



OPEN

Assessing the role of initial conditions in the local structural identifiability of large dynamic models

Dominique Joubert[✉], J. D. Stigter & Jaap Molenaar

Structural identifiability is a binary property that determines whether or not *unique* parameter values can, in principle, be estimated from error-free input–output data. The many papers that have been written on this topic collectively stress the importance of this a priori analysis in the model development process. The story however, often ends with a structurally unidentifiable model. This may leave a model developer with no plan of action on how to address this potential issue. We continue this model exploration journey by identifying one of the possible sources of a model's unidentifiability: problematic initial conditions. It is well-known that certain *initial values* may result in the loss of local structural identifiability. Nevertheless, literature on this topic has been limited to the analysis of small toy models. Here, we present a systematic approach to detect problematic initial conditions of real-world systems biology models, that are usually not small. A model's identifiability can often be reinstated by changing the value of such problematic initial conditions. This provides modellers an option to resolve the “unidentifiable model” problem. Additionally, a good understanding of which initial values should rather be avoided can be very useful during experimental design. We show how our approach works in practice by applying it to five models. First, two small benchmark models are studied to get the reader acquainted with the method. The first one shows the effect of a zero-valued problematic initial condition. The second one illustrates that the approach also yields correct results in the presence of input signals and that problematic initial conditions need not be zero-values. For the remaining three examples, we set out to identify key initial values which may result in the structural *unidentifiability*. The third and fourth examples involve a systems biology Epo receptor model and a JAK/STAT model, respectively. In the final Pharmacokinetics model, of which its global structural identifiability has only recently been confirmed, we indicate that there are still sets of initial values for which this property does not hold.

We have grown to appreciate the importance of model accuracy as societies increasingly depend on model predictions to answer difficult questions, such as which measures should be taken to fight the Covid-19 pandemic. Clearly, mathematical models are powerful tools. They often contain unknown parameters which need to be estimated from experimental data since they are not measurable. Model developers often run into practical difficulties whilst estimating the values of such parameters. These range from significant computational memory requirements to insufficient funding to perform required experiments¹. Here we highlight *model identifiability*, one of the primary challenges encountered during in parameter estimation. We focus on the role *initial conditions* play in a model's local structural identifiability.

Parameter identifiability assesses whether it is possible to infer *unique* parameter values from gathered input–output data. It can be divided into two main categories with structural or a priori identifiability the prerequisite for practical identifiability. Structural identifiability is based on a model's structure and is a binary property and so supposes that experimental measurements are error free. In contrast, practical identifiability characterises the ability to estimate parameters from observed data containing measurement errors and so takes both data quality and availability into account².

Developing dynamic models using ordinary differential equations requires the formulation of differential (and possibly accompanying algebraic equations) and calls for the definition of measurable outputs and initial

Biometris, Department of Mathematical and Statistical Methods, Wageningen University and Research, 6708 PD Wageningen, The Netherlands. ✉email: dominique.joubert@wur.nl

conditions. A model's structural identifiability can be affected by these chosen initial values^{3–5}. It may therefore be impossible to uniquely estimate certain parameters if a system that evolves from a “problematic set of initial conditions”. In 2001 Denis-Vidal and co-authors alluded to this by stressing the importance of an appropriate choice of initial values³. This was confirmed by Saccomani et al.⁴ who studied the role initial conditions play in the identifiability of controlled models. They commented that “it happens frequently in global identifiability analyses that the property only holds generically, i.e. except for a ‘thin’ set of initial conditions. In these situations the system is (incorrectly but forgivably) declared to be (global) identifiable”⁴. Finally, Villaverde and Banga remarked that both the symbolic differential algebra and differential geometry methods, which are often used to assess a model's identifiability, do not recognise the loss of local structural identifiability for very specific initial values⁵. Furthermore, they list a number of methods that can detect these problematic initial conditions, including their own differential geometry based approach. These methods include⁵:

- Exact Arithmetic Rank method (EAR)—This method works for rational systems was introduced in 2012 by Karlsson et al.⁶. It allows for the definition of specific initial values and uses these in efficient numerical calculations.
- DAISY—This method also analyses rational systems exclusively and uses the differential algebra method⁷.
- STRIKE-GOLDD—This method adopts a differential geometry approach to analyse a model's local structural identifiability and so entails the symbolical calculation of successive Lie derivatives⁸.

It should be mentioned that some methods do not always yield correct results for certain special cases. In the “Discussion” we mention such a case explicitly. The method we present in this paper has a higher degree of reliability since it combines 2 analyses, one numerical and one symbolic. It is also not restricted to rational or relatively small models. As extensively described in the “Methods” section, our approach is based on the identifiability algorithm presented by Stigter and Molenaar⁹. The method starts with a numerical analysis of the rank of the so-called sensitivity matrix and is available as a downloadable application¹⁰. This initial step pinpoints potentially unidentifiable parameters and initial conditions. In the second step a symbolic analysis is performed to check whether these parameters and initial conditions are indeed causing identifiability problems. Since this analysis only involves a restricted number of parameters and initial conditions, the computational demand which is often found to be the curse of a symbolic analysis of the full model, is significantly reduced. The numerical results from the first step can attractively be summarised in a so-called “identifiability signature”¹¹. This signature contains a graphical presentation of the singular values resulting from a Singular Value Decomposition (SVD) of the sensitivity matrix. Structural unidentifiability is indicated by a *clear gap in the displayed singular values*. Furthermore, the signature shows the components of the singular vectors corresponding to the close-to-zero singular values. The nonzero components of these vectors reveal which parameters and initial conditions are expected to be unidentifiable.

An advantage of structural identifiability is that it can be assessed before the experimental phase. Since the method presented in this paper allows for the efficient identification of problematic initial conditions, the values of these may be altered before any expenses on wasteful experiments are incurred. So, it may play an essential role in the design-of-experiment stage that should precede any experimental endeavour.

Results

The discussion of problematic initial conditions has in the past been limited to small toy models. To show the effectiveness and power of the present approach, we apply the method to 5 models. These examples have been carefully chosen to demonstrate different aspects of the approach. First, 2 small benchmark models are studied to get the reader acquainted with the method. The first one shows the effect of a zero-valued problematic initial condition. The second illustrates that the approach also yields correct results in the presence of input signals and that problematic initial values need not be zero-values. The third example comprises a realistic systems biology model describing Epo receptor dynamics. For this Epo model we show that it is possible - thanks to the efficiency of the approach - to identify the precise (zero-valued) initial conditions that result in model unidentifiability. The fourth example is an analysis of the well-known JAK/STAT model, for which the cause of its unidentifiability was not yet published in the literature. In the last example, of which its global structural identifiability has only recently been confirmed, we indicate that there are still sets of initial values for which this property does not hold.

Example 1 Small benchmark model (M_1).

In example 1 we analyse a small academic model published by Denis-Vidal et al.³. This example illustrates the potential role initial values play in the structural identifiability of uncontrolled models. The model contains two state equations:

$$\frac{dx_1}{dt} = p_1x_1^2 + p_2x_1x_2, \quad x_1(0) = x_{10} \neq 0, \quad (1)$$

$$\frac{dx_2}{dt} = p_3x_1^2 + x_1x_2, \quad x_2(0) = 0. \quad (2)$$

The first state is measured directly, so the output is $y = x_1$. The three system parameters are assumed unknown and therefore the parameter vector is $\theta = [p_1, p_2, p_3]$. In Fig. 1 we present the identifiability signature of this model resulting from the numerical analysis; see “Methods” section for details. It shows a clear gap between the second and third singular values, which indicates that the model is structurally unidentifiable given the set of

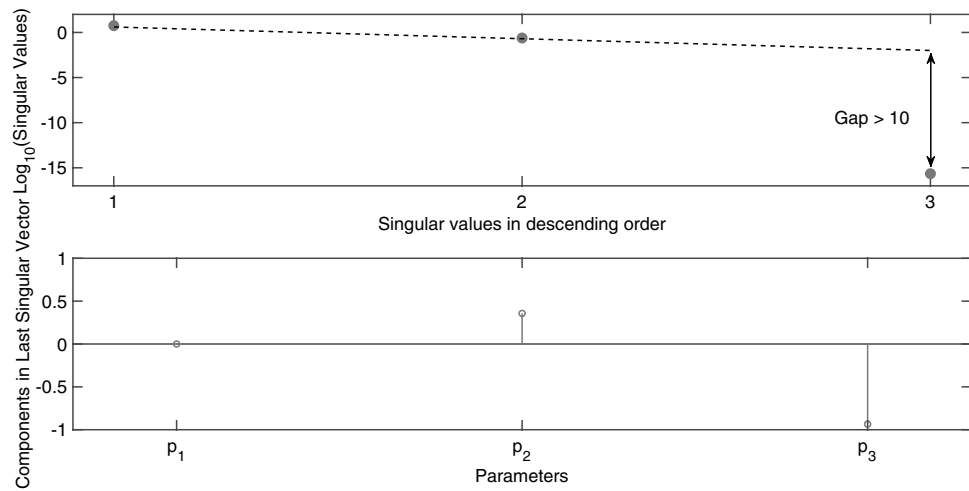


Figure 1. Identifiability signature of Example 1: small benchmark model (M_1). (Top) Singular values: For $x_2(0) = 0$ and $y = x_1$, the rank deficiency of the sensitivity matrix is indicated by a distinct gap between the second and third singular values. This suggests that this model is structurally unidentifiable and that there is a single set of totally correlated parameters. (Bottom) Elements of the singular vector related to the zero valued singular value. The nonzero components indicate that parameters p_2 and p_3 are structurally unidentifiable.

initial conditions in Eqs. (1) and (2). Our result is in line with the analysis by Denis-Vidal et al. which shows that parameters p_2 and p_3 are structurally unidentifiable³. At the bottom of Fig. 1 the components of the singular vector corresponding to the close-to-zero singular value of the sensitivity matrix are given. The nonzero components correspond to parameters p_2 and p_3 respectively; this indicates that these parameters are indeed unidentifiable. The second step in our method is to check these numerical suggestions symbolically. The parameter set to be analysed is now reduced to $\theta^{unid} = \{p_2, p_3\}$ (obtained from Fig. 1). The Jacobi matrix needed in this second step (see the “Methods” section), is given in Eq. (3) as a 4×2 matrix, with each column related to a parameter in θ^{unid} . It is computed by calculating partial derivatives of successive Lie derivatives, defined in Eqs. (56) and (57).

$$\frac{dG}{d\theta^{unid}}(x(0), \theta) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ p_3 & p_2 \\ (1 + 6p_1)p_3 & (1 + 6p_1)p_2 \end{pmatrix}. \tag{3}$$

The null-space of the matrix in Eq. (3) is $\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(x(0), \theta)\right) = \left\{-\frac{p_2}{p_3}, 1\right\}$. Entries of the base-vector of this nontrivial null-space are the coefficients of the partial differential equation that describes the linear dependence between the 2 columns of this Jacobi matrix. This partial differential equation for some function $\phi(p_2, p_3)$ reads as

$$-\frac{p_2}{p_3} \frac{\partial \phi}{\partial p_2} + \frac{\partial \phi}{\partial p_3} = 0. \tag{4}$$

A possible solution to Eq. (4) is $\phi = p_2 p_3$. One option for reinstating this model’s identifiability is to reduce the number of parameters from 3 (p_1, p_2, p_3) to 2 (p_1, ϕ). By introducing the scaled variable $\tilde{x}_2 \equiv x_2/p_3$, we obtain the following reparameterised, identifiable model:

$$\frac{dx_1}{dt} = p_1 x_1^2 + \phi x_1 \tilde{x}_2, \quad x_1(0) = x_{10} \neq 0, \tag{5}$$

$$\frac{d\tilde{x}_2}{dt} = x_1^2 + x_1 \tilde{x}_2, \quad \tilde{x}_2(0) = 0. \tag{6}$$

In view of the topic of this paper, we set out to find an alternative option for reinstating this model’s identifiability by investigating the role of the initial conditions. It is easy to confirm that the model’s unidentifiability can also be turned into identifiability simply by setting $x_2(0) \neq 0$. The Jacobi matrix computed for the scenario where $x_2(0) \neq 0$ is given in Eq. (7). Similar to Eq. (3), its 2 columns are related to system parameters p_2 and p_3 respectively. In contrast with the matrix in Eq. (3), the matrix in Eq. (7) has rank 2, so the linear dependence between its 2 columns is destroyed by setting $x_2(0) \neq 0$.

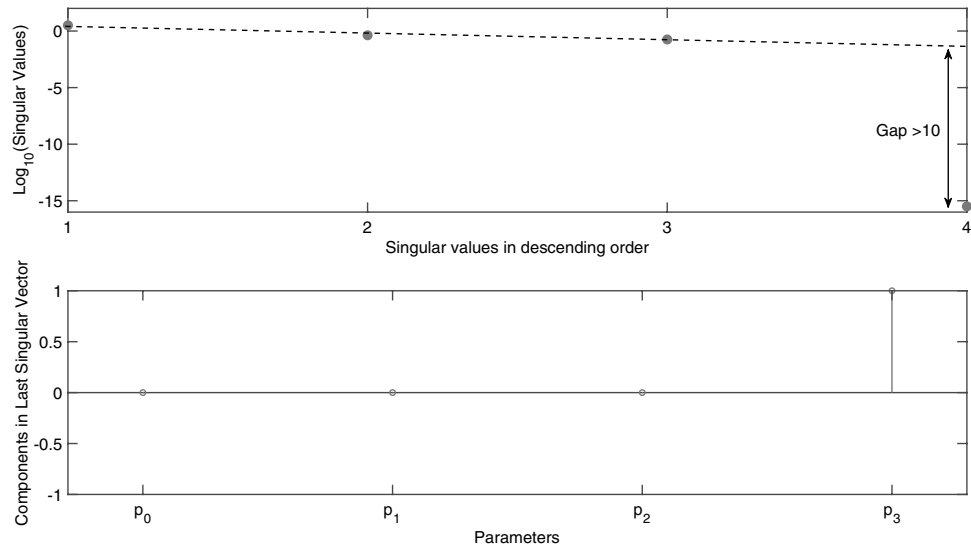


Figure 2. Identifiability signature of the benchmark model with input (M_2). (Top) Singular values: When state x_1 is measured and $x_2(0) = p_1/p_3$, the clear gap between the third and fourth singular values suggests rank deficiency of the sensitivity matrix. (Bottom) Entries in the last column of the right singular matrix: The nonzero entry corresponding to p_3 indicates that parameter p_3 is structurally unidentifiable.

$$\frac{dG}{d\theta^{unid}}(x(0), \theta) = \begin{pmatrix} 0 & 0 & 0 \\ x_1(0)x_2(0) & 0 & 0 \\ x_1(0)(p_3x_1(0)^2 + x_2(0)(x_1(0) + 3p_1x_1(0) + 2p_2x_5(0))) & p_2x_1(0)^2 & 0 \end{pmatrix}. \quad (7)$$

Example 2 Benchmark model with input signal (M_2)

We now turn our attention to a benchmark model that has an input signal and was previously analysed by Saccomani et al.⁴. Here we analyse the local structural identifiability of the 4 unknown system parameters and so $\theta = [p_0, p_1, p_2, p_3]$. This example shows that problematic initial conditions are not necessarily zero values.

$$\frac{dx_1}{dt} = -p_0u - p_2x_1 - p_3x_2, \quad x_1(0) = x_{10}, \quad (8)$$

$$\frac{dx_2}{dt} = p_3x_1x_2 - p_1x_1, \quad x_2(0) = \frac{p_1}{p_3}. \quad (9)$$

Given the measured output $y = x_1$, Saccomani et al. show that when $x_2(0) = p_1/p_3$, p_3 is structurally unidentifiable. Our numerical results shown in Fig. 2, corroborate this result. Figure 2 shows entries of the last column of the V matrix which related to the singular value beyond the gap. Its nonzero entry shows that parameter p_3 is not identifiable. To verify this numerical result symbolically, one begins by computing a set of Fliess series coefficients using Eq. (59) defined in the “Methods” section, $G(x(0), \theta) = \{x_1(0), -p_2x_1(0) - p_3x_2(0), -p_0p_2, -p_0p_2^2 + p_0p_3(-p_1 + p_3x_2(0))\}$. Next, one substitutes the initial condition $x_2(0) = p_1/p_3$ into this series, and calculates partial derivatives of G with respect to the unidentifiable parameter. For illustration, we compute the Jacobi matrix in this example with respect to all 4 system parameters. The columns are related to the parameters p_0, p_1, p_2 , and p_3 , respectively. Here we show the 5×4 matrix,

$$\frac{dG}{d\theta}(x(0), \theta) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ -p_2 & 0 & -p_0 & 0 \\ -p_2^2 & 0 & -2p_0p_2 & 0 \\ -p_2^3 & 0 & -3p_0p_2^2 & 0 \end{pmatrix}. \quad (10)$$

The last column of Eq. (10) contains only zeros and accordingly system parameter p_3 is not structurally identifiable. The nontrivial null-space of the Jacobi matrix is $\mathcal{N}\left(\frac{dG}{d\theta}(\theta)\right) = \{0, 0, 0, 1\}$. Our method confirms that this model’s identifiability is reinstated when $x_2(0) \neq p_1/p_3$. For example, if we would choose as initial condition $x_2(0) = p_2/p_3$, we would obtain the Jacobi matrix in Eq. (11), which has rank 4. Accordingly, we conclude that for this choice all 4 system parameters are locally identifiable.

$$\frac{d\mathbf{G}}{d\boldsymbol{\theta}}(\mathbf{x}(\mathbf{0}), \boldsymbol{\theta}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ -p_2 & p_0 & 0 & 0 \\ -p_2^2 + (p_2 - p_1)p_3 & -2p_0p_2 + p_0p_3 & -p_0p_3 & p_0(p_2 - p_1) \\ p_2(p_2 - p_1)p_3 + (p_2 - p_1)p_3^2 & -p_0p_2^2 + p_0p_2p_3 & & \\ +p_2[-p_2^2 + (p_2 - p_1)p_3] & +2p_0(p_2 - p_1)p_3 + p_0p_3^2 & -2p_0p_2p_3 - p_0p_3^2 & 2p_0p_2(p_2 - p_1) \\ & +p_2(-2p_0p_2 + p_0p_3) & & +2p_0(p_2 - p_1)p_3 \end{pmatrix}. \quad (11)$$

Example 3 Erythropoietin (Epo) and Epo receptor (EpoR) interaction and trafficking (M_3)

We now analyse a real-world systems biology model. In erythroid progenitor cells, which give rise to erythrocytes (commonly known as red blood cells), the dynamic properties of the Epo receptors determine how signals in the concentration of the ligand Epo are processed at the receptor level. This ultimately indicates how downstream signalling cascades such as the JAK2-STAT5 pathway are activated, which in turn leads to cellular responses such as differentiation and proliferation of erythrocytes¹². The structural unidentifiability of this six state model which describes Erythropoietin (Epo) and Epo receptor (EpoR) interaction and trafficking was previously confirmed¹². Given that four of the six initial conditions are zero, the aim here is to see whether changing these values from an experimental design perspective could address this model's unidentifiability. The six model equations are^{12,13},

$$\frac{d[Epo]}{dt} = -k_{on}[Epo][EpoR] + k_{on}k_D[Epo_EpoR] + k_{ex}[Epo_EpoR_i], \quad (12)$$

$$\frac{d[EpoR]}{dt} = -k_{on}[Epo][EpoR] + k_{on}k_D[Epo_EpoR] + k_t[EpoR(0)] - k_t[EpoR] + k_{ex}[Epo_EpoR_i], \quad (13)$$

$$\frac{d[Epo_EpoR]}{dt} = k_{on}[Epo][EpoR] - k_{on}k_D[Epo_EpoR] - k_e[Epo_EpoR], \quad (14)$$

$$\frac{d[Epo_EpoR_i]}{dt} = k_e[Epo_EpoR] - k_{ex}[Epo_EpoR_i] - k_{di}[Epo_EpoR_i] - k_{de}[Epo_EpoR_i], \quad (15)$$

$$\frac{d[dEpo_i]}{dt} = k_{di}[Epo_EpoR_i], \quad (16)$$

$$\frac{d[dEpo_e]}{dt} = k_{de}[Epo_EpoR_i]. \quad (17)$$

The measured output defined by Raue et al.¹² contains one additional unknown scaling parameter, *scale*:

$$y_1 = scale([Epo] + [dEpo_e]), \quad (18)$$

$$y_2 = scale[Epo_EpoR]. \quad (19)$$

The initial values of [Epo] and [EpoR] are assumed to be unknown and so $\boldsymbol{\theta}$ contains these and the 8 system parameters: $\boldsymbol{\theta} = [k_{on}, k_D, k_{ex}, k_t, k_e, k_{di}, k_{de}, scale, [Epo](0), [EpoR](0)]$. The initial conditions of all the remaining model states are zero:

$$\mathbf{x}(\mathbf{0}) = [[Epo](0), [EpoR](0), 0, 0, 0, 0]. \quad (20)$$

The structural and practical identifiability of this model was assessed in a 2010 paper by calculating the profile likelihood related to each of its 10 unknown parameters¹². Five structurally unidentifiable parameters were identified. Their identifiability was reinstated by assuming the value of [Epo](0) to be known¹².

An alternative way to reinstate the model's structural identifiability is the addition of one or more sensors to the model's measured output. Which sensors might be added can efficiently be solved by determining a model's minimal sensor set, the minimal set of sensors that needs to be measured to ensure model identifiability¹⁴. In this example, the identifiability can be reinstated by adding either state [dEpo_i] or [dEpo_e] to its measured output. In other words, measuring either $\mathbf{y} = \{scale.([Epo] + [dEpo_e]), scale.[Epo_EpoR], [dEpo_i]\}$ or $\mathbf{y} = \{scale.([Epo] + [dEpo_e]), scale.[Epo_EpoR], [dEpo_e]\}$ would ensure model identifiability. The physical limitations associated with this experimental implementation are not considered here.

To understand the role this model's initial conditions play in its unidentifiability, we analyse the model for the conditions stipulated in Eq. (20), with both [Epo](0), [EpoR](0) assumed to be nonzero. The resulting identifiability signature is given in Fig. 3. The results indicate that $\boldsymbol{\theta}^{unid} = \{k_{on}, k_D, scale, [Epo](0), [EpoR](0)\}$ are

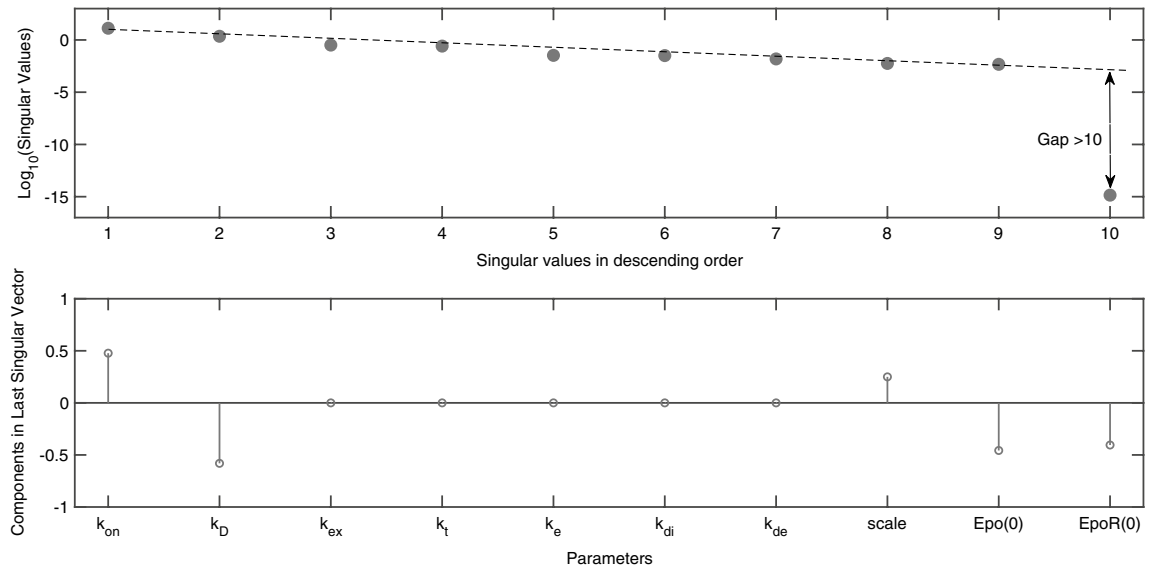


Figure 3. Identifiability signature of the Erythropoietin (Epo) and Epo receptor (EpoR) model (M_3). (Top) Singular values: When output $y = \{y_1, y_2\}$ is measured, and for the values defined in Eq. (20), with both $[Epo](0)$, $[EpoR](0)$ nonzero. The clear gap between the spectrum of singular values suggests that the model is not structurally identifiable. (Bottom) Entries in the last column of the V matrix related to the singular value beyond the gap: Nonzero entries indicate that parameters k_{on} , k_D , $scale$, $[Epo](0)$ and $[EpoR](0)$ are structurally unidentifiable and apparently totally correlated.

Experimental scenario	Number of nonzero initial values	Nonzero initial value
1	1	$[Epo_EpoR](0)$
2	1	$[Epo_EpoR_i](0)$
3	3	$[Epo](0), [EpoR](0), [Epo_e](0)$

Table 1. List of plausible nonzero initial value combinations that ensure the structural identifiability of the Erythropoietin (Epo) and Epo receptor (EpoR) model (M_3).

unidentifiable. This result is symbolically verified by the base-vector spanning the nontrivial null-space: $\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right) = \left\{ -\frac{k_{on}}{[EpoR](0)}, \frac{k_D}{[EpoR](0)}, -\frac{scale}{[EpoR](0)}, \frac{[Epo](0)}{[EpoR](0)}, 1 \right\}$.

We suspect that certain of the zero initial conditions reduce the dynamic information required to estimate accurate parameter values. To assess whether or not certain initial conditions contribute to the unidentifiability of the five parameters, we apply our numerical analysis of the model, each time selecting different combinations of nonzero initial values. The numerous iterations required in this analysis are made possible by the computational efficiency of the algorithm⁹. Assuming that the two unknown initial values, $[Epo](0)$ and $[EpoR](0)$ can also be zero, we find that the model's identifiability can indeed be reinstated by setting certain initial conditions at nonzero values. The results in Table 1 reveal that there are three plausible experimental setups which would restore model identifiability. Scenarios 1 and 2 require that only a single initial condition holds a nonzero value. The first scenario is associated with the measured sensor in (19). In the third, a set of three specific nonzero initial conditions is defined. The Jacobi matrices for these three scenarios can be computed similarly as done in Examples 1 and 2, given in Eqs. (7) and (11), respectively.

Example 4 JAK/STAT model (M_4).

Here, we consider the well-known unidentifiable JAK/STAT model^{15,16}. No literature has been published investigating the source of its unidentifiability and we address this question here by investigating the role of its initial conditions defined as¹⁵

$$x(0) = [x_1(0), \dots, x_{14}(0)] = [1.3, x_2(0), 0, 0, 0, 2.8, 0, 165, 0, 0, 0.34, 0, 0, 0]. \tag{21}$$

The 14 model equations are¹⁵

$$\dot{x}_1 = -\theta_1 u_1 c_1 x_1 - \theta_5 x_1 + \theta_6 x_2, \tag{22}$$

$$\dot{x}_2 = \theta_5 x_1 - \theta_6 x_2, \tag{23}$$

$$\dot{x}_3 = \theta_1 u_1 c_1 x_1 - \theta_2 x_3 x_7, \quad (24)$$

$$\dot{x}_4 = \theta_2 x_3 x_7 - \theta_3 x_4, \quad (25)$$

$$\dot{x}_5 = \theta_3 x_4 - \theta_4 x_5, \quad (26)$$

$$\dot{x}_6 = -\frac{\theta_7 x_3 x_6}{(1 + \theta_{13} x_{13})} - \frac{\theta_7 x_4 x_6}{(1 + \theta_{13} x_{13})} + \theta_8 c_2 x_7, \quad (27)$$

$$\dot{x}_7 = \frac{\theta_7 x_3 x_6}{(1 + \theta_{13} x_{13})} + \frac{\theta_7 x_4 x_6}{(1 + \theta_{13} x_{13})} - \theta_8 c_2 x_7, \quad (28)$$

$$\dot{x}_8 = -\theta_9 x_8 x_7 + c_2 \theta_{10} x_9, \quad (29)$$

$$\dot{x}_9 = \theta_9 x_8 x_7 - c_2 \theta_{10} x_9, \quad (30)$$

$$\dot{x}_{10} = \theta_{11} x_9, \quad (31)$$

$$\dot{x}_{11} = -\theta_{12} c_1 u_1 x_{11}, \quad (32)$$

$$\dot{x}_{12} = \theta_{12} c_1 u_1 x_{11}, \quad (33)$$

$$\dot{x}_{13} = \frac{\theta_{14} x_{10}}{(\theta_{15} + x_{10})} - \theta_{16} x_{13}, \quad (34)$$

$$\dot{x}_{14} = \theta_{17} x_9. \quad (35)$$

The defined output contains additional parameters $\theta_{18}, \dots, \theta_{22}$

$$y_1 = x_1 + x_3 + x_4, \quad (36)$$

$$y_2 = \theta_{18}(x_3 + x_4 + x_5 + x_{12}), \quad (37)$$

$$y_3 = \theta_{19}(x_4 + x_5), \quad (38)$$

$$y_4 = \theta_{20} x_7, \quad (39)$$

$$y_5 = \theta_{21} x_{10}, \quad (40)$$

$$y_6 = \theta_{22} x_{14}, \quad (41)$$

$$y_7 = x_{13}, \quad (42)$$

$$y_8 = x_9. \quad (43)$$

With $x_2(0)$ assumed unknown, the identifiability of 23 unknown parameters in total must be analysed. The numerical results corroborate that this model is indeed unidentifiable for the initial conditions defined in Eq. (21), where $x_2(0) \neq 0$. This is apparent from the significant gap in the spectrum of singular values in Fig. 4. The two singular values beyond this gap suggest that the null-space contains two base-vectors and that there are two sets of totally correlated parameters. The unidentifiable parameters are the nonzero elements in the bottom of Fig. 4 and so the union of the elements in these sets is $\theta^{und} = \{\theta_{11}, \theta_{15}, \theta_{17}, \theta_{21}, \theta_{22}\}$. The 2 base-vectors spanning the nontrivial null-space: $\mathcal{N}\left(\frac{dG}{d\theta^{und}}(\theta)\right) = \{0, 0, -\theta_{17}/\theta_{22}, 0, 1\}$ and $\{-\theta_{11}/\theta_{21}, -\theta_{15}/\theta_{21}, 0, 1, 0\}$, give a clear indication as to which parameters are totally correlated. Using numerical results, the symbolic calculations associated with this example are significantly reduced since we only need to compute and analyse a Jacobi matrix with 5 columns instead of 23.

Systematically changing one or more of the zero values of the initial conditions in Eq. (21), we find that the identifiability is reinstated if we choose $x_{10}(0) \neq 0$ and $x_{14}(0) \neq 0$, leaving the other conditions at zero values. In hindsight, this conclusion can be understood from the output equations, $y_5 = \theta_{21} x_{10}$ and $y_6 = \theta_{22} x_{14}$, which contain the unidentifiable parameters θ_{21} and θ_{22} , respectively. Taking $x_{10}(0) \neq 0$ and $x_{14}(0) \neq 0$ will ensure that sufficient dynamics is observed to enable the accurate estimation of all unknown parameters. We remark that setting $x_{10}(0) \neq 0$ destroys the base-vector $\{0, 0, -\theta_{17}/\theta_{22}, 0, 1\}$ and setting $x_{14}(0) \neq 0$ the base-vector $\{-\theta_{11}/\theta_{21}, -\theta_{15}/\theta_{21}, 0, 1, 0\}$.

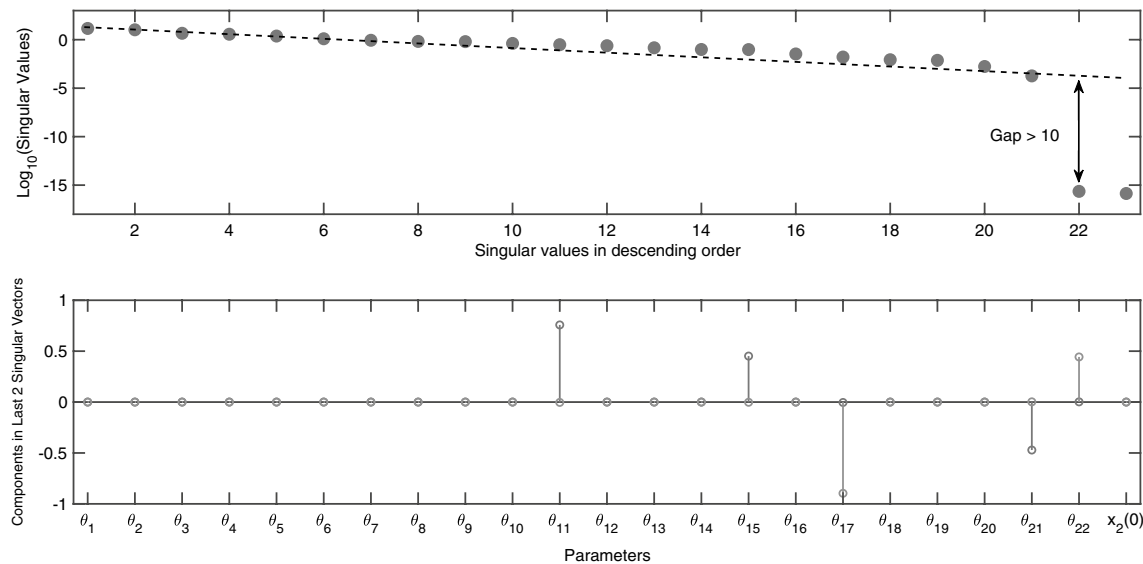


Figure 4. Identifiability signature of the JAK/STAT model (M_4). (Top) The 2 singular values beyond the gap calculated whilst measuring Eqs. (36)–(43) and using the initial conditions defined in Eq. (21). These suggest model unidentifiability and that there are two sets of totally correlated parameters. (Bottom) Entries in the last two columns of matrix V related to the 2 zero singular values. The nonzero elements indicate the union between two potential sets of totally correlated parameters, $\{\theta_{11}, \theta_{15}, \theta_{17}, \theta_{21}, \theta_{22}\}$.

Example 5 Pharmacokinetics model (M_5).

Our final example contains only four model states and its global identifiability has been a topic of investigation since 2005. Unlike the previous four examples, this model is identifiable for the defined set of initial conditions. However, this does not imply that this holds for all initial conditions. Below we shall show that this model is an excellent example of the statement from Maria Saccomani, “It happens frequently in the global identifiability applications that the property holds only generically, i.e. except for a ‘thin’ set of initial conditions. In these situations the system is (incorrectly but forgivably) nevertheless declared to be (global) identifiable, excluding certain subsets of initial states”⁴.

We suspect that the model has sets of zero-valued problematic initial conditions that should be avoided during experimental design. To identify these sets, we perform an iterative search for zero-valued initial conditions which might render the model *unidentifiable*. This search is not exhaustive in the sense that we would search for all problematic combinations of initial values, as our aim is to find “a thin set of values” for which this model’s global identifiability as recorded in the literature, does not hold.

The model which comprises 4 model equations describes the ligands of the macrophage mannose receptor¹⁷

$$\frac{dx_1}{dt} = \alpha_1(x_2 - x_1) - \frac{k_a v_m x_1}{(k_c k_a + k_c x_3 + k_a x_1)}, \quad x_1(0) = x_{10}, \quad (44)$$

$$\frac{dx_2}{dt} = \alpha_2(x_1 - x_2), \quad x_2(0) = 0, \quad (45)$$

$$\frac{dx_3}{dt} = \beta_1(x_4 - x_3) - \frac{k_c v_m x_3}{(k_c k_a + k_c x_3 + k_a x_1)}, \quad x_3(0) = x_{30}, \quad (46)$$

$$\frac{dx_4}{dt} = \beta_2(x_3 - x_4). \quad x_4(0) = 0. \quad (47)$$

The first state is measured, so the model’s output is defined as $y = x_1$. State x_1 represents the plasma enzyme concentration, x_2 its concentration in compartment 2, x_3 is the plasma concentration of the mannose polymer that acts as a competitor of glucose oxidase for the mannose receptor of macrophages, and x_4 is the concentration of this competitor in the extra vascular fluid of the organs accessible to this macro molecule¹⁸.

A 2005 publication¹⁷ on the topic of identifiability analysed this model using the differential algebra method. This method requires that functions f and h , defined in the general descriptions Eqs. (49) and (51), be rational. This is clearly the case for the present model. The model was found to be globally identifiable. The analysis comprised two steps, where the unknown parameters were divided into two subsets $\{\alpha_1, \alpha_2, V_m, k_c\}$ and $\{\beta_1, \beta_2, k_a\}$, respectively. In this analysis the initial conditions were not taken into account. In a 2010 publication¹⁹ the model was reported to be globally identifiable, but only under the assumption that parameter α_2 was known. No results could be obtained for the case with α_2 unknown. The model was once more included in a 2011 publication¹⁸ which compared seven different identifiability analysis approaches. The local structural identifiability of six of

Number of zero-valued initial conditions	States involved
2	$x_1(0) = x_2(0) = 0$
2	$x_3(0) = x_4(0) = 0$

Table 2. Zero-valued initial condition combinations that result in the *loss* of structural identifiability of the Pharmacokinetics model (M_5).

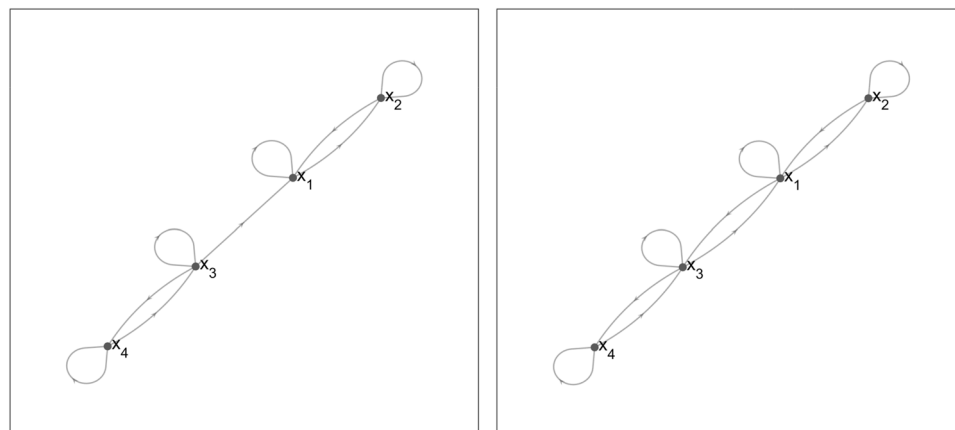


Figure 5. Directed graph of the Pharmacokinetics model (M_5). (Left) For the scenario $x_3(0) = x_4(0) = 0$. The model is divided into 2 strongly connected components. (Right) For the scenario $x_3(0) \neq 0$ and $x_4(0) = 0$. The additional connection between nodes x_1 and x_3 reduces the model down to a single component. The result is that a greater amount of information is transferred between the individual model states.

the seven system parameters $\alpha_1, k_a, V_m, k_c, \beta_1$ and β_2 , could be confirmed with the Taylor series method. In a recent publication²⁰, the global identifiability result of Saccomani et al. was confirmed, this time also including the four initial conditions to the set of unknown parameters.

In our analysis we include α_2 and all initial conditions as unknown parameters, so we take as vector of parameters to be estimated: $\theta = [\alpha_1, \alpha_2, k_a, V_m, k_c, \beta_1, \beta_2, x_1(0), x_2(0), x_3(0), x_4(0)]$. We proceed by performing an iterative search to identify problematic zero-valued initial conditions. Table 2 contains the initial value combinations for which we found the model to be not identifiable. These combinations may perhaps not be realistic in the context of this particular example. However, the emphasis here is to illustrate that the notion of global identifiability does not hold for this model. It also illustrates that the fast numerical method used in this paper allows for these kind of searches.

Figure 5 shows the change in the structure of the directed graph for the scenario where $x_3(0) = x_4(0) = 0$ (left). For this case, the graph is divided into two strongly connected components with state x_1 and x_2 forming the first component and x_3 and x_4 the second. Notice that when $x_3(0) \neq 0$, that there is an additional connection between the nodes related to states x_1 and x_3 (right). This reduces the model structure down to one single component and therefore when $x_3(0) \neq 0$, information can flow from the measured state x_1 to the model equations pertaining to x_3 and x_4 . As apparent from this directed graph, the identifiability signature in Fig. 6 shows that parameters k_a, β_1 and β_2 , of which the latter two are exclusively related to the differential equations of states x_3 and x_4 , are not identifiable. This result is corroborated by the symbolically computed 5×7 Jacobi matrix, in which each column is related to 1 of the 7 unknown system parameters, $\alpha_1, k_a, V_m, k_c, \alpha_2, \beta_1, \beta_2$. Since the columns related to parameters k_a, β_1 and β_2 contain only zero elements, we may immediately conclude that these parameters cannot be estimated. We symbolically calculated the base-vectors spanning the nontrivial null-space and found that $\mathcal{N}\left(\frac{dG}{d\theta}(\theta)\right) = \{0, 1, 0, 0, 0, 0, 0\}, \{0, 0, 0, 0, 0, 1, 0\}$, and $\{0, 0, 0, 0, 0, 0, 1\}$. This tells us that there is no other totally correlated set of parameters. This is also in complete agreement with the identifiability signature in Fig. 6 which shows precisely 3 numerically zero-valued singular values. This result was obtained within 0.5 s on an Intel Core i7 processor with 8 GB RAM using the application described in¹⁰. This is comparable with EAR when the initial conditions are *not* parametrised. For parametrised initial conditions, differences in computation times were observed in favour of the sensitivity-based application from¹⁰.

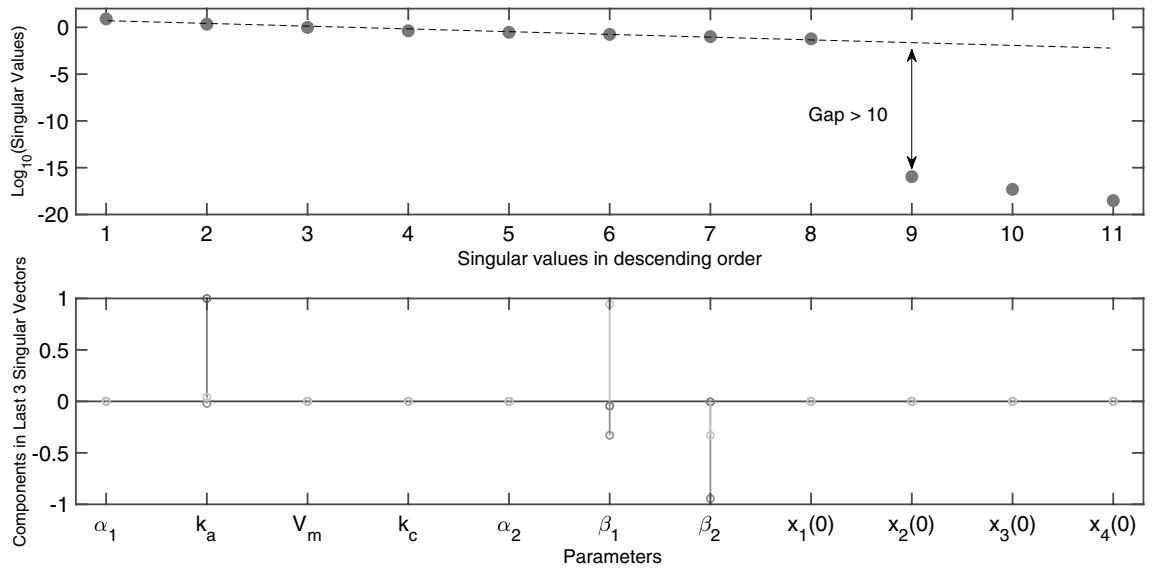


Figure 6. Identifiability signature of the pharmacokinetics model (M_5) for the experimental setup with $x_3(0) = x_4(0) = 0$. (Top) The three singular values beyond the gap suggest that the sensitivity matrix is rank deficient and so the model is not identifiable. (Bottom) Entries in the last 3 columns of the V matrix, related to the 3 zero valued singular values. The nonzero entries suggest that parameters $\{k_a, \beta_1, \beta_2\}$ are unidentifiable.

$$\frac{dG}{d\theta}(x(0), \theta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -x_1(0) + x_2(0) & 0 & \frac{-x_1(0)}{(k_c + x_1(0))} & \frac{V_m x_1(0)}{(k_c + x_1(0))^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{((2\alpha_1 + \alpha_2)(k_c + x_1(0))^2(x_1(0) - x_2(0)) + V_m(2k_c x_1(0) + x_1(0)^2 - k_c x_2(0)))}{(k_c + x_1(0))^2} & 0 & \dots & \dots & \alpha_1(x_1(0) - x_2(0)) & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & 0 & \dots & \dots & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & 0 & \dots & \dots & \dots & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Discussion

The results in this paper show the importance of including initial values in the identifiability analysis of any model. We showed how to identify problematic initial conditions that may result in loss of identifiability. The insight gained from this approach provides modellers with an extra tool to turn an unidentifiable model into a identifiable one, e.g., by avoiding initial values that may cause problems. This is especially useful during the design of the experiments. We also showed that thanks to the algorithm’s computational efficiency, one can detect these problematic values not only for small toy models but also for realistic system biology models that are usually fairly large.

Highlights include the identification of problematic initial conditions of the well-known JAK/STAT model comprising 14 model equations (Example 4) and the identification of problematic values that result in the loss of local identifiability of a Pharmacokinetics model, a model classified as globally identifiable in the past (Example 5).

Here, we mainly used our method to identify problematic sets of zero-valued initial conditions. The computational efficiency of the method allows for extensive searches in the space of initial conditions. However, nonzero initial conditions may also cause loss of identifiability and then the question that begs to be answered is “How does one go about identifying potential problematic sets of nonzero initial values?” This remains a challenging problem since in general the space of initial conditions is infinitely dimensional. The insight that the problematic sets form a thin subspace may be of help here. To find such manifolds one could think of first detecting one problematic point, after which the rest of the manifold could be traced via continuation. Another promising avenue to be pursued in the future is the investigation of systems in *steady state*. It is well-known that the estimation of certain parameters of such systems might be problematic, since the data do not contain enough informative dynamics for the accurate parameter estimation. Initial conditions that give rise to a steady state could thus rather be avoided. Given the size of modern systems biology models, any future research requires an algorithm that can quickly analyse identifiability. The approach followed in this paper provides such an algorithm.

As a final point, we wish to mention that caution should always be taken when analysing and giving judgement on a model’s structural identifiability. Given that this is a difficult property to analyse, for non-linear models in particular, no single method exists that can be applied to all models without fail. As an example, consider the following 1 state model:

$$\frac{dx_1}{dt} = 1 + \theta_1 x_1^2, \quad x_1(0) = 0. \tag{48}$$

For the defined output $y = x_1$, some of the methods mentioned in the introduction incorrectly classify parameter θ_1 as unidentifiable, when in actual fact it is identifiable. By first applying our numerical step, one finds that the model is indeed identifiable and accordingly, no further analyses are required, eliminating the risk of obtaining

incorrect results. This proves that using different methods in tandem will not only minimise computational times but will also improve computational correctness.

Methods

Model description. Ordinary differential equation models may be written in standard state-space form²¹:

$$\dot{\mathbf{x}}(t) = \mathbf{f}_0(\mathbf{x}(t), \boldsymbol{\theta}) + \sum_{i=1}^k u_i(t) \mathbf{f}_i(\mathbf{x}(t), \boldsymbol{\theta}), \quad (49)$$

$$\mathbf{x}(0) = \mathbf{x}_0, \quad (50)$$

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \boldsymbol{\theta}). \quad (51)$$

State variables are contained in vector $\mathbf{x}(t)$ with dimension n , system parameters in vector $\boldsymbol{\theta}$ ($\dim(\boldsymbol{\theta}) = p$) and the measured model outputs in vector $\mathbf{y}(t)$ ($\dim(\mathbf{y}) = m$). The initial values of states may be unknown and in such instances, the initial condition vector may be parameterised through some additional parameters that then become part of the identification problem. The resulting unknown parameter vector then has $\dim(\boldsymbol{\theta}) = p + n$ ²².

The state vector, \mathbf{x} , evolves in time in R^n . Functions \mathbf{f}_i , $i = 0, \dots, k$, and \mathbf{h} are assumed to be analytical and C^∞ functions, so that their partial derivatives of any order exist and are continuous²³. Finally, a model's input functions are contained in the vector $\mathbf{u}(t) \equiv \{u_1, \dots, u_k\}$ ²².

Local structural identifiability analysis. The identifiability detection method we propose combines both numerical and symbolic analyses. The methods starts with a numerical analysis of the rank of the so-called sensitivity matrix. This initial step pinpoints potentially unidentifiable parameters and initial conditions. In a second step a symbolic analysis is performed to check whether these parameters and initial conditions are indeed causing identifiability problems. In this discussion we deal with both steps separately. Note that at its core both the factorisation of the sensitivity matrix via SVD and the computation of the null-space of the Jacobi matrix have the same task: i.e. identifying linear dependencies between the partial derivatives of the output sensors with respect to the unknown parameters. The difference being that the sensitivity matrix represents the functions via values at different points whilst the matrix generated using Lie derivatives uses values of the derivatives at one point.

Since the second, symbolic step of our method usually involves only the analysis of a restricted number of parameters and initial conditions, the computational demand which is often found to be the curse of a symbolic analysis of the full model, is significantly reduced. The numerical results obtained from the first step can attractively be summarised in a so-called “identifiability signature”. This signature contains of a graphical presentation of the singular values resulting from a Singular Value Decomposition (SVD) of the sensitivity matrix. Structural unidentifiability is indicated by a *clear gap in the displayed singular values*¹¹. Furthermore, the signature represents the components of the singular vectors corresponding with the close-to-zero singular values. The nonzero components reveal which parameters and initial conditions are expected to be problematic.

Numerical analysis. The numerical step of our method uses the sensitivity matrix function $\partial \mathbf{y} / \partial \boldsymbol{\theta}$ of the model output with respect to individual unknown model parameters. These sensitivities are calculated using the following 2 equations⁹:

$$\frac{d}{dt} \left(\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} \right) = \frac{\partial \mathbf{f}_0}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{f}_0}{\partial \boldsymbol{\theta}} + \sum_{i=1}^k \left(\frac{\partial \mathbf{f}_i}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{f}_i}{\partial \boldsymbol{\theta}} \right) u_i, \quad (52)$$

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}. \quad (53)$$

One obtains $\partial \mathbf{y} / \partial \boldsymbol{\theta}$ as a function of time by simultaneously integrating Eqs. (49) and (52) and substituting the solution into Eq. (53)²⁴. By evaluating these sensitivities at discrete time points on an interval $[t_0, \dots, t_N]$ one constructs a sensitivity matrix, \mathcal{S} . Matrix \mathcal{S} has $p + n$ columns when all initial values of model states are also unknown, with each column related to a specific parameter or initial condition, θ_i , $i = 1, \dots, p + n$. The sensitivity matrix thus reads as:

$$\mathbf{S}(t_0, \dots, t_N, \boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial y_1}{\partial \theta_1}(t_0) & \dots & \frac{\partial y_1}{\partial \theta_{p+n}}(t_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial \theta_1}(t_0) & \dots & \frac{\partial y_m}{\partial \theta_{p+n}}(t_0) \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial \theta_1}(t_N) & \dots & \frac{\partial y_1}{\partial \theta_{p+n}}(t_N) \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial \theta_1}(t_N) & \dots & \frac{\partial y_m}{\partial \theta_{p+n}}(t_N) \end{pmatrix}. \quad (54)$$

A full ranked matrix \mathbf{S} is a sufficient condition for local structural identifiability^{25,26}. The rank deficiency of the sensitivity matrix can be attributed to two factors: (1) an output may be insensitive to a specific parameter and so all entries in the matrix column pertaining to this parameter are zero. Accordingly the parameter is not identifiable. This phenomenon is observed in Examples 2 and 5 in the “Results” section. (2) Alternatively, a model output may be sensitive to a particular parameter, but this sensitivity is related to the sensitivity of the output to one or more other parameters²⁴. The result is that certain columns of the sensitivity matrix are linearly dependent and so the parameters are totally correlated and not identifiable²⁷. This can be seen in Examples 2, 3 and 4 in the “Results” section. We determine the numerical rank of the sensitivity matrix by applying a Singular Value Decomposition (SVD), in which the matrix \mathbf{S} is written as a sum of equally sized matrices that decrease in dominance¹¹:

$$\mathbf{S}(t_0, \dots, t_N, \boldsymbol{\theta}) = u_1 \sigma_1 v_1^T + \dots + u_{p+n} \sigma_{p+n} v_{p+n}^T. \quad (55)$$

If all parameters and initial conditions are involved, there are $p + n$ singular values σ_i , arranged in descending order. The rank of \mathbf{S} is given by the number of nonzero singular values. Therefore zero-valued singular values indicate the rank-deficiency of \mathbf{S} ²⁸. Due to numerical errors, singular values are seldom exactly zero and accordingly the following practical definition is used: zero-valued singular values are values that fall beyond a *distinct gap* in the spectrum of singular values²⁹. It is up to the user to define a reliable width for the gap. In our examples, we take 10 decades on the logarithmic scale as reliable gap width. The presence of close-to-zero singular values, which are located beyond such a gap, indicate that the model may be unidentifiable. The parameters and initial values that may be involved follow from the nonzero entries in the columns, v_i , of the right singular matrix, that correspond with the close-to-zero singular values. The singular values and the unidentifiable parameters are graphically illustrated in an identifiability signature¹¹. We show the signatures of each of the models in the 5 Examples in the Results section (cf. Fig. 1).

Symbolic calculations. The second step in our procedure is to verify the numerical results symbolically. This entails the symbolic calculation of the Jacobi matrix of a model. The computational demand often associated with computing this matrix is greatly reduced by using the preceding numerical outcomes⁹. Available software packages that can be used include amongst others: COMBOS³⁰ and GenSSI2.0³¹. We use Lie derivatives and accordingly, only compute derivatives of the Lie derivatives with respect to the parameters that are suggested to be unidentifiable from the numerical analysis. We indicate them by $\boldsymbol{\theta}^{unid}$. We use the rank condition for local structural identifiability presented by Tunali and Tarn³².

The Jacobi matrix of a model with no control inputs ($u_1, u_2, \dots, u_k = 0$) can be computed using Lie derivatives, where a Lie derivative is the directional derivative of the smooth function, \mathbf{h} , in the direction of the drift vector field, \mathbf{f}_0 ²⁴. Mathematically it is defined as

$$\mathcal{L}_{\mathbf{f}_0} \mathbf{h} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \mathbf{f}_0. \quad (56)$$

Successive Lie derivatives are calculated as

$$\mathcal{L}_{\mathbf{f}_0}^i \mathbf{h} = \frac{\partial \mathcal{L}_{\mathbf{f}_0}^{i-1} \mathbf{h}}{\partial \mathbf{x}} \mathbf{f}_0. \quad (57)$$

The symbolic algebra package Kwatny’s ProPac add-on for Mathematica can be used to calculate these Lie derivatives²². In a generating series expansion, successive Lie derivatives of the output vector function \mathbf{h} are calculated. Parameterising the unknown initial conditions and so regarding them as additional parameters, the Jacobi matrix may have up to $p + n$ columns, each related to an individual parameter. Finally, the Jacobi matrix is computed by calculating partial derivatives of the generating series coefficients with respect to the unknown parameters. The symbolic matrix associated with the analysis of all system parameters and initial conditions reads as⁹,

$$\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \mathbf{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathbf{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_0} \mathbf{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_0} \mathbf{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_0}^2 \mathbf{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_0}^2 \mathbf{h}}{\partial \theta_{p+n}} \\ \vdots & \cdots & \vdots \end{pmatrix}. \tag{58}$$

A sufficient condition for structural identifiability is that $\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta})$, defined in Eq. (58), has full rank $p + n^9$. A lower rank is equivalent to it having a nontrivial null-space³³. The elements in such a nontrivial null-space reveal which parameters and initial conditions are involved in one or more correlated sets.

When evaluating models of the form defined in Eqs. (49)–(51) with $u_1, u_2, \dots, u_k \neq 0$, individual input functions must be incorporated into the symbolic calculations^{21,34}. An output is now expanded in a so-called Fliess series³² with respect to time and *input signals*. The coefficients of such a series are $\mathbf{h}(\mathbf{x}(0), \boldsymbol{\theta})$ and

$$\mathcal{L}_{f_{j_0}} \dots \mathcal{L}_{f_{j_q}} \mathbf{h}(\mathbf{x}(t), \boldsymbol{\theta})|_0. \tag{59}$$

The notation f_{j_0}, \dots, f_{j_q} represents all possible combinations of the vector fields $\{f_j, j = 0, \dots, k\}^{9,21}$. Furthermore, $|_0$ specifies that this Jacobi matrix is evaluated in the point $\mathbf{x}(0)$. It is clear that this procedure leads to a fast expanding number of terms if the values of k (the number of input signals) and p (the number of system parameters) increase. To give an impression of the Jacobi matrix associated with the full model in Eq. (52) for $k = 1$, we restrict the number of parameters to 1, e.g. θ_1 , and the number of initial values also to 1, e.g. x_0 . In that very basic case, the Jacobi matrix reads as⁹:

$$\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathbf{h}}{\partial x_1(0)} \\ \frac{\partial \mathcal{L}_{f_0} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_0} \mathbf{h}}{\partial x_1(0)} \\ \frac{\partial \mathcal{L}_{f_1} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_1} \mathbf{h}}{\partial x_1(0)} \\ \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathbf{h}}{\partial x_1(0)} \\ \frac{\partial \mathcal{L}_{f_0} \mathcal{L}_{f_1} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_0} \mathcal{L}_{f_1} \mathbf{h}}{\partial x_1(0)} \\ \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathbf{h}}{\partial x_1(0)} \\ \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathcal{L}_{f_0} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathcal{L}_{f_0} \mathbf{h}}{\partial x_1(0)} \\ \vdots & \vdots \\ \frac{\partial \mathcal{L}_{f_{j_0}} \dots \mathcal{L}_{f_{j_q}} \mathbf{h}}{\partial \theta_1} & \frac{\partial \mathcal{L}_{f_{j_0}} \dots \mathcal{L}_{f_{j_q}} \mathbf{h}}{\partial x_1(0)} \\ \vdots & \vdots \end{pmatrix}, \tag{60}$$

where $j_0, j_1, \dots, j_q \in [0, 1]$. Usually, a null-space will emerge if this matrix has only few rows. However, in the process of adding additional rows, two things may happen: either the basis vectors of this null-space are destroyed at some stage, indicating that the system is identifiable, or this null-space will persist regardless of the number of rows added, indicating that the system is structurally unidentifiable.

Code availability

MATLAB code is available at <https://sourceforge.net/projects/structural-identifiability/files/> with the user application available upon request from the author hans.stigter@wur.nl.

Received: 5 February 2021; Accepted: 5 August 2021

Published online: 19 August 2021

References

1. Yao, K. Z., Shaw, B. M., Kou, B., McAuley, K. B. & Bacon, D. W. Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. *Polymer React. Eng.* **11**, 563–588 (2003).

2. Raue, A. *et al.* Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929 (2009).
3. Denis-Vidal, L., Joly-Blanchard, G. & Noiret, C. Some effective approaches to check the identifiability of uncontrolled nonlinear systems. *Math. Comput. Simul.* **57**, 35–44. [https://doi.org/10.1016/S0378-4754\(01\)00274-9](https://doi.org/10.1016/S0378-4754(01)00274-9) (2001).
4. Saccomani, M. P., Audoly, S. & D'Angiò, L. Parameter identifiability of nonlinear systems: The role of initial conditions. *Automatica* **39**, 619–632. [https://doi.org/10.1016/S0005-1098\(02\)00302-3](https://doi.org/10.1016/S0005-1098(02)00302-3) (2003).
5. Villaverde, A. F. & Banga, J. R. Structural properties of dynamic systems biology models: Identifiability, reachability and initial conditions. *Processes* **5**(2), 29. <https://doi.org/10.3390/pr5-20029> (2017).
6. Karlsson, J., Anguelova, M. & Jirstrand, M. An efficient method for structural identifiability analysis of large dynamic systems. In *Proceedings of 16th IFAC Symposium on System Identification: The International Federation of Automatic Control, Brussels, Belgium* **45**, 941–946. <https://doi.org/10.3182/20120711-3-BE-2027.00381> (2012).
7. Bellu, G., Saccomani, M. P., Audoly, S. & D'Angiò, L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* **81**(1), 52–61. <https://doi.org/10.1016/j.cmpb.2007.07.002> (2007).
8. Villaverde, A. F., Barreiro, A. & Papachristodoulou, A. Structural identifiability of dynamic systems biology models. *PLOS Comput. Biol.* **20**, 1–22. <https://doi.org/10.1371/journal.pcbi.1005153> (2016).
9. Stigter, J. D. & Molenaar, J. A fast algorithm to assess local structural identifiability. *Automatica* **58**, 118–124. <https://doi.org/10.1016/j.automatica.2015.05.004> (2015).
10. Stigter, J. D. & Joubert, D. Computing measures of identifiability, observability, and controllability for a dynamic system model with the StrucID App. In *Proceedings of 19th IFAC Symposium System Identification: Learning Models for Decision and Control, Padova, Italy* (2021).
11. Stigter, J. D., Joubert, D. & Molenaar, J. Observability of complex systems: Finding the gap. *Sci. Rep.* **7**, 1–9. <https://doi.org/10.1038/s41598-017-16682-x> (2017).
12. Raue, A., Becker, V., Klingmüller, U. & Timmer, J. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos* **20**, 045105. <https://doi.org/10.1063/1.3528102> (2010).
13. Saccomani, M. P. Structural vs practical identifiability in system biology. 305–313. In *Proceedings of International Work-Conference on Bioinformatics and Biomedical Engineering IWBBIO, Grenada, Spain* (2013).
14. Joubert, D., Stigter, J. D. & Molenaar, J. Determining minimal output sets that ensure structural identifiability. *PLoS One* **13**, e0207334 (2018).
15. Raue, A., Karlsson, J., Saccomani, M. P., Jirstrand, M. & Timmer, J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **30**, 1440–1448. <https://doi.org/10.1093/bioinformatics/btu006> (2014).
16. Quaiser, T., Dittrich, A., Schaper, F. & Mönnigmann, M. A simple work flow for biologically inspired model reduction-application to early Jak-Stat signaling. *BMC Syst. Biol.* **30**. <https://doi.org/10.1186/1752-0509-5-30> (2011).
17. Verdière, N., Denis-Vidal, L., Joly-Blanchard, G. & Domurado, D. Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrophage mannose receptor. *Int. J. Appl. Math. Comput. Sci.* **15**, 517–526 (2005).
18. Chis, B. J. O. T. & Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE* **6**, e27755. <https://doi.org/10.1371/journal.pone.0027755> (2011).
19. Saccomani, M. P., Audoly, S., Bellu, G. & D'Angiò, L. Examples of testing global identifiability of biological and biomedical models with the DAISY software. *Comput. Biol. Med.* **40**(4), 402–407. <https://doi.org/10.1016/j.combiomed.2010.02.004> (2010).
20. Hong, H., Ovchinnikov, A., Pogudin, G. & Yap, C. Global identifiability of differential models. *Commun. Pure Appl. Math.* **73**, 1831–1879 (2020).
21. Walter, E. & Pronzato, L. On the identifiability and distinguishability of nonlinear parametric models. *Math. Comput. Simul.* **42**, 125–134 (1996).
22. Stigter, J. D. & Peeters, R. L. M. On a geometric approach to the structural identifiability problem and its application in a water quality case study. In *Proceedings of European Control Conference (ECC), Kos, Greece* (2007).
23. Henson, M. & Seborg, D. *Nonlinear Process Control* (Prentice Hall, 1997).
24. Joubert, D., Stigter, J. D. & Molenaar, J. An efficient procedure to assist in the re-parametrization of structurally unidentifiable models. *Math. Biosci.* **323**, 108328. <https://doi.org/10.1016/j.mbs.2020.108328> (2020).
25. Bard, Y. *Nonlinear Parameter Estimation* (Academic Press Inc, 1974).
26. Vidyasagar, M. *Nonlinear Systems Analysis* (Prentice Hall, 1993).
27. Gábor, A., Villaverde, A. F. & Banga, J. R. Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC Syst. Biol.* **11**, 1–16. <https://doi.org/10.1186/s12918-017-0428-y> (2017).
28. Golub, G. H. & Van Loan, C. F. *Matrix Computations* 4th edn. (The Johns Hopkins University Press, 2013).
29. Quintana-Ortí, G. & Quintana-Ortí, E. S. Parallel codes for computing the numerical rank. *Linear Algebra Appl.* **275–276**, 451–470. [https://doi.org/10.1016/S0024-3795\(97\)10032-5](https://doi.org/10.1016/S0024-3795(97)10032-5) (1998).
30. Meshkat, N., Er-zhen Kuo, C. & DiStefano, J. III. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: A novel web implementation. *PLOS ONE* **9**, 1–14. <https://doi.org/10.1371/journal.pone.0110261> (2014).
31. Ligon, T. *et al.* Genssi 2.0: Multi-experiment structural identifiability analysis of sbml models. *Bioinformatics* **34**, 1421–1423. <https://doi.org/10.1093/bioinformatics/btx735> (2018).
32. Tunali, E. T. & Tarn, T. J. New results for identifiability of nonlinear systems. *IEEE Trans. Autom. Control* **32**, 146–154. <https://doi.org/10.1109/TAC.1987.1104544> (1987).
33. Stigter, J. D., Beck, M. B. & Molenaar, J. Assessing local structural identifiability for environmental models. *Environ. Model. Softw.* **93**, 398–408. <https://doi.org/10.1016/j.envsoft.2017.03.006> (2017).
34. Iggidr, A. Controllability, observability, and stability of mathematical models. In *Encyclopedia of Life Support Systems (EOLSS)* (Filar, J. A. ed.), Vol. Mathematical Models (UNESCO, Eolss Publishers, 2004).

Acknowledgements

We wish to thank both reviewers for their comments. The example suggested by the first reviewer was included in the “Discussion”.

Author contributions

D.J. contributed in the writing of the paper, theoretical analysis, providing case studies and programming/computations. J.D.S. contributed in the discussion and analysis of results. J.M. contributed in the discussion and analysis of results. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021