistry of the system. Often in these models the system parameters contain key information but these parameters can only be measured indirectly as it is usually not possible to measure directly the dynamics of every portion of the system. The recovery of parameter values can then only be approached indirectly as a parameter estimation problem starting from external, input-output measurements [20]. *Global* (*unique*) *identifiability* [6,12,20] concerns the possibility of uniquely determining the model parameters from input-output data, under ideal conditions of noise-free observations and error-free model structure. If, even in such an ideal situation, it turns out that the parameters of the postulated model are not uniquely identifiable, then there is no way that the parameters can be identified in a real-life situation, where errors in the model structure and noise in the data are inevitably present. If one tries to numerically estimate the parameters of the model, the optimization algorithm will provide unreliable (essentially random) numbers, not informative about the physiological process under analysis. In case of non-identifiability, one has to simplify the model structure, which may be too complex for the particular experiment setup and/or, if it is possible, enrich the planned input-output experiment with additional sampling sites or measurements.

Global (unique) identifiability is often neglected by many researchers, who start from the experimental data and then try to fit a model structure to the acquired data to identify parameter values. Serious problems may happen if a nonuniquely identifiable model is used for clinical studies: it may provide ambiguous estimates of some key model parameters, i.e. informative about the normal vs pathological state, leading the researcher physician to draw totally erroneous conclusions on the health state of the patient. An example of this situation has been described [4] for an

☆This work was partially supported by National Institutes of Health Grant R01 GM070635.

* Corresponding author. Tel.: +39 49 8277628; fax: +39 49 8277699.
E-mail address: pia@dei.unipd.it (M.P. Saccomani).

0010-4825/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2010.02.004

extremely popular model of oral drug dosing, where estimates of absorption and elimination rates are ambiguous and can be switched irrespectively of model fit to experimental data.

Identifiability analysis can be helpful also to provide guidelines to deal with non-identifiability, either providing hints on how to simplify the model structure or indicating when more information (measurable data) are needed to allow unique identifiability [15]. In many biological, especially clinical studies, checking a priori global identifiability of the underlying model may save resources in performing expensive and/or difficult experiments which may not be sufficiently informative for parameter identification. Although necessary, however, global identifiability is obviously not sufficient to guarantee an accurate identification of the model parameters from real input/output data.

The need to investigate the identifiability properties of nonlinear systems is unquestionable. In fact nonlinear models are very common in applied sciences and don't only concern enzyme kinetics and drug metabolism. Referring to the more recent literature, different approaches have been proposed which can be grouped in three different classes: the revisited Taylor Series approach [11], the methods based on the local state isomorphism theorem [21,6,5,9], and the differential algebra based methods [12,14,1,17,19].

The advantages and disadvantages of these methods have been long discussed in the literature. Yet there seem to be no general agreement on which method should be suggested to approach specific classes of problems.

For completeness one should mention also data-driven (i.e. simulation based) methods, which work on simulated model data coupled with parameter estimation routines. Although in principle these methods may give an answer in case of local identifiability, they are prone to error in case of indistinguishable multiple parameter values, see e.g. the pharmacokinetics model [4]. Moreover, it may be impossible to distinguish between non-identifiability and lack of convergence of the iterative optimization algorithm used for parameter estimation.

Our software tool, DAISY (Differential Algebra for Identifiability of SYstems) [2], recently developed and available on http://www.dei.unipd.it/~pia/ is designed to automatically check identifiability of (linear and) nonlinear dynamic models described by polynomial or rational functions. It implements a differential algebra algorithm based on [1,17] and it is written in the symbolic language REDUCE version 3.8. We refer the reader to [2] where this algorithm is described in detail. Despite of several other proposed methods it seems to be the only software tool currently available to the purpose. Naturally most other proposed methods can in principle be implemented in a symbolic language, but, to the best of our knowledge, this needs to be done on a *case by case* basis. The user has to implement by herself the identifiability algorithm adapting it to the particular system at hand. The reason for the particular choice of the symbolic language (REDUCE) used in this software tool is that it is particularly adapted to Gröbner basis computations. This has been pointed out in several publications [2]. We should mention that some of the differential algebra methods proposed in the literature [19] are implemented by using differential algebra packages available in Maple or in other symbolic languages [3].

In this paper, we present a direct comparison in testing the identifiability of two benchmark models which have been analyzed by some of the identifiability methods mentioned above. In addition to the advantage of DAISY in providing a completely automatized tool, we stress its flexibility and its capability to simplify the calculations and to provide a global, not only a local, answer to the identifiability problem. Furthermore, the identifiability of a third model is presented to show how the knowledge of initial conditions in testing identifiability can play a role.

Ideally, one would like to establish the domain of validity of the identifiability algorithm rigorously in terms of the model structure complexity (number of states, number of parameters, number of essential nonlinearities, number of measured variables). This is however very difficult, if not impossible. It would in fact require defining the limits of applicability of the Buchberger algorithm for solving the exhaustive summary of the model, which does not only depend on the number of parameters, but also, and in a complicated way, on the other variables. For example the computational cost decreases with the number of observed variables and a model with twenty states and twenty observed variables may turn out to be easier to analyze than a very interconnected model with four states and only one output. We have tested the algorithm on a library of complex linear and nonlinear models used in the biological and biomedical literature. In particular, we have evaluated the performance of DAISY in dealing with the most general input-output experimental configurations, i.e. characterized by multi input-multi output experiments with possible nonlinear constraints linking the input, the output and the model parameters, in order to maximize the likelihood of success and impact in biology and medicine.

The computation of the identifiability results is usually fast (few seconds) for nonlinear models of order 4 or 5 (even more if many of these states are measured) in a commercially available PC. However for very complex models, with more than one nonlinearity, the algorithm may not successfully terminate due to a lack of memory. In this paper, a fourth example of a high dimensional model is presented where the computational times required by DAISY to check the identifiability of this model and of those obtained by modifying the number of states are reported. This analysis shows, by way of an example, both potentiality and limitations of DAISY in terms of computational complexity/time.

The layout of the paper is the following:

- In Section 2, the basic mathematical background is introduced and some definitions regarding parameter identifiability are recalled.
- In Sections 3 and 4 we present the DAISY software checking the a priori global identifiability of two nonlinear biological models: a tumor model and a complex pharmacokinetics model. A comparison with the previous methods used to test identifiability of the same models is presented.
- In Section 5 a three compartment model with known initial conditions is presented to show the role of initial conditions in parameter identifiability.
- In Section 6 the identifiability analysis of a model with twenty states and twenty-two unknown parameters is discussed to show both potentiality and limitations of DAISY in terms of computational complexity/time. All these examples bring up the relevance of testing global identifiability in biological studies, thus showing the advantages of a software tool able to check it in a fully automatic way, as DAISY does.

2. Checking a priori identifiability

This section provides the reader with a brief description of the theory behind the software tool. A detailed documentation of the theory is reported in [1,17].

2.1. The system-experiment model

Many dynamic models of biological and physiological systems can be described by a general nonlinear system $\Sigma_{\mathbf{p}}$ in state space

form, depending on an unknown parameter vector $\mathbf{p}$:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{p}) + \sum_{i=1}^{m} \mathbf{g}_i(\mathbf{x}(t), \mathbf{p}) u_i(t) \tag{1}$$

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{u}(t), \mathbf{x}(t), \mathbf{p}) \tag{2}$$

where $\mathbf{x}$ is the n-dimensional state variable, representing, e.g. masses, concentrations, etc.; $\mathbf{u}$ the m-dimensional input vector made of smooth functions; $\mathbf{y}$ is the r-dimensional output; $\mathbf{p} \in \mathcal{P}$ is a v- dimensional vector of unknown parameters. The entries of $\mathbf{f}$, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_m]$ and $\mathbf{h}$ are assumed to be *rational functions* of their arguments. The dependence on $\mathbf{p}$ may be rational.

Note that (1) and (2) describe the system together with a certain choice of input-output variables which is designed for the identification experiment. In biological/medical applications, identification experiments are often performed on systems at rest or started from known (equilibrium) initial conditions. If initial conditions are specified, the relevant equation $\mathbf{x}(t_0)=\mathbf{x}_0$ (or $\mathbf{x}(t_0)=\mathbf{x}_0(\mathbf{p})$ if initial conditions are parameter dependent) is added to the system.

Equality constraints (linear or nonlinear) of the form $E(\mathbf{p})=0$, where E is a polynomial or a rational vector function, may be also present.

2.2. Definitions

A priori global identifiability addresses the following question: given arbitrary parameters $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}$, $\mathbf{p}_1 \neq \mathbf{p}_2$ and assuming the systems $\Sigma_{\mathbf{p}_1}$ and $\Sigma_{\mathbf{p}_2}$ are initialized at the same initial state; does there exist (at least) one input function $\mathbf{u}$ for which systems $\Sigma_{\mathbf{p}_1}$ and $\Sigma_{\mathbf{p}_2}$ yield different outputs? If so, $\Sigma_{\mathbf{p}_1}$ and $\Sigma_{\mathbf{p}_2}$ are distinguishable from the experiment and the system $\Sigma_{\mathbf{p}}$ is said to be *a priori globally* (*or uniquely*) *identifiable*. If instead systems $\Sigma_{\mathbf{p}_1}$ and $\Sigma_{\mathbf{p}_2}$ yield the same output no matter which input function is applied and no matter what initial state the system is started at, the parameter values $\mathbf{p}_1$ and $\mathbf{p}_2$ can never be distinguished and are said to be *indistinguishable*.

If there are at most a finite number of indistinguishable parameter values, the system is called *locally* (*or nonuniquely*) *identifiable*.

Finally, if there is an infinite number of different indistinguishable parameter values, the system is commonly called *nonidentifiable* or *unidentifiable* [6,12,21].

To check global identifiability with DAISY, the dynamical model should be provided in a separate file that follows a specific format described in the instructions of the software.

The input of the algorithm is provided by the differential polynomials defining the dynamic system (1) and (2), the number of inputs and outputs, the ranked list of input, output and state variables, the list of the unknown parameters, and, if present, the equality constraints among the parameters and the initial conditions of the system. DAISY can provide identifiability answers for both generic and fixed initial conditions, under a mild assumption of algebraic observability [17].

A note for the user. Most of the discussion below is a "theoretical" discussion of the mathematical procedures attempted in the literature for the considered examples versus the procedure used by DAISY, to motivate the superiority of our method/software in dealing with non-trivial examples. This discussion should actually not concern the practical user at all. She/he should only read off the last lines of the program output where the identifiability results are reported.

3. Identifiability of a tumor model

In this section we shall analyze the identifiability of a model of a tumor dealt with antibody presented in [18]. The model is mathematically described by the following polynomial nonlinear differential equations:

$$\begin{cases} \dot{x}_1(t) = -(k_3+k_7)x_1 + k_4 x_2 \\ \dot{x}_2(t) = k_3 x_1 - (k_4 + a k_5 + bd k_5)x_2 + k_6 x_3 + k_6 x_4 + k_5 x_2 x_3 + k_5 x_2 x_4 \\ \dot{x}_3(t) = a k_5 x_2 - k_6 x_3 - k_5 x_2 x_3 \\ \dot{x}_4(t) = bd k_5 x_2 - k_6 x_4 - k_5 x_2 x_4 \\ \dot{x}_5(t) = k_7 x_1 \\ y(t) = x_5 \end{cases} \tag{3}$$

where x_i, $i=1,\ldots,5$, are the compartmental concentrations, k_3, k_4, k_5, k_6 and k_7 are the unknown rate parameters, a, b and d are known constants and y is the measured output in the plasma.

The initial conditions of the model are $\mathbf{x}(0)=[x_{10},0,0,0,0]$ where x_{10} is known.

The main question to be addressed is whether the unknown parameter vector $\mathbf{p}=[k_3,k_4,k_5,k_6,k_7]$ is globally identifiable from the input-output experiment.

The a priori identifiability of this model has been first analyzed by using the local state isomorphism theorem [7]. As pointed out by [9], this method requires a rather limiting linearity assumption. In order to overcome this limitation, the method has been generalized by [9]. The method however can only tests local identifiability of the model.

The identifiability of the tumor model has been also checked via the Taylor series method coupled with the method of the input-output equation based on the so-called "Logiciel Identifiabilité" [13] but this complex approach (surely difficult to be automatized) requires some assumption in the optimization procedure which makes it less reliable [13].

We have tested the global identifiability of the tumor model by using DAISY. It is easy to appreciate the advantages of using this computer algebra tool: in about one second DAISY provides all the required calculations to give the global identifiability answer. In addition, with respect to the previous cited methods, DAISY tests global (not local) identifiability of the model, issue particularly critical in physiological systems where a different numerical estimate can characterize a pathological from a normal state. Furthermore, no linearity assumptions are required. As theoretically justified in [17], DAISY automatically includes, if available, the knowledge of the initial conditions of the model in the identifiability analysis. The input file is reported, as an example, in the Appendix.

After choosing the standard ranking of the input, output and state variables, e.g. $y < x_1 < x_2 < x_3 < x_4 < x_5$ the program automatically starts the pseudodivision algorithm and calculates the characteristic set of the model [17]:

$$\begin{aligned} &\dddot{y}\ddot{y}k_4k_5k_7 + \dddot{y}\ddot{y}k_4k_5k_7(k_3+k_7) + \dddot{y}\dot{y}k_4^2k_6k_7^2 - \ddot{y}^2k_4k_5k_7 + \ddot{y}\dot{y}^2k_5^2 \\ &+ 2\ddot{y}\ddot{y}\dot{y}k_5^2(k_3+k_7) + \ddot{y}\ddot{y}k_4k_5k_7(-k_3+2k_6-k_7) + \ddot{y}\dot{y}^2k_5^2(k_3^2 \\ &+ 2k_3k_7+k_7^2) + \ddot{y}\dot{y}k_4k_5k_7(k_3^2+k_3k_4+2k_3k_6+2k_3k_7+2k_6k_7 \\ &+ k_7^2) + \ddot{y}\dot{y}k_4^2k_6k_7^2(k_3+k_4+(a+bd)k_5+k_6+k_7) + \ddot{y}^3k_5^2(k_3+k_4+k_7) \\ &+ \ddot{y}^2\dot{y}k_5^2(2k_3^2+2k_3k_4+4k_3k_7+3k_4k_7+2k_7^2) + \ddot{y}^2k_4k_5k_7(-k_3^2-k_3k_4 \\ &+ 2k_3k_6-2k_3k_7+2k_4k_6+2k_6k_7-k_7^2) + \ddot{y}\dot{y}^2k_5^2(k_3^3+k_3^2k_4+3k_3^2k_7 \\ &+ 4k_3k_4k_7+3k_3k_7^2+3k_4k_7^2+k_7^3) + 2\ddot{y}\dot{y}k_4k_5k_6k_7(k_3^2+k_3k_4+2k_3k_7 \\ &+ 2k_4k_7+k_7^2) + \dot{y}k_4^2k_6k_7^2((a+bd)k_3k_5+k_3k_6+k_4k_6+k_4k_7 \\ &+ (a+bd)k_5k_7+k_6k_7) + \dot{y}^3k_4k_5^2k_7(k_3^2+2k_3k_7+k_7^2) + 2\dot{y}^2k_4^2k_5k_6k_7^2(k_3 \\ &+ k_7) + \dot{y}k_4^3k_6^2k_7^3, \\ &\dot{y} - x_1k_7, \end{aligned}$$

$$
\begin{aligned}
&\ddot{y}+\dot{y}(k_3+k_7)-x_2k_4k_7,\\
&\ddot{y}-\dot{y}x_3k_5-\dot{y}x_4k_5\\
&+\dot{y}(k_3+k_4+(a+bd)k_5+k_7)-\dot{y}x_3k_5(k_3+k_7)-\dot{y}x_4k_5(k_3+k_7)\\
&+\dot{y}((a+bd)k_3k_5+k_4k_7+(a+bd)k_5k_7)-x_3k_4k_6k_7-x_4k_4k_6k_7,\\
&-x_5+y,\\
&\dot{x}_3k_4k_7+\ddot{y}x_3k_5-a\ddot{y}k_5+\dot{y}x_3k_5(k_3+k_7)-a\dot{y}k_5(k_3+k_7)\\
&+x_3k_4k_6k_7
\end{aligned}
\tag{4}
$$

After having normalized the input-output differential relation, i.e. the first polynomial of (4), by the coefficient $k_4\,k_5\,k_7$ in order to set it monic, the so-called *exhaustive summary* of the model, namely the coefficients of the input-output relation, are extracted. DAISY checks identifiability by solving for the unknown parameters the algebraic nonlinear equations (not reported here for brevity) obtained by equating these coefficients to a set of pseudo-randomly chosen numerical values in their range set. To do this DAISY applies the Buchberger algorithm which provides the following Gröbner basis:

$$
\begin{aligned}
&k_3-55\\
&k_4-87\\
&k_5-83\\
&k_6-8\\
&k_7-70
\end{aligned}
\tag{5}
$$

To make sure that the pseudo-randomly chosen numerical point does not belong to a "singular" set of zero measure in the parameter space, we repeat this procedure for a "reasonably large" set of randomly chosen parameter values. This confirms the above identifiability result and shows that all the parameters k_3, k_4, k_5, k_6 and k_7 are uniquely identifiable from input-output experiments.

The construction of the characteristic set ignores initial conditions and the identifiability result above holds irrespective of which initial conditions the system may have started from. Therefore, in this case, knowledge of the initial conditions is redundant to assess identifiability, given that all the parameters turn out to be globally identifiable even without the knowledge of the initial conditions. If, for example, the initial condition of the model x_{10} was not known, DAISY would consequently deal with this unknown initial condition as an extra parameter to be identified. In more technical words, DAISY would check the algebraically observability [17] of the state x_1 from the characteristic set. In this case x_1 was algebraically observable and thus DAISY would provide only one solution for x_{10}, which would then turn out to be globally identifiable from the experimental data.

As stated in the introduction, once a priori global identifiability is assessed, this is not sufficient to guarantee an accurate identification of the model parameters from real input/output data, where error in the model structure and noise in the data are inevitably present. In fact, the accuracy of the estimates will eventually depend on other different factors, as for example, noise, the paucity of the available data and/or the complexity of the model with respect to the available data.

Remark 1. DAISY is able to automatically provide, in the same output file, the identifiability answer both for the unknown parameters and for the initial conditions. However, from a theoretical point of view, this answer comes from two different steps: the identifiability of the model parameters comes from the investigation of the input-output relations of the characteristic set. The identifiability of the initial conditions comes from the investigation of the whole characteristic set. The procedure is however internal to the algorithm and the user is not required to be aware of these theoretical steps.

4. Identifiability of a pharmacokinetics model

In this section we shall analyze a nonlinear biological model [8] proposed for the study of the ligands of the macrophage mannose receptor and employed as a benchmark model in identifiability studies of the last ten years. It is mathematically described by the following rational differential equations:

$$
\begin{cases}
\dot{x}_1(t)=a_1(x_2-x_1)-k_aV_mx_1/(k_ck_a+k_cx_3+k_ax_1)\\
\dot{x}_2(t)=a_2(x_1-x_2)\\
\dot{x}_3(t)=b_1(x_4-x_3)-k_cV_mx_3/(k_ck_a+k_cx_3+k_ax_1)\\
\dot{x}_4(t)=b_2(x_3-x_4)\\
y_1(t)=x_1
\end{cases}
\tag{6}
$$

where x_1 and x_2 are glucose oxidase concentrations, x_3 and x_4 are the concentrations of the mannosylated polymer in plasma and in the part of the extravascular fluid of the organs accessible to this macromolecule, respectively. The initial conditions are $\mathbf{x}(0)=[C_0, 0, x_{30}, 0]$, where C_0 is the known initial concentration while x_{30} is supposed to be unknown. The unknown parameters are a_1, a_2, b_1, b_2, k_a, V_m, k_c (in the original version a and b were denoted by the corresponding Greek letters). Finally y_1 is the measured output.

In the literature, the a priori identifiability has been first analyzed with the method based on the local state isomorphism theorem, [21,6,5], by using "piece of information one after another about the Jacobian matrix" as declared in [9]. As stated in [19], this method requires the introduction of a fictitious control which does not correspond to reality. Successively, in [11] and [10] the identifiability of the model has been tested via the Taylor series method by considering more than one experiment, the so-called "augmented" system. This method suffers of the same problems of that based on the Taylor series and its implications to the estimation procedure are not clear. This procedure has been criticized by [10] since it is able to test only local but not global identifiability and as discussed by the authors themselves in [19], since it requires some unrealistic hypotheses on the model parameters.

More recently this model has been analyzed with a differential algebra approach by [19]; in particular the analysis in [19] was carried on in two steps: first the global identifiability of a submodel having only the four unknown parameters a_1, a_2, V_m, k_c was tested; second, the whole model (6) was re-defined as follows: parameters a_1, a_2, V_m, k_c were assumed to be known from the previous submodel and thus the unknown parameter vector was b_1, b_2, k_a [19]. In particular, the authors themselves have used this assumption to introduce what they call an "artificial" output [19]:

$$
y_2(t)=x_2(t) \tag{7}
$$

This could be done only because, by assuming a_2 known from the previous model, also the state variable x_2 could be assumed to be known (remember that $x_2(0)$ was known). The idea of introducing this extra output has allowed to considerably simplify the identifiability calculations of the whole model (6). In particular the identifiability of the remaining three unknown parameters has been successfully assessed. Finally also the third unknown initial condition x_{30} has been deduced.

The procedure seems to be a little *ad hoc* and also, in the numerical estimation procedure, it requires to fix four of the seven unknown parameter values to estimated values obtained from data of a different model-experiment. This seems to render less reliable the identification results.

We shall analyze the model of [19] with the software tool DAISY.

Since we want to compare our results with those obtained in [19], we start directly with the whole model just described but,

unlike [19], of the four parameters a_1,a_2, V_m, k_c only a_2 is assumed to be known (in order to have a second output function y_2). DAISY calculates the exhaustive summary of the model from the characteristic set, which is not reported here for lack of space, solves it by using the Gröbner basis algorithm and shows global identifiability of all the unknown parameters of the model and of the unknown initial condition x_{30}.

This result seems to fully agree with that obtained in [19], but here only one parameter is assumed fixed, in place of four, thus allowing a more reliable performance of the model. Incidentally, we believe that using our differential algebra tool the result is obtained directly, very quickly and following a conceptually clean line of reasoning. DAISY is in fact able to solve all the calculations required to test the global identifiability of all the unknown parameters a_1, b_1, b_2, k_a, V_m, k_c plus x_{30} in few seconds.

5. Identifiability of a model with known initial conditions

In this section we shall show by way of an example, how initial conditions can play a role in parameter identifiability. We shall analyze the identifiability of the following three compartments model:

$$\begin{cases} \dot{x}_1(t) = -t_1x_1 + t_2x_2 + u_1 \\ \dot{x}_2(t) = t_3x_1 - t_4x_2 + t_5x_3 \\ \dot{x}_3(t) = t_6x_1 - t_7x_3 \\ y_1(t) = x_1 \end{cases} \tag{8}$$

where x_1, x_2, x_3 are concentrations, t_i, $i=1,\ldots,7$ are the unknown parameters and y_1 is the measured output of the system. Initial conditions $\mathbf{x}(0)=[x_{10}, x_{20}, x_{30}]$ are known.

With the standard ranking of the variables, i.e. $u_1 < y_1 < x_1 < x_2 < x_3$, DAISY calculates the characteristic set of this model:

$$\begin{aligned} &-\ddot{u}_1 - \dot{u}_1(t_4+t_7) + \dddot{y}_1 + \ddot{y}_1(t_1+t_4+t_7) + \dot{y}_1(t_1t_4+t_1t_7-t_2t_3+t_4t_7) \\ &-u_1t_4t_7 + y_1(t_1t_4t_7 - t_2t_3t_7 - t_2t_5t_6), \\ &x_1 - y_1, \\ &\dot{y}_1 - u_1 - x_2t_2 + y_1t_1, \\ &-\dot{u}_1 + \ddot{y}_1 + \dot{y}_1(t_1+t_4) - u_1t_4 - x_3t_2t_5 + y_1(t_1t_4 - t_2t_3) \end{aligned} \tag{9}$$

its exhaustive summary:

$$\begin{aligned} &-t_4t_7 + 224, \\ &-t_4 - t_7 + 36, \\ &t_1 + t_4 + t_7 - 47, \\ &t_1t_4t_7 - t_2t_3t_7 - t_2t_5t_6 - 1889, \\ &t_1t_4 + t_1t_7 - t_2t_3 + t_4t_7 - 555 \end{aligned} \tag{10}$$

and the Gröbner basis solution:

$$\begin{aligned} &\{t_5 = 55/(t_2t_6), \quad t_3 = 65/t_2, \quad t_1 = 11, \quad t_4 = 28, \quad t_7 = 8\} \\ &\{t_5 = (-1245)/(t_2t_6), \quad t_3 = 65/t_2, \quad t_1 = 11, \quad t_4 = 8, \quad t_7 = 28\} \end{aligned} \tag{11}$$

showing that the model is nonidentifiable.

So far the knowledge of the initial conditions has not be included in parameter identifiability.

In order to include these known initial conditions, DAISY calculates at time 0 all the polynomials of the characteristic set (by excluding the input/output relation which was already known at each time). These polynomials, given the knowledge of the initial conditions, becomes known and thus can be added to the previous exhaustive summary of the model (10). Thus the new exhaustive summary is given by the previous polynomials (10) plus the following ones:

$$\begin{aligned} &t_1x_{10} - t_2x_{20} - 11x_{10} + 5x_{20}, \\ &t_1t_4x_{10} + t_1u_{10} - 11t_1x_{10} + 5t_1x_{20} - t_2t_3x_{10} - t_2t_5x_{30} - 11t_4x_{10} + 5t_4x_{20} - \\ &11u_{10} + 186x_{10} - 195x_{20} + 55x_{30} \end{aligned} \tag{12}$$

DAISY reapplies the Buchberger algorithm and obtains the following solutions:

$$\begin{aligned} &\{t_6 = 249x_{30}/(20x_{20} - 11x_{30}), \quad t_5 = (-20x_{20} + 11x_{30})/x_{30}, t_3 = 13, \\ &t_2 = 5, \quad t_1 = 11, \quad t_7 = 28, \quad t_4 = 8\} \\ &\{t_1 = 11, \quad t_2 = 5, \quad t_3 = 13, \\ &t_4 = 28, \quad t_5 = 11, \quad t_6 = 1, \quad t_7 = 8\} \end{aligned} \tag{13}$$

showing that now the system has a finite number of solutions. DAISY provides thus the answer "SYSTEM LOCALLY IDENTIFIABLE". In conclusion, the knowledge of initial conditions has allowed to locally identify the unknown model parameters.

6. Identifiability of a high dimensional nonlinear model

In this section, to show both potentiality and limitations of DAISY in terms of computational complexity/time, we shall analyze the identifiability of a nonlinear model with twenty states and twenty-two unknown parameters. The unidirectional structure is a classical structure adopted to describe many biochemical reaction systems. The model is described by the following equations:

$$\begin{cases} \dot{x}_1(t) = -v_{max}x_1/(k_m + x_1) - p_1x_1 + u_1 \\ \dot{x}_2(t) = p_1x_1 - p_2x_2 \\ \dot{x}_3(t) = p_2x_2 - p_3x_3 \\ \dot{x}_4(t) = p_3x_3 - p_4x_4 \\ \dot{x}_5(t) = p_4x_4 - p_5x_5 \\ \dot{x}_6(t) = p_5x_5 - p_6x_6 \\ \dot{x}_7(t) = p_6x_6 - p_7x_7 \\ \dot{x}_8(t) = p_7x_7 - p_8x_8 \\ \dot{x}_9(t) = p_8x_8 - p_9x_9 \\ \dot{x}_{10}(t) = p_9x_9 - p_{10}x_{10} \\ \dot{x}_{11}(t) = p_{10}x_{10} - p_{11}x_{11} \\ \dot{x}_{12}(t) = p_{11}x_{11} - p_{12}x_{12} \\ \dot{x}_{13}(t) = p_{12}x_{12} - p_{13}x_{13} \\ \dot{x}_{14}(t) = p_{13}x_{13} - p_{14}x_{14} \\ \dot{x}_{15}(t) = p_{14}x_{14} - p_{15}x_{15} \\ \dot{x}_{16}(t) = p_{15}x_{15} - p_{16}x_{16} \\ \dot{x}_{17}(t) = p_{16}x_{16} - p_{17}x_{17} \\ \dot{x}_{18}(t) = p_{17}x_{17} - p_{18}x_{18} \\ \dot{x}_{19}(t) = p_{18}x_{18} - p_{19}x_{19} \\ \dot{x}_{20}(t) = p_{19}x_{19} - p_{20}x_{20} \end{cases} \tag{14}$$

where x_i, $i=1,\ldots,20$ are the states. The unknown parameters are p_i, $i=1,\ldots,20$, plus the two Michaelis-Menten parameters v_{max} and k_m. All the states are assumed to be measured. The model is fed to DAISY running on an Intel Core Duo processor at 3.16 GHz and 3 GB RAM. In about 150 min DAISY provides the answer: "SYSTEM GLOBALLY IDENTIFIABLE". Simplifying the model by considering only the first eighteen states, DAISY provides the identifiability answer in about 30 min. This computational time decreases to 15 s if the model is reduced to 10 states and to 3 s for 8 states.

7. Conclusions

In this paper we have presented applications of a differential algebra based approach for checking a priori identifiability, as described in [17]. We discuss two nonlinear biological models

from the literature: a tumor model and a pharmacokinetics model, both widely used as benchmark models to test and compare the different identifiability methods. A third model has been analyzed to show the role of initial conditions in parameter identifiability. Finally a twenty states nonlinear model has been considered to assess the potentiality/limitations of the approach. The DAISY software described in [2] has proven to be a very useful and easy to use tool for this purpose.

To our best knowledge DAISY is the only software tool which, although being based on a rather sophisticated mathematical tools, does not require expertise on mathematical modeling by the experimenter. For this reason it is believed to be an useful instrument for the medical/biological investigator.

Conflict of interest statement

None declared.

Appendix A. Input file of the tumor model

```
WRITE "A TUMOR MODEL"$
% B_ IS THE VARIABLE VECTOR
B_:={y,x1,x2,x3,x4,x5}$
FOR EACH EL_ IN B_ DO DEPEND EL_,T$
%B1_ IS THE UNKNOWN PARAMETER VECTOR
B1_:={k3,k4,k5,k6,k7}$
%NUMBER OF STATES
NX_:=5$
%NUMBER OF OUTPUTS
NY_:=1$
%MODEL EQUATIONS
C_:={df(x1,t)=-(k3+k7)*x1+k4*x2,
    df(x2,t)=k3*x1-(k4+a*k5+b*d*k5)*x2+
    k6*x3+k6*x4+k5*x2*x3+k5*x2*x4,
    df(x3,t)=a*k5*x2-k6*x3-k5*x2*x3,
    df(x4,t)=b*d*k5*x2-k6*x4-k5*x2*x4,
    df(x5,t)=k7*x1,
    y=x5}$
SEED_:=120$
DAISY ()$
%VALUES OF INITIAL CONDITIONS ARE GIVEN
IC_:={x1=x10,x2=0,x3=0,x4=0,x5=0}$
CONDINIZ()$
```

References

[1] S. Audoly, G. Bellu, L. D'Angiò, M.P. Saccomani, C. Cobelli, Global identifiability of nonlinear models of biological systems, IEEE Trans. Biomed. Eng. 48 (1) (2001) 55–65.

[2] G. Bellu, M.P. Saccomani, S. Audoly, L. D'Angiò, DAISY: A new software tool to test global identifiability of biological and physiological systems, Comput. Methods Programs Biomed. 88 (2007) 52–61.

[3] F. Boulier, D. Lazard, F. Ollivier, M. Petitot, Representation for the radical of a finitely generated differential ideal, Proc. Int. Symp. Symbolic and Algebraic Computation, ISSAC'95, Montréal Canada, 1995, 158–166.

[4] C. Cobelli, M.P. Saccomani, Unappreciation of a priori identifiability in software packages causes ambiguities in numerical estimates, Am. J. Physiol. 21 (1990) E1058–E1059 (Letter to the Editor).

[5] M.J. Chapman, K.R. Godfrey, M.J. Chappell, N.D. Evans, Structural identifiability of non-linear systems using linear/non-linear splitting, Int. J. Control 76 (3) (2003) 209–216.

[6] M.J. Chappell, K.R. Godfrey, Structural identifiability of the parameters of a nonlinear batch reactor model, Math. Biosci. 108 (1992) 245–251.

[7] M.J. Chappell, K.R. Godfrey, S. Vadja, Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods, Math. Biosci. 102 (1990) 41–73.

[8] S. Demignot, D. Domurado, Effect of prosthetic sugar groups on the pharmacokinetics of glucose-oxidase, Drug Design Deliv. 1 (1987) 333–348.

[9] L. Denis-Vidal, G. Joly-Blanchard, Identifiability of some nonlinear kinetics, in: Proceedings of the Third Workshop on Modelling of Chemical Reaction Systems, Heidelberg, 1996.

[10] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, Some effective approaches to check the identifiability of uncontrolled nonlinear systems, Math. Comput. Simulation 57 (2001) 35–44.

[11] G. Joly-Blanchard, L. Denis-Vidal, Some remarks about identifiability of controlled and uncontrolled nonlinear systems, Automatica 34 (1998) 1151–1152.

[12] L. Ljung, S.T. Glad, On global identifiability for arbitrary model parameterizations, Automatica 30 (2) (1994) 265–276.

[13] C. Noiret, Utilisation du calcul formel pour l'identifiabilité de modèles paramétriques et nouveaux algorithmes en estimation des paramétres, Ph.D. Thesis, Université de Technologie de Compiègne, 2000.

[14] F. Ollivier, *Le problème de l'identifiabilité structurelle globale: étude théorique, méthodes effectives et bornes de complexité*, Thèse de Doctorat en Science, École Polytéchnique, Paris, France, 1990.

[15] M.P. Saccomani, C. Cobelli, A minimal input-output configuration for a priori identifiability of a compartmental model of leucine metabolism, IEEE Trans. Biomed. Eng. 40 (1993) 797–803.

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

[18] G.D. Thomas, M.J. Chappell, P.W. Dykes, D.B. Ramsden, K.R. Godfrey, J.R.M. Ellis, Effect of dose, molecular size, affinity, and protein binding on tumor uptake of antibody or ligand: a biomathematical model, Cancer Res. 49 (1989) 3290–3296.

[19] N. Verdière, L. Denis-Vidal, G. Joly-Blanchard, D. Domurado, Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrophage mannose receptor, Int. J. Appl. Math. Comput. Sci. 15 (4) (2005) 517–526.

[20] E. Walter, Identifiability of State Space Models, Springer, Berlin, 1982.

[21] E. Walter, Y. Lecourtier, Global approaches to identifiability testing for linear and nonlinear state space models, Math. Comput. Simulation 24 (1992) 472–482.