An efficient procedure to assist in the re-parametrization of structurally unidentifiable models

3 authors:

Dominique Joubert
Wageningen University & Research
24 PUBLICATIONS   **54** CITATIONS

Hans Stigter
Wageningen University & Research
142 PUBLICATIONS   **957** CITATIONS

Jaap Molenaar
Wageningen University & Research
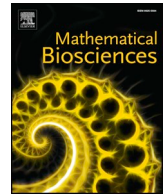**278** PUBLICATIONS   **1,839** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project SMART Detection and Real-time Learning in Water Distribution: An Integrated Data-Model Approach View project

Project Polymer processing rheology View project

Original research article

An efficient procedure to assist in the re-parametrization of structurally unidentifiable models

D. Joubert*, J.D. Stigter, J. Molenaar

Mathematical and Statistical Methods Group, Wageningen University & Research, P.O.Box 16, Wageningen 6700AV, the Netherlands

A B S T R A C T

An efficient method that assists in the re-parametrization of structurally unidentifiable models is introduced. It significantly reduces computational demand by combining numerical and symbolic identifiability calculations. This hybrid approach facilitates the re-parametrization of large unidentifiable ordinary differential equation models, including models where state transformations are required. A model is first assessed numerically, to discover potential structurally unidentifiable parameters. We then use symbolic calculations to confirm the numerical results, after which we describe the algebraic relationships between the unidentifiable parameters. Finally, the unidentifiable parameters are substituted with new parameters and simplification ensures that all the unidentifiable parameters are eliminated from the original model structure. The novelty of this method is its utilisation of numerical results, which notably reduces the number of symbolic calculations required. We illustrate our procedure and the detailed re-parametrization process in 5 examples: (1) an immunological model, (2) a microbial growth model, (3) a lung cancer model, (4) a JAK/STAT model, and (5) a small linear model with a non-scalable re-parametrization.

1. Introduction

System biology models often utilize ordinary differential equations to describe physical phenomena. These models may contain large numbers of parameters and at times, also initial conditions that need to be inferred from experimental data. However, in some cases the statistical inference fails. This may be due to insufficient or low quality data, which is referred to as practical unidentifiability, or due to the inherent structure of the model, referred to as structural unidentifiability [1]. One way of characterizing structural unidentifiability is to say that at least 1 parameter estimate has a confidence interval that spans the interval $(-\infty, \infty)$. Any form of unidentifiability, also referred to as aliasing, calls into question the predictive power of a model and urges its user to interpret all results with caution.

If a model is classified as structurally unidentifiable, a modeller may wish to know "how can this model be made identifiable?" In this paper, we present a method that can answer this question, also for large models, in an efficient way.

The central concept underpinning this method is that *unidentifiable parameters may be divided into different subsets of totally correlated or aliased parameters* [2]. From a linear algebra perspective, this implies that the *columns* of a sensitivity matrix, each pertaining to a specific parameter, are *linearly dependent*. Structural identifiability can only be

obtained if *the correlation between parameters in each of these correlated sets is destroyed*. This can be done in different ways:

1. Measure additional model outputs. Although there may be some practical restrictions to this approach, a good understanding of a model's minimal output sets may result in easily obtainable structural identifiability [3,4]. For example, if a model has 1 set of totally correlated parameters that includes a certain state's unknown initial condition, measuring this initial condition will destroy the correlation and result in identifiability of all parameters.

2. Deduce the values of unidentifiable parameters from other sources. If the value of one of the unknown parameters in a totally correlated set is known, this correlation is destroyed. To destroy all correlations, the value of one parameter from each of the different totally correlated sets must be known. Such a parameter value could be obtained either from literature or from a separate experiment. It should however be noted that even if parameter values are obtained from literature, they may still require re-calibration with experimental data and so this should always be done with caution [5].

3. Re-parametrize the model to remove redundant parameters. In this approach, the total number of model parameters is reduced by the number of totally correlated sets. For example, if a model has P parameters and M ($<P$) sets of totally correlated parameters, the re-

* Corresponding author.
E-mail address: dominique.joubert@wur.nl (D. Joubert).

parametrized, structurally identifiable model will have $P - M$ parameters.

4. Start an experiment from different initial conditions. It might be that the initial values are taken from a thin set of singular points that give rise to structural unidentifiability [6]. In such a case, it may be possible to regain structural identifiability by changing the initial experimental conditions.

In this paper, we introduce a method related to point 3. The novelty of the work presented here is that we can re-parametrize large ordinary differential equation (ODE) models and, because our calculations allow for the inclusion of unknown initial conditions of model states as additional parameters, the obtained results will also reveal state transformations if required.

Although this topic has been covered to some extent in the past, re-parametrization of large models including state transformations has remained elusive. Previous work on this topic includes: (a) symbolic methods based on exhaustive summary [1,7–9] and examples including state transformations [10,11], (b) numerical methods involving the Fisher Information Matrix and profile likelihood [12] and (c) hybrid methods where a Jacobi matrix is symbolically calculated and analyzed numerically [1], and where a sensitivity matrix is computed numerically and analyzed using a singular value decomposition (SVD) [13]. In the latter approach, the unidentifiable parameters are subsequently fixed at nominal values and so the model structure remains unchanged.

Before we discuss our approach, we want to emphasize that *a re-parametrization process is never unique*. It is always up to the modeller to decide which redundant parameters to eliminate from a model. In this respect, caution should be taken as not all re-parametrizations are biologically relevant [14].

This paper is divided as follows: in Section 2 the steps of our re-parametrization method, including the underlying theoretical concepts, are discussed. These steps are also illustrated in an example. Section 3 contains 4 examples that show the re-parametrization process in detail. Concluding remarks are given in Section 4.

2. Methods

2.1. Model definition

We begin with the definition of a typical ordinary differential equation model, as used in systems biology. These models often describe mass balances of certain cellular constituents and can be very detailed, containing numerous model states and vast numbers of unknown parameters which need to be inferred from experimental data. In this paper, we analyze dynamic models that can be written in the standard state-space form:

$$\dot{x}(t) = f_0(x(t), \theta) + \sum_{i=1}^{k} u_i(t) f_i(x(t), \theta), \tag{1}$$

$$x(0) = x_0, \tag{2}$$

$$y(t) = h(x(t), \theta). \tag{3}$$

State variables are contained in vector $x(t)$ ($\dim(x) = n$), system parameters in vector θ ($\dim(\theta) = p$), and measured sensors or outputs in vector $y(t)$ ($\dim(y) = m$). The state vector, $x(t)$, evolves on a manifold V in R^n. Functions f_i, $i = 0, ..., k$, and output function h are assumed to be analytical on V and C^∞ functions and so, their partial derivatives of any order exist and are continuous [15]. A model's input functions are contained in vector . Initial values of the model states may also be unknown and in such cases, the initial condition vector may be parametrized through some additional parameters that then become part of the identification problem. The resulting parameter vector, θ, then has $\dim(\theta) = p + n$ [16].

2.2. The 5 step re-parametrization process

The steps required to re-parametrize a structurally unidentifiable model are discussed here. *Step 1:* numerical identifiability analysis (pinpoint potential unidentifiable parameters), *Step 2:* symbolic identifiability analysis (identify sets of totally correlated parameters), *Step 3:* define new model parameters (compute the algebraic relationship between totally correlated parameters), *Step 4:* use the newly computed parameters to simplify the model, and *Step 5:* re-evaluate the identifiability of the re-parametrized model.

Step 1: Numerical identifiability analysis

Our numerical identifiability method uses sensitivity based calculations of model outputs with respect to model parameters [17,18]. Reid introduced the concept of sensitivity based identifiability analysis for linear models [19]. In his paper, he defines a sensitivity matrix, **S**, with elements depicting the sensitivities of the model output with respect to model parameters, $\partial y / \partial \theta$. For non-linear models, the sensitivities of model outputs to individual model parameters can be calculated from the following equations that can be obtained from (1) to (3):

$$\frac{d}{dt}\left(\frac{\partial x}{\partial \theta}\right) = \frac{\partial f_0}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f_0}{\partial \theta} + \sum_{i=1}^{k}\left(\frac{\partial f_i}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f_i}{\partial \theta}\right)u_i, \tag{4}$$

$$\frac{\partial y}{\partial \theta} = \frac{\partial h}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial h}{\partial \theta}. \tag{5}$$

One obtains $\partial y / \partial \theta$ as a function of time by integrating Eqs. (1) and (4) and substituting the solution into (5). By calculating these sensitivities at discrete time points on an interval $[t_0, ..., t_N]$, one can construct a sensitivity matrix, **S**.

$$S = \begin{pmatrix} \frac{\partial y_1}{\partial \theta_1}(t_0) & \cdots & \frac{\partial y_1}{\partial \theta_{p+n}}(t_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial \theta_1}(t_0) & \cdots & \frac{\partial y_m}{\partial \theta_{p+n}}(t_0) \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial \theta_1}(t_N) & \cdots & \frac{\partial y_1}{\partial \theta_{p+n}}(t_N) \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial \theta_1}(t_N) & \cdots & \frac{\partial y_m}{\partial \theta_{p+n}}(t_N) \end{pmatrix}. \tag{6}$$

If any of the initial values of model states are unknown, their identifiability can easily be assessed by regarding them as additional unknown parameters. In such cases, S may have up to $p + n$ columns, each related to a specific parameter, θ_i, $i = 1, ..., p + n$.

A full ranked sensitivity matrix, S, is a sufficient condition for local structural identifiability [20,21]. The rank deficiency of S can be attributed to 2 factors: (1) a model output may be insensitive to a specific parameter and in this instance, the column of S pertaining to this parameter contains only zeros and the parameter is directly classified as unidentifiable and/or (2) a model output may be sensitive to a particular parameter, but this sensitivity is counteracted by the sensitivity of the model output to other parameters. In this case the columns of S, related to these parameters will be linearly dependent. This implies that these parameters are totally correlated and thus unidentifiable [22].

The rank of the sensitivity matrix, S, is numerically determined using singular value decomposition (SVD):

$$S = U\Sigma V^T. \tag{7}$$

If S has $p + n$ columns, matrix Σ will have $p + n$ singular values on its diagonal and these are arranged in descending order. The rank of S is the number of non-zero singular values and conversely rank-deficiency is indicated by the presence of zero-valued singular values [23]. Due to numerical rounding errors, singular values are rarely exactly zero and

so one uses as practical definition: zero-valued singular values are values that fall beyond a distinct gap in the spectrum of singular values [24].

Once possible unidentifiability suggested by the presence of zero-valued singular values has been established, the unidentifiable parameters responsible for the unidentifiability are recognized from the non-zero entries in the columns of the matrix V, related to the vanishing singular values. These are indicated in a set θ^{unid} and can comprise both system parameters and initial conditions. There may be multiple sets of totally correlated parameters within this collection of unidentifiable parameters. Both the singular values and the unidentifiable parameters can graphically be illustrated in an easy to interpret identifiability signature [18].

As an example, consider the identifiability signature of an unidentifiable immunological model describing mastitis in dairy cows. This small model contains 2 states and 7 unknown parameters, 5 system parameters and 2 initial conditions. Its non-linear model structure reads [25,26]:

$$\dot{x}_1(t) = \theta_1 x_1(t) - \theta_2 x_1(t) x_2(t), \tag{8}$$

$$\dot{x}_2(t) = \theta_3 x_2(t)(1 - \theta_4 x_2(t)) + \theta_5 x_1(t) x_2(t). \tag{9}$$

This model was found to be unidentifiable if only $x_1(t)$ is measured, so $y(t) = \{x_1(t)\}$, with parameters θ_2, θ_4 and the initial state $x_2(0)$ classified as unidentifiable [7,26]. The identifiability signature is given in Fig. 1. The lone singular value beyond the gap suggests unidentifiability, with $\theta^{unid} = \{\theta_2, \theta_4, x_2(0)\}$, deduced from the bottom plot in Fig. 1. *Step 2: Symbolic identifiability calculations*

Symbolic calculations may be computationally demanding and so there is a significant need for applications that can reduce this computing burden. Our method uses the numerical results obtained in step 1 to reduce the number of required symbolic calculations. More concretely, a Jacobi matrix of the coefficients in a Fliess series expansion, which will be discussed in this section, only needs to be computed with respect to the parameters in θ^{unid} resulting in substantial computational savings.

The work of Fliess [27] and Tunali and Tarn [28] made it possible to analyze non-linear systems symbolically. It requires the symbolic calculation of a so-called Jacobi matrix. We use the rank condition for local structural identifiability presented by Tunali and Tarn [28]. The calculations required to compute this Jacobi matrix depend on whether or not the model contains input functions. Here, we describe the computations associated with and without input functions.

The Jacobi matrix of a model with no control input, i.e. with $u_i(t) \equiv 0$ ($i = 1, ..., k$) in Eq. (1), can be computed using Lie derivatives. A Lie derivative is the derivative of the smooth output function, h, in the direction of the drift vector field, f_0, given in (1):

$$\mathcal{L}_{f_0} h = \frac{\partial h}{\partial x} f_0. \tag{10}$$

Successive Lie derivatives are computed as:

$$\mathcal{L}_{f_0}^i h = \frac{\partial \mathcal{L}_{f_0}^{i-1} h}{\partial x} f_0. \tag{11}$$

We use a symbolic algebra package, Kwatny's ProPac add-on for Mathematica, to calculate successive Lie derivatives of the vector h to form a generating series expansion [16]. By parametrizing the unknown initial conditions and therefore regarding them as additional parameters, the Jacobi matrix may also have up to $p + n$ columns. The Jacobi matrix is defined as [29,30]:

$$\frac{\partial G}{\partial \theta}(\theta) = \begin{pmatrix} \frac{\partial h}{\partial \theta_1} & \cdots & \frac{\partial h}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_0} h}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_0} h}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_0}^2 h}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_0}^2 h}{\partial \theta_{p+n}} \\ \vdots & \cdots & \vdots \end{pmatrix}. \tag{12}$$

For models of the form defined in (1), the individual input functions should be incorporated into calculations [31,32]. An output can be expanded in a Fliess series [28] with respect to time and *inputs*. The coefficients of this series are $h(x(0), \theta)$ and:

$$\mathcal{L}_{f_{j_0}} \cdots \mathcal{L}_{f_{j_q}} h(x(t), \theta)|_0, \tag{13}$$

where $f_{j_0}, \dots f_{j_q}$ represent all possible combinations of the input functions $\{f_j, j = 0, ..., k\}$ [17,31]. The notation $|_0$ indicates that this matrix is evaluated in the point $x(0)$. For example, the Jacobi matrix associated with the full model in (1) if $k = 1$, calculated with respect to the unknown parameters in the augmented parameter vector $\theta = \begin{pmatrix} \theta \\ x_0 \end{pmatrix}$ is shown in (14) [17]. Here, $j_0, j_1, \cdots j_q \in \{0, 1\}$.

Fig. 1. *Identifiability signature for the immunological model* - (Top) Singular values - the numerically zero-valued singular value suggests rank deficiency of the sensitivity matrix. This singular value falls beyond a gap larger than 3 decades and indicates that the model is unidentifiable and that there is 1 set of totally correlated parameters. (Bottom) Components of the last column of the right singular matrix corresponding to the singular value beyond the gap - the non-zero elements in this column indicate that parameters θ_2, θ_4 and initial condition $x_2(0)$ may be totally correlated and consequently unidentifiable.

$$\frac{\partial \boldsymbol{G}}{\partial \theta}(\theta) = \begin{pmatrix} \frac{\partial \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \boldsymbol{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_1} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_1} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_0} \mathcal{L}_{f_1} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_0} \mathcal{L}_{f_1} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_1} \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_1} \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_1} \mathcal{L}_{f_0} \mathcal{L}_{f_0} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \vdots & \cdots & \vdots \\ \frac{\partial \mathcal{L}_{f_{j_0}} \cdots \mathcal{L}_{f_{j_q}} \boldsymbol{h}}{\partial \theta_1} & \cdots & \frac{\partial \mathcal{L}_{f_{j_0}} \cdots \mathcal{L}_{f_{j_q}} \boldsymbol{h}}{\partial \theta_{p+n}} \\ \vdots & \cdots & \vdots \end{pmatrix}. \tag{14}$$

For structural identifiability, it is sufficient for $\frac{\partial G}{\partial \theta}(\theta)$ to have rank $p + n$, implying that all initial values and system parameters can uniquely be determined. As was the case with the sensitivity matrix, $\boldsymbol{S}$, a rank deficient Jacobi matrix indicates that: (1) one or more of the columns of this matrix contain only zeros and/or (2) two or more of the columns of the matrix are linearly dependent. It is known from linear algebra that rank deficiency of a matrix is equivalent to it having a nontrivial null-space [33]. The elements in such a nontrivial null-space can be used to characterize the linear dependence between individual columns of the matrix. As an example, consider the immunological model introduced in (8) and (9). The Jacobi matrix for this model is computed with respect to $\theta^{unid} = \{\theta_2, \theta_4, x_2(0)\}$ only since the numerical SVD analysis suggested these parameters as potentially correlated.

This matrix has a nontrivial null-space, which confirms the numerical findings in step 1. This null-space is spanned by a single base vector, $\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right) = \{-\theta_2/x_2(0), -\theta_4/x_2(0), 1\}$, also confirming that these 3 parameters are totally correlated.

The entries of this base vector are used as coefficients in the partial differential equation describing the linear dependence between the columns of the Jacobi matrix. This partial differential equation for some function $\phi(\theta_2, \theta_4, x_2(0))$ reads as

$$-\frac{\theta_2}{x_2(0)}\frac{\partial \phi}{\partial \theta_2} - \frac{\theta_4}{x_2(0)}\frac{\partial \phi}{\partial \theta_4} + \frac{\partial \phi}{\partial x_2(0)} = 0. \tag{15}$$

Step 3: Compute new model parameters

In general, suggestions for new model parameters are obtained by computing solutions of partial differential equations like the one given in (15) [14]. Since no unique solutions to these equations exist, solutions can be expressed in terms of different parameters and ultimately it is up to a modeller's discretion which new model parameters to implement. Special attention should be paid to ensure that these are biologically sound.

We use the standard DSolve function in Mathematica to calculate solutions of partial differential equations. For the immunological model, two possible solutions to (15) are $\phi_1 = \theta_4/\theta_2$ and $\phi_2 = \theta_2 x_2(0)$.

Step 4: Simplify the original model using newly defined parameters

The new parameters from step 3, characterize the algebraic relationships between the totally correlated parameters, and these have to be substituted in the original model structure. As a result the unidentifiable parameters will be eliminated from the model. We show this process in detail since this can be quite a challenging endeavour.

For the immunological model the single base vector of the null-space indicates that the total number of unknown model parameters has to be reduced by 1. More concretely, parameters $\theta^{unid} = \{\theta_2, \theta_4, x_2(0)\}$

have to be replaced by $\theta^{id} = \{\phi_1, \phi_2\}$.

The new parameter $\phi_2 = \theta_2 x_2(0)$ suggests the required state transformation, $\tilde{x}_2(t) = \theta_2 x_2(t)$. Multiplying (9) by θ_2 we obtain:

$$\dot{x}_1(t) = \theta_1 x_1(t) - x_1(t)\tilde{x}_2(t), \tag{16}$$

$$\dot{\tilde{x}}_2(t) = \theta_3 \tilde{x}_2(t)(1 - \theta_4 x_2(t)) + \theta_5 x_1(t)\tilde{x}_2(t), \tag{17}$$

Next, we use the relation $\theta_4 = \phi_1 \theta_2$ obtained from step 3 to arrive at:

$$\dot{x}_1(t) = \theta_1 x_1(t) - x_1(t)\tilde{x}_2(t), \tag{18}$$

$$\dot{\tilde{x}}_2(t) = \theta_3 \tilde{x}_2(t)(1 - \phi_1 \tilde{x}_2(t)) + \theta_5 x_1(t)\tilde{x}_2(t). \tag{19}$$

The final re-parametrized identifiable model has 6 parameters instead of 7: $\theta_1, \theta_3, \theta_5, \phi_1, x_1(0)$ and $\tilde{x}_2(0) = \phi_2 = \theta_2 x_2(0)$.

Step 5: Check the identifiability of the re-parametrized model

In the final step the structural identifiability of the re-parametrized model should be re-evaluated. In the supplementary material we show that the original and re-parametrized immunological models in (8) and (9), and (18) and (19) respectively, have the same outputs and that the re-parametrized model is structurally identifiable.

3. Results

Example 1 (*Microbial growth model (2 states, 6 parameters)*). In this example we consider a benchmark model that also requires a state transformation to obtain an identifiable re-parametrization [11]. This model describes the microbial growth in a batch reactor and has 2 states and 6 unknown parameters [34]:

$$\dot{x}(t) = \frac{\mu x(t)s(t)}{K_s + s(t)} - K_d x(t), \tag{20}$$

$$\dot{s}(t) = -\frac{\mu x(t)s(t)}{Y(K_s + s(t))}. \tag{21}$$

Step 1: This model is unidentifiable if the output $y(t) = \{x(t)\}$ is measured, with parameters K_s, Y, and $s(0)$ not estimable [11]. The calculated singular values are given in Fig. 2 and show a single singular value beyond the gap. This suggests that the null-space of the sensitivity matrix contains only 1 base vector and so all unidentifiable parameters are totally correlated. The parameters in this set, $\theta^{unid} = \{K_s, Y, s(0)\}$, can be deduced from the bottom plot in Fig. 2. **Step 2:** Using the results from step 1, the calculated Jacobi matrix now only has 3 columns. The non-trivial null-space of this matrix is calculated as $\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right) = \{K_s/s(0), -Y/s(0), 1\}$. Since the null-space is 1-dimensional, the number of model parameters may be reduced from 6 to 5. This null-space leads to the following partial differential equation for some function $\phi\{K_s, Y, s(0)\}$:

$$\frac{K_s}{s(0)}\frac{\partial \phi}{\partial K_s} - \frac{Y}{s(0)}\frac{\partial \phi}{\partial Y} + \frac{\partial \phi}{\partial s(0)} = 0. \tag{22}$$

Step 3: Possible solutions to this equation are $\phi_1 = K_s/s(0)$ and $\phi_2 = Ys(0)$. **Step 4:** Solution ϕ_2 in step 3 suggests that we could try the state transformation $\tilde{s}(t) = Ys(t)$. Substituting this transformation in (20) and multiplying (21) by Y, we find:

$$\dot{x}(t) = \frac{\mu x(t)\tilde{s}(t)}{YK_s + \tilde{s}(t)} - K_d x(t), \tag{23}$$

$$\dot{\tilde{s}}(t) = -\frac{\mu x(t)\tilde{s}(t)}{YK_s + \tilde{s}(t)}. \tag{24}$$

Both Eqs. (23) and (24) still contain the unidentifiable parameters Y and K_s. Using the algebraic relationships calculated in step 3, where, $K_s = \phi_1 s(0)$ and $Y = \phi_2/s(0)$, one finds:

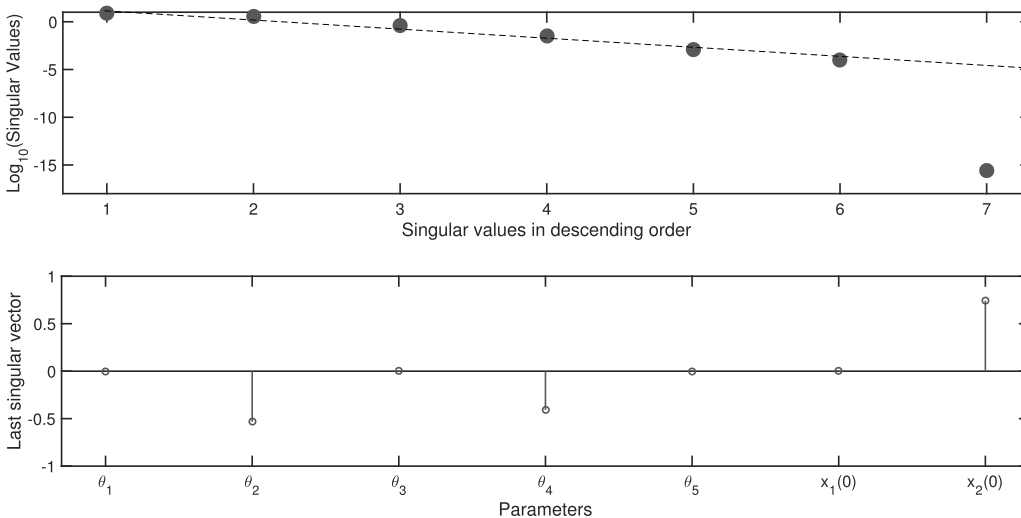$$\dot{x}(t) = \frac{\mu x(t)\tilde{s}(t)}{\phi_2 \phi_1 + \tilde{s}(t)} - K_d x(t), \tag{25}$$
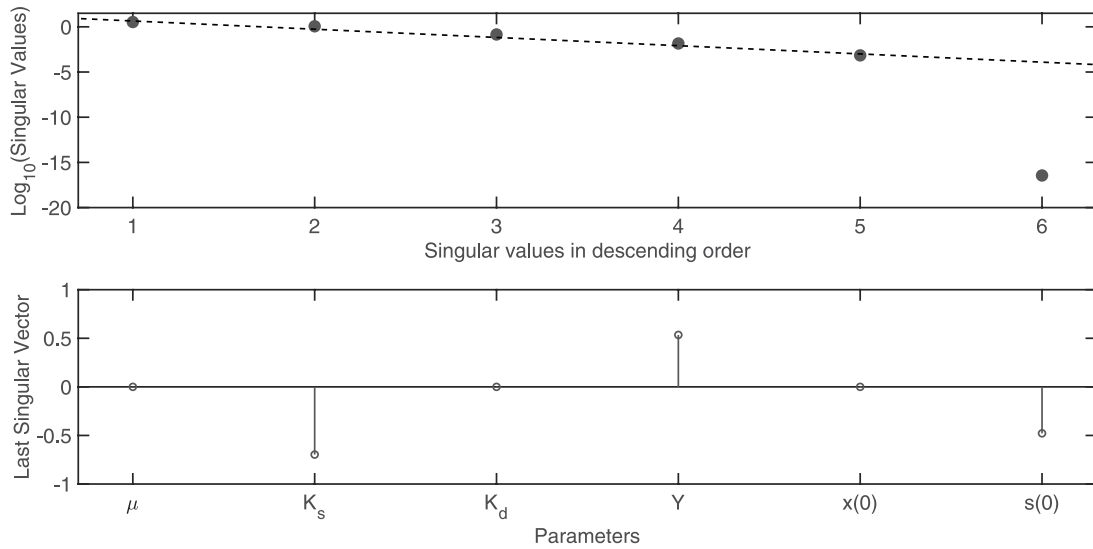
Fig. 2. *Identifiability signature for the microbial growth model* - (Top) Singular values - the numerically zero-valued singular value suggests rank deficiency of the sensitivity matrix. This singular value falls beyond a gap larger than 3 decades and indicates that there is 1 set with totally correlated parameters. (Bottom) Components of the last column of the right singular matrix corresponding to the singular value beyond the gap - non-zero elements in this column indicate that parameters K_s, Y and initial condition $s(0)$ might be totally correlated and therefore unidentifiable.

$$\dot{\tilde{s}}(t) = -\frac{\mu x(t)\tilde{s}(t)}{\phi_2\phi_1 + \tilde{s}(t)}. \tag{26}$$

The re-parametrized model has 5 unknown parameters: μ, K_d, ϕ_1, $x(0)$, and $\tilde{s}(0) = \phi_2 = Ys(0)$. As remarked above, re-parametrization is not a unique procedure. This is demonstrated by comparing our result with that in [11]. The re-parametrization reported in [11] is obtained by following a different route. They use the relations from step 3, $K_s = \phi_1 s(0)$ and $Y = \phi_2/s(0)$, to arrive at:

$$\dot{x}(t) = \frac{\mu x(t)s(t)}{\phi_1 s(0) + s(t)} - K_d x(t) = \frac{\mu x(t)s(t)}{s(0)(\phi_1 + \frac{s(t)}{s(0)})} - K_d x(t), \tag{27}$$

$$\dot{s}(t) = -\frac{\mu x(t)s(t)}{\frac{\phi_2}{s(0)}(\phi_1 s(0) + s(t))} = -\frac{\mu x(t)s(t)}{\phi_2(\phi_1 + \frac{s(t)}{s(0)})}. \tag{28}$$

As final step they apply the state transformation $\tilde{s}(t) = s(t)/s(0)$, which results in:

$$\dot{x}(t) = \frac{\mu x(t)\tilde{s}(t)}{\phi_1 + \tilde{s}(t)} - K_d x(t), \tag{29}$$

$$\dot{\tilde{s}}(t) = -\frac{\mu x(t)\tilde{s}(t)}{\phi_2(\phi_1 + \tilde{s}(t))}. \tag{30}$$

The final model contains 5 parameters instead of 6: μ, K_d, ϕ_1, ϕ_2, and $x(0)$. The initial condition of $\tilde{s}(t)$ is now known since $\tilde{s}(0) = s(0)/s(0) = 1$ [11]. **Step 5:** In the supplementary material we show that the original and new models have exactly the same outputs.

Example 2 (*Lung cancer model (21 states, 75 parameters)*). This lung cancer model describes the biology of the Epidermal Growth Factor Receptor (EGFR) and type 1 Insulin-like Growth Factor (IGF1R) pathways in non-small cell lung cancer (NSCLC) [35,36]. The authors present a detailed in silico ordinary differential equation model, which consists of 21 states and 54 system parameters. The full model is given in the supplementary material. In the original paper, the values of all but 5 of these parameters values are assumed known. Here, we assume that the values of all 54 system parameters and 21 initial conditions are unknown. In addition, we assume that only certain model states can be measured as output sensors. An insightful starting point in the identifiability analysis of a model is to study its directed graph. This

gives a visual queue as to the interaction between model states as well as hinting towards sensors that need to be measured to ensure structural identifiability [37]. However, as models get larger, this visual analysis becomes cumbersome and so one has to proceed with a more theoretical approach. From the directed graph we know that sensors x_1 and x_9, related to EGFRative and IGFRactive respectively, are both root strongly connected components and so need to be measured to ensure the model's identifiability. To illustrate our re-parametrization procedure, we will assume that these key states cannot be measured and so we analyze the model measuring as output, $y(t) = \{x_2(t), x_3(t), x_4(t), x_5(t), x_6(t), x_7(t), x_8(t), x_{10}(t), x_{11}(t), x_{12}(t), x_{13}(t), x_{14}(t), x_{15}(t), x_{16}(t), x_{17}(t), x_{18}(t), x_{19}(t), x_{20}(t), x_{21}(t)\}$.

Step 1: Measuring $y(t)$, we observe 2 singular values beyond the gap in Fig. 3. These suggest that the model is unidentifiable with a null-space spanned by 2 base vectors corresponding with 2 sets of totally correlated parameters. This tells us that by re-parametrizing this model, the total number of unknown parameters will reduce by 2. The unidentifiable parameters, $\Theta^{unid} = \{\theta_6, \theta_{17}, \theta_{20}, \theta_{23}, x_1(0), x_9(0)\}$, can be deduced from the bottom plot in Fig. 3. **Step 2:** The Jacobi matrix now only contains 6 columns, each related to an unidentifiable parameter identified in step 1. If we were to omit step 1, we would have to calculate this matrix with 75 columns instead. As expected, the symbolically calculated null-space of this matrix is spanned by 2 base vectors, confirming our findings in step 1, so $\mathcal{N}(\Theta^{unid})$ is spanned by the vectors $\{0, -\theta_{17}/x_9(0), -\theta_{20}/x_9(0), 0, 0, 1\}$ and $\{-\theta_6/x_1(0), 0, 0, -2\theta_{23}/x_1(0), 1, 0\}$. The linear dependence between the individual columns of the Jacobi matrix is characterized by the following 2 partial differential equations for the functions $\phi_1 = \phi_1(\theta_{17}, \theta_{20}, x_9(0))$ and $\phi_2 = \phi_2(\theta_6, \theta_{23}, x_1(0))$:

$$-\frac{\theta_{17}}{x_9(0)}\frac{\partial \phi_1}{\partial \theta_{17}} - \frac{\theta_{20}}{x_9(0)}\frac{\partial \phi_1}{\partial \theta_{20}} + \frac{\partial \phi_1}{\partial x_9(0)} = 0, \tag{31}$$

$$-\frac{\theta_6}{x_1(0)}\frac{\partial \phi_2}{\partial \theta_6} - 2\frac{\theta_{23}}{x_1(0)}\frac{\partial \phi_2}{\partial \theta_{23}} + \frac{\partial \phi_2}{\partial x_1(0)} = 0. \tag{32}$$

Step 3: Potential solutions of (31) include $\phi_{1,1} = \theta_{20}/\theta_{17}$ and $\phi_{1,2} = \theta_{17}x_9(0)$ and $\phi_{2,1} = \theta_{23}/\theta_6^2$ and $\phi_{2,2} = \theta_6 x_1(0)$ are solutions to (32). In the re-parametrized model parameters $\Theta^{unid} = \{\theta_6, \theta_{17}, \theta_{20}, \theta_{23}, x_1(0), x_9(0)\}$ should be replaced by $\Theta^{id} = \{\phi_{1,1}, \phi_{1,2}, \phi_{2,1}, \phi_{2,2}\}$.

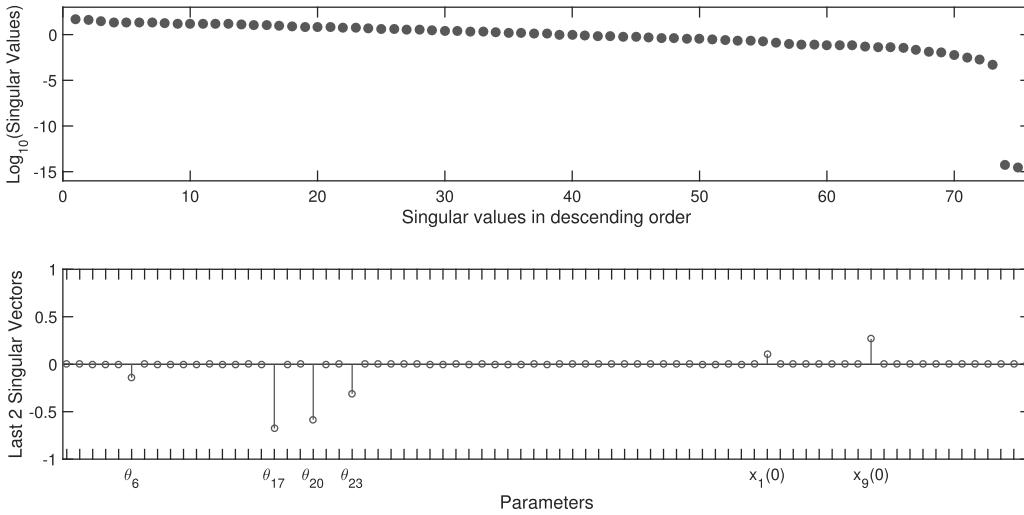**Step 4:** Here, the relevant state equations are given, with the

Fig. 3. *Identifiability signature for the lung cancer model model* - (Top) Singular values - the two vanishing singular values suggest rank deficiency of the sensitivity matrix with 2 sets of totally correlated parameters. (Bottom) Components of the last 2 columns of the right singular matrix corresponding to the 2 singular values beyond the gap - the non-zero elements suggest that parameters θ_6, θ_{17}, θ_{20}, θ_{23} and initial conditions $x_1(0)$ and $x_9(0)$ are structurally unidentifiable.

unidentifiable parameters in bold:

$$\dot{x}_1(t) = -\theta_5 x_1(t), \tag{33}$$

$$\dot{x}_2(t) = -\frac{\boldsymbol{\theta_6} x_1(t) x_2(t)^{\theta_7}}{\theta_8^{\theta_7} - x_2(t)^{\theta_7}} + \frac{x_{21}(t)\theta_{14} x_3(t)^{\theta_{15}}}{\theta_{16}^{\theta_{15}} + x_3(t)^{\theta_{15}}} - \frac{\boldsymbol{\theta_{17}} x_9(t) x_2(t)^{\theta_{18}}}{\theta_{19}^{\theta_{18}} + x_2(t)^{\theta_{18}}}, \tag{34}$$

$$\dot{x}_3(t) = \frac{\boldsymbol{\theta_6} x_1(t) x_2(t)^{\theta_7}}{\theta_8^{\theta_7} - x_2(t)^{\theta_7}} - \frac{x_{21}(t)\theta_{14} x_3(t)^{\theta_{15}}}{\theta_{16}^{\theta_{15}} + x_3(t)^{\theta_{15}}} + \frac{\boldsymbol{\theta_{17}} x_9(t) x_2(t)^{\theta_{18}}}{\theta_{19}^{\theta_{18}} + x_2(t)^{\theta_{18}}}, \tag{35}$$

$$\dot{x}_9(t) = -\theta_1 x_9(t), \tag{36}$$

$$\dot{x}_{10}(t) = -\frac{\boldsymbol{\theta_{20}} x_9(t) x_{10}(t)^{\theta_{21}}}{\theta_{22}^{\theta_{21}} + x_{10}(t)^{\theta_{21}}} - \frac{\boldsymbol{\theta_{23}} x_1(t)^2 x_{10}(t)^{\theta_{24}}}{\theta_{25}^{\theta_{24}} + x_{10}(t)^{\theta_{24}}} - \frac{x_5(t)\theta_{33} x_{10}(t)^{\theta_{34}}}{\theta_{35}^{\theta_{34}} + x_{10}(t)^{\theta_{34}}} + \theta_2 x_{11}(t), \tag{37}$$

$$\dot{x}_{11}(t) = \frac{\boldsymbol{\theta_{20}} x_9(t) x_{10}(t)^{\theta_{21}}}{\theta_{22}^{\theta_{21}} + x_{10}(t)^{\theta_{21}}} + \frac{\boldsymbol{\theta_{23}} x_1(t)^2 x_{10}(t)^{\theta_{24}}}{\theta_{25}^{\theta_{24}} + x_{10}(t)^{\theta_{24}}} + \frac{x_5(t)\theta_{33} x_{10}(t)^{\theta_{34}}}{\theta_{35}^{\theta_{34}} + x_{10}(t)^{\theta_{34}}} - \theta_2 x_{11}(t). \tag{38}$$

The solutions from step 3 suggest the 2 states transformations, $\tilde{x}_1(t) = \theta_6 x_1(t)$ and $\tilde{x}_9(t) = \theta_{17} x_9(t)$. These are already present in Eqs. (34) and (35). Using these state transformations and the relations $\theta_{20} = \phi_{1,1}\theta_{17}$ and $\theta_{23} = \phi_{2,1}\theta_6^2$, we arrive at the re-parametrized model:

$$\dot{\tilde{x}}_1(t) = -\theta_5 \tilde{x}_1(t), \tag{39}$$

$$\dot{x}_2(t) = -\frac{\tilde{x}_1(t) x_2(t)^{\theta_7}}{\theta_8^{\theta_7} - x_2(t)^{\theta_7}} + \frac{x_{21}(t)\theta_{14} x_3(t)^{\theta_{15}}}{\theta_{16}^{\theta_{15}} + x_3(t)^{\theta_{15}}} - \frac{\tilde{x}_9(t) x_2(t)^{\theta_{18}}}{\theta_{19}^{\theta_{18}} + x_2(t)^{\theta_{18}}}, \tag{40}$$

$$\dot{x}_3(t) = \frac{\tilde{x}_1(t) x_2(t)^{\theta_7}}{\theta_8^{\theta_7} - x_2(t)^{\theta_7}} - \frac{x_{21}(t)\theta_{14} x_3(t)^{\theta_{15}}}{\theta_{16}^{\theta_{15}} + x_3(t)^{\theta_{15}}} + \frac{\tilde{x}_9(t) x_2(t)^{\theta_{18}}}{\theta_{19}^{\theta_{18}} + x_2(t)^{\theta_{18}}}, \tag{41}$$

$$\dot{\tilde{x}}_9(t) = -\theta_1 \tilde{x}_9(t), \tag{42}$$

$$\dot{x}_{10}(t) = -\frac{\phi_{1,1}\tilde{x}_9(t) x_{10}(t)^{\theta_{21}}}{\theta_{22}^{\theta_{21}} + x_{10}(t)^{\theta_{21}}} - \frac{\phi_{2,1}\tilde{x}_1(t)^2 x_{10}(t)^{\theta_{24}}}{\theta_{25}^{\theta_{24}} + x_{10}(t)^{\theta_{24}}} - \frac{x_5(t)\theta_{33} x_{10}(t)^{\theta_{34}}}{\theta_{35}^{\theta_{34}} + x_{10}(t)^{\theta_{34}}} + \theta_2 x_{11}(t), \tag{43}$$

$$\dot{x}_{11}(t) = \frac{\phi_{1,1}\tilde{x}_9(t) x_{10}(t)^{\theta_{21}}}{\theta_{22}^{\theta_{21}} + x_{10}(t)^{\theta_{21}}} + \frac{\phi_{2,1}\tilde{x}_1(t)^2 x_{10}(t)^{\theta_{24}}}{\theta_{25}^{\theta_{24}} + x_{10}(t)^{\theta_{24}}} + \frac{x_5(t)\theta_{33} x_{10}(t)^{\theta_{34}}}{\theta_{35}^{\theta_{34}} + x_{10}(t)^{\theta_{34}}} - \theta_2 x_{11}(t). \tag{44}$$

The unknown initial conditions are now defined as $\tilde{x}_1(0) = \phi_{2,2} = \theta_6 x_1(0)$ and $\tilde{x}_9(0) = \phi_{1,2} = \theta_{17} x_9(0)$ respectively. **Step 5:** We refer to the supplementary material for the full re-parametrized model structure.

Example 3 (*JAK/STAT model (14 states, 23 parameters)*). This example is unique in the sense that the required state transformations are not related to unidentifiable initial conditions and are therefore not apparent from obtained solutions in step 3. Instead, they are required since the measured output vector contains additional unidentifiable parameters. A re-parametrization of this type has to our knowledge not previously been illustrated. The constitutive activation of the JAK (Janus kinase)/STAT signalling pathway forms part of both the primary mediastinal B-cell lymphoma (PMBL) and the classical Hodgkin lymphoma (cHL) [38]. Raue et al. investigated the identifiability of this benchmark model using three different approaches and concluded that the model is unidentifiable [39]. We treat the unknown initial value of state x_2 as an additional parameter and so in total, the values of 23 unknown parameters need to be inferred [39,40]:

$$\dot{x}_1(t) = -\theta_1 u_1 c_1 x_1(t) - \theta_5 x_1(t) + \theta_6 x_2(t), \tag{45}$$

$$\dot{x}_2(t) = \theta_5 x_1(t) - \theta_6 x_2(t), \tag{46}$$

$$\dot{x}_3(t) = \theta_1 u_1 c_1 x_1(t) - \theta_2 x_3(t) x_7(t), \tag{47}$$

$$\dot{x}_4(t) = \theta_2 x_3(t) x_7(t) - \theta_3 x_4(t), \tag{48}$$

$$\dot{x}_5(t) = \theta_3 x_4(t) - \theta_4 x_5(t), \tag{49}$$

$$\dot{x}_6(t) = -\frac{\theta_7 x_3(t) x_6(t)}{(1 + \theta_{13} x_{13}(t))} - \frac{\theta_7 x_4(t) x_6(t)}{(1 + \theta_{13} x_{13}(t))} + \theta_8 c_2 x_7(t), \tag{50}$$

$$\dot{x}_7(t) = \frac{\theta_7 x_3(t) x_6(t)}{(1 + \theta_{13} x_{13}(t))} + \frac{\theta_7 x_4(t) x_6(t)}{(1 + \theta_{13} x_{13}(t))} - \theta_8 c_2 x_7(t), \tag{51}$$

$$\dot{x}_8(t) = -\theta_9 x_8(t) x_7(t) + c_2 \theta_{10} x_9(t), \tag{52}$$

$$\dot{x}_9(t) = \theta_9 x_8(t) x_7(t) - c_2 \theta_{10} x_9(t), \tag{53}$$

$$\dot{x}_{10}(t) = \theta_{11} x_9(t), \tag{54}$$

$$\dot{x}_{11}(t) = -\theta_{12} c_1 u_1 x_{11}(t), \tag{55}$$

$$\dot{x}_{12}(t) = \theta_{12} c_1 u_1 x_{11}(t), \tag{56}$$

$$\dot{x}_{13}(t) = \frac{\theta_{14} x_{10}(t)}{(\theta_{15} + x_{10}(t))} - \theta_{16} x_{13}(t), \tag{57}$$

$$\dot{x}_{14}(t) = \theta_{17} x_9(t). \tag{58}$$

The model output contains 5 additional parameters:

$$y_1(t) = x_1(t) + x_3(t) + x_4(t), \tag{59}$$

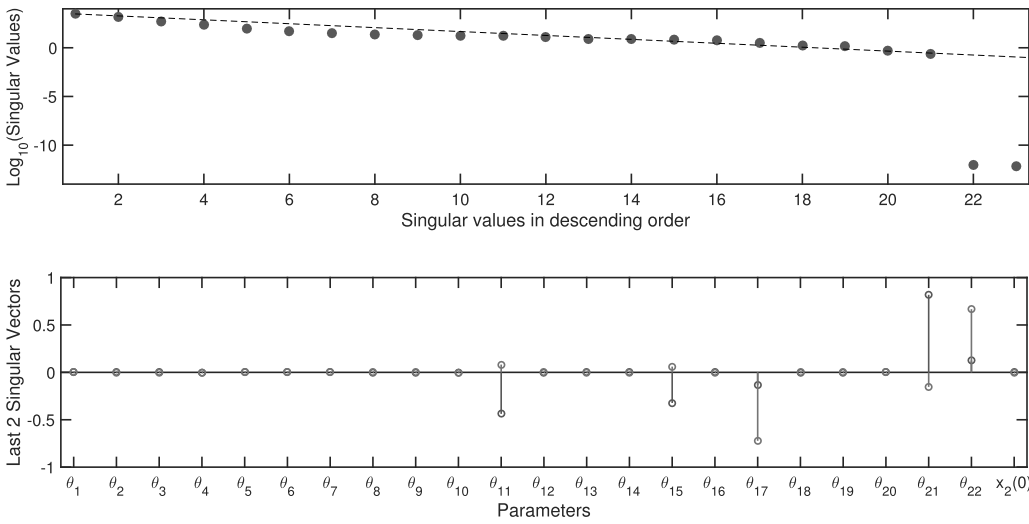$$y_2(t) = \theta_{18}(x_3(t) + x_4(t) + x_5(t) + x_{12}(t)), \tag{60}$$

Fig. 4. *Identifiability signature for the JAK/STAT model model* - (Top) Singular values - the 2 singular values beyond the gap suggest that the model is unidentifiable, with the numerically zero-valued singular values alluding to rank deficiency of the sensitivity matrix, **S**. These values also suggest that there are 2 sets of totally correlated parameters. (Bottom) Components of the 2 last columns of the right singular matrix corresponding to the 2 singular values beyond the gap - these columns are related to the 2 singular values beyond the gap. The non-zero elements indicate that parameters θ_{11}, θ_{15}, θ_{17}, θ_{21} and θ_{22} might be unidentifiable.

$$y_3(t) = \theta_{19}(x_4(t) + x_5(t)), \tag{61}$$

$$y_4(t) = \theta_{20}x_7(t), \tag{62}$$

$$y_5(t) = \theta_{21}x_{10}(t), \tag{63}$$

$$y_6(t) = \theta_{22}x_{14}(t), \tag{64}$$

$$y_7(t) = x_{13}(t), \tag{65}$$

$$y_8(t) = x_9(t). \tag{66}$$

The initial values of the individual model states are $\boldsymbol{x}(0) = \{1.3, \theta_{23}, 0, 0, 0, 2.8, 0, 165, 0, 0, 0.34, 0, 0, 0\}$ [39]. The constants c_1, c_2 are assumed known. **Step 1:** The model's unidentifiability is evident from the large gap between the singular values in Fig. 4. The 2 singular values beyond the gap suggest that the null-space contains 2 base vectors and that there are 2 sets of totally correlated parameters. The union of the elements in these 2 sets, $\boldsymbol{\theta}^{unid} = \{\theta_{11}, \theta_{15}, \theta_{17}, \theta_{21}, \theta_{22}\}$, follows from the non-zero elements in the bottom plot of Fig. 4. This is in agreement with the findings of previous authors [39]. **Step 2:** The calculated Jacobi matrix now only has 5 columns, instead of 23, each corresponding to an unidentifiable parameter in $\boldsymbol{\theta}^{unid}$. Its null-space, $\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right)$, spanned by the vectors $\{0, 0, -\theta_{17}/\theta_{22}, 0, 1\}$ and $\{-\theta_{11}/\theta_{21}, -\theta_{15}/\theta_{21}, 0, 1, 0\}$, substantiates the numerical findings in Fig. 4. The 2 base-vectors spanning this null-space indicate the existence of 2 sets of totally correlated parameters. Entries of these respective base vectors are used in 2 separate partial differential equations, described for the functions $\phi_1 = \phi_1(\theta_{17}, \theta_{22})$ and $\phi_2 = \phi_2(\theta_{11}, \theta_{15}, \theta_{21})$:

$$-\frac{\theta_{17}}{\theta_{22}}\frac{\partial\phi_1}{\partial\theta_{17}} + \frac{\partial\phi_1}{\partial\theta_{22}} = 0, \tag{67}$$

$$-\frac{\theta_{11}}{\theta_{21}}\frac{\partial\phi_2}{\partial\theta_{11}} - \frac{\theta_{15}}{\theta_{21}}\frac{\partial\phi_2}{\partial\theta_{15}} + \frac{\partial\phi_2}{\partial\theta_{22}} = 0. \tag{68}$$

Step 3: A possible solution to Eq. (67) is $\phi_{1,1} = \theta_{17}\theta_{22}$, while $\phi_{2,1} = \frac{\theta_{15}}{\theta_{11}}$ and $\phi_{2,2} = \theta_{11}\theta_{21}$ are solutions to (68). Accordingly, parameters $\boldsymbol{\theta}^{unid} = \{\theta_{11}, \theta_{15}, \theta_{17}, \theta_{21}, \theta_{22}\}$ need to be replaced by $\boldsymbol{\theta}^{id} = \{\phi_{1,1}, \phi_{2,1}, \phi_{2,2}\}$. **Step 4:** Here we only show the relevant model equations, with substituted parameters indicated in bold. Substituting $\theta_{15} = \phi_{2,1}\theta_{11}$ into (57) yields:

$$\dot{x}_{10}(t) = \theta_{11}x_9(t), \tag{69}$$

$$\dot{x}_{13}(t) = \frac{\theta_{14}x_{10}(t)}{\boldsymbol{\phi_{2,1}\theta_{11}} + x_{10}(t)} - \theta_{16}x_{13}(t), \tag{70}$$

$$\dot{x}_{14}(t) = \theta_{17}x_9(t). \tag{71}$$

Substituting $\theta_{21} = \frac{\phi_{2,2}}{\theta_{11}}$ and $\theta_{22} = \frac{\phi_{1,1}}{\theta_{17}}$ into output Eqs. (63) and (64) yields:

$$y_5(t) = \frac{\boldsymbol{\phi_{2,2}}}{\boldsymbol{\theta_{11}}}x_{10}(t), \tag{72}$$

$$y_6(t) = \frac{\boldsymbol{\phi_{1,1}}}{\boldsymbol{\theta_{17}}}x_{14}(t). \tag{73}$$

After these substitutions, parameters θ_{11} and θ_{17} are still present in the model. An inspection and simplification of the model equations reveals that 2 state transformations are required to facilitate the elimination of these parameters. The form of these transformations are suggested by (72) and (73) as $\tilde{x}_{10}(t) = x_{10}(t)/\theta_{11}$ and $\tilde{x}_{14}(t) = x_{14}(t)/\theta_{17}$ respectively. After some additional algebra, we arrive at:

$$\dot{\tilde{x}}_{10}(t) = x_9(t), \tag{74}$$

$$\dot{\tilde{x}}_{13}(t) = \frac{\theta_{14}\tilde{x}_{10}(t)}{\phi_{2,1} + \tilde{x}_{10}(t)} - \theta_{16}x_{13}(t), \tag{75}$$

$$\dot{\tilde{x}}_{14} = x_9(t), \tag{76}$$

$$y_5(t) = \phi_{2,2}\tilde{x}_{10}(t), \tag{77}$$

$$y_6(t) = \phi_{1,1}\tilde{x}_{14}(t). \tag{78}$$

Step 5: We refer to the supplementary material for the full re-parametrized model structure.

Example 4 (*Linear model (3 states, 10 parameters)*). This linear model comprises 3 states and 7 system parameters and so there are 10 unknown parameters in total, $\theta = \{a_{11}, a_{12}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}, x_1(0), x_2(0), x_3(0)\}$ [41]:

$$\dot{x}_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + u, \tag{79}$$

$$\dot{x}_2(t) = a_{22}x_2(t) + a_{23}x_3(t), \tag{80}$$

$$\dot{x}_3(t) = a_{31}x_1(t) + a_{32}x_2(t) + a_{33}x_3(t). \tag{81}$$

In this example we obtain an non-scalable state transformation and associated model re-parametrization by repeating steps 2-4, thereby simplifying the model at each iteration. More concretely, one zero-singular-value of the output sensitivity matrix is destroyed at each iteration, thereby discarding unidentifiable model parameters one-by-one. As it turns out, the model re-parametrization then becomes manageable in a reasonable number of steps. It is interesting to observe in this process how the state transformations facilitate the destruction of a particular parameter in the model structure. This is shown in more detail below. **Step 1:** The model is structurally
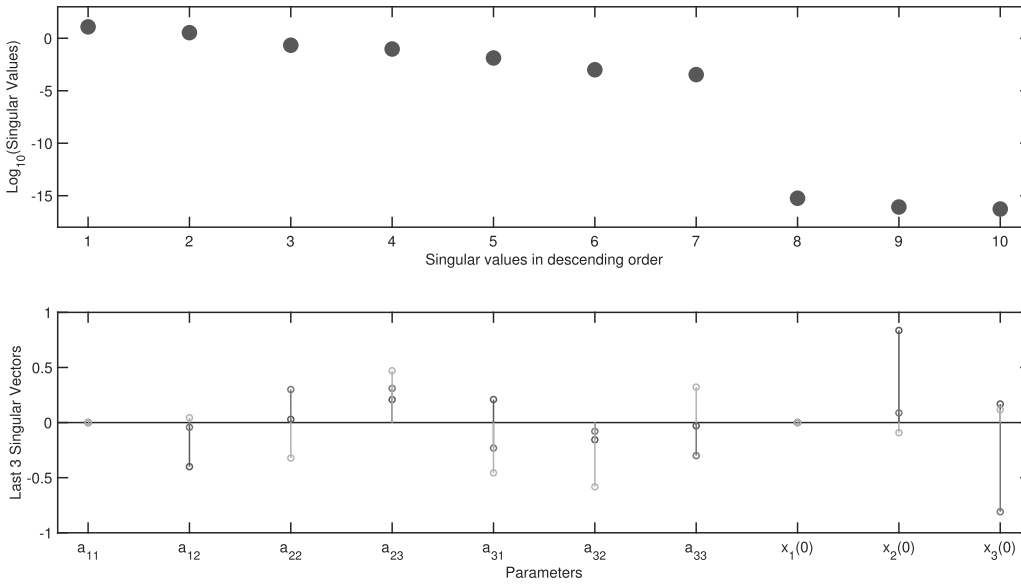
Fig. 5. *Identifiability signature for the linear* model *- (Top) Singular values - three numerically zero-valued singular value suggest rank deficiency of the sensitivity matrix. (Bottom) Components of the last 3 columns of the right singular matrix corresponding to the 3 singular values beyond the gap - non-zero elements in this plot indicate that parameters* a_{12}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33} *and initial conditions* $x_2(0)$ *and* $x_3(0)$ *might be unidentifiable.*

unidentifiable measuring $y(t) = \{x_1(t)\}$ as output, with 3 singular values beyond the gap as can be seen in the top plot in Fig. 5. This suggests that the total number of unknown parameters can be reduced from 10 to 7. The unidentifiable parameters $\theta^{unid} = \{a_{12}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}, x_2(0), x_3(0)\}$, are deduced from the bottom plot in Fig. 5. **Step 2**: This model is known to have a non-scalable re-parametrization [41] and our task is now to find this re-parametrization in more detail. To start the reduction procedure, a null-space was computed for the original model, first taking successive Lie-derivatives of the output $y(t) = \{x_1(t)\}$, and then a Jacobi matrix was derived for the unidentifiable parameter set

$$\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right)=$$

$$\begin{pmatrix} 0 & 0 & -\frac{a_{23}}{x_3(0)} & \frac{a_{31}}{x_3(0)} & & \frac{a_{32}}{x_3(0)} & & 0 & 0 & 1 \\ -\frac{a_{12}}{x_2(0)} & 0 & \frac{a_{23}}{x_2(0)} & 0 & & -\frac{a_{32}}{x_2(0)} & & 0 & 1 & 0 \\ 0 & -1 & \frac{x_2(0)}{x_3(0)} & -\frac{a_{31}x_2(0)}{a_{23}x_3(0)} & \frac{a_{22}x_3(0)-a_{33}x_3(0)-a_{32}x_2(0)}{a_{23}x_3(0)} & & 1 & 0 & 0 \end{pmatrix}.$$

A first observation is that we can separate the unidentifiable set, θ^{unid}, into three groups namely, $\theta_1^u = \{a_{23}, a_{31}, a_{32}, x_3(0)\}$, $\theta_2^u = \{a_{12}, a_{23}, a_{32}, x_2(0)\}$, and $\theta_3^u = \{a_{22}, a_{23}, a_{31}, a_{32}, a_{33}\}$. This is apparent from the rows of the Jacobi matrix. Second, a partial differential equation can be associated with each of the above null-space basis vectors (or rows in the Jacobi matrix). The partial differential equation associated with the first row is:

$$-\frac{a_{23}}{x_3(0)}\frac{\partial\phi}{\partial a_{23}} + \frac{a_{31}}{x_3(0)}\frac{\partial\phi}{\partial a_{31}} + \frac{a_{32}}{x_3(0)}\frac{\partial\phi}{\partial a_{32}} + \frac{\partial\phi}{\partial x_3(0)} = 0. \tag{82}$$

Step 3: Three candidate solutions for (82) are:

$$\phi_1 = a_{23}x_3(0), \tag{83}$$

$$\phi_2 = a_{23}a_{31}, \tag{84}$$

$$\phi_3 = a_{23}a_{32}. \tag{85}$$

Step 4: The combination of parameters in (83) suggests that a good candidate for a first state transformation is $z_3(t) = a_{23}x_3(t)$. We leave x_1 and x_2 unchanged, i.e. $z_1(t) = x_1(t)$ and $z_2(t) = x_2(t)$. Substituting this and ϕ_2 and ϕ_3 into Eqs. (79)–(81), we get:

$$\dot{z}_1(t) = a_{11}z_1(t) + a_{12}z_2(t) + u, \tag{86}$$

$$\dot{z}_2(t) = a_{22}z_2(t) + z_3(t), \tag{87}$$

$$\dot{z}_3(t) = \phi_2 z_1(t) + \phi_3 z_2(t) + a_{33}z_3(t). \tag{88}$$

The new output equation clearly becomes $y(t) = \{z_1(t)\}$. **Step 2**: Again, we generated Lie-derivatives for this new output equation, computed the Jacobi matrix with respect to the new unidentifiable parameter set $\theta^{unid} = \{a_{12}, a_{22}, \phi_2, \phi_3, a_{33}, z_2(0), z_3(0)\}$ and found

$$\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right) = \begin{pmatrix} 0 & -\frac{1}{z_2(0)} & 0 & \frac{a_{22}-a_{33}}{z_2(0)} & \frac{1}{z_2(0)} & 0 & 1 \\ * & \cdots & \cdots & \cdots & \cdots & \cdots & * \end{pmatrix}. \tag{89}$$

Step 3: Solving the partial differential equation associated with the first null-space vector, we get the following 3 re-parametrization candidates:

$$\phi_4 = a_{22} + a_{33}, \tag{90}$$

$$\phi_5 = \phi_3 - a_{22}a_{33}, \tag{91}$$

$$\phi_6 = z_3(0) + a_{22}z_2(0). \tag{92}$$

Step 4: This suggests that a good candidate for our second state transformation is $v_3(t) = z_3(t) + a_{22}z_2(t)$, while $v_1(t) = z_1(t)$ and $v_2(t) = z_2(t)$. When substituting this in the Eqs. (86)–(88), and recognizing the new parameter combinations ϕ_4 and ϕ_5, we get:

$$\dot{v}_1(t) = a_{11}v_1(t) + a_{12}v_2(t) + u, \tag{93}$$

$$\dot{v}_2(t) = v_3(t), \tag{94}$$

$$\dot{v}_3(t) = \phi_2 v_1(t) + \phi_5 v_2(t) + \phi_4 v_3(t). \tag{95}$$

The new output for this system becomes $y(t) = \{v_1(t)\}$. **Step 2**: We continued by generating Lie-derivatives for this new output equation, this time computing the Jacobi matrix with respect to the new unidentifiable parameter set $\theta^{unid} = \{a_{12}, \phi_2, \phi_2, v_2(0), v_3(0)\}$, and found the null-space:

$$\mathcal{N}\left(\frac{dG}{d\theta^{unid}}(\theta)\right) = \left(-\frac{a_{12}}{v_3(0)} \quad \frac{\phi_2}{v_3(0)} \quad 0 \quad \frac{v_2(0)}{v_3(0)}\right). \tag{96}$$

Step 3: Solving the associated partial differential equation we find as a final re-parametrization set:

$$\phi_7 = \phi_2 a_{12}, \tag{97}$$

$$\phi_8 = a_{12}v_2(0), \tag{98}$$

$$\phi_9 = a_{12}v_3(0). \tag{99}$$

Step 4: The last two parameters suggest as a final state transformation $w_2(t) = a_{12}v_2(t)$ and $w_3(t) = a_{12}v_3(t)$. The resulting final set of ordinary

differential equations is:

$$\dot{w}_1(t) = a_{11} w_1(t) + w_2(t) + u, \tag{100}$$

$$\dot{w}_2(t) = w_3(t), \tag{101}$$

$$\dot{w}_3(t) = \phi_7 w_1(t) + \phi_5 w_2(t) + \phi_4 w_3(t). \tag{102}$$

Clearly, the original model has now been reduced to a system with only 4 (instead of 7) system parameters. The transformed model in (100)–(102) has 3 new identifiable parameters, ϕ_4, ϕ_5, ϕ_7. The only original parameter left is a_{11}. The final state transformations ($w_1(t)$, $w_2(t)$, $w_3(t)$) expressed in the original state variables can now be summarized as:

$$w_1(t) = x_1(t), \tag{103}$$

$$w_2(t) = a_{12} x_2(t), \tag{104}$$

$$w_3(t) = a_{12} a_{22} x_2(t) + a_{12} a_{23} x_3(t). \tag{105}$$

Important to observe is that, clearly, the third parameter transformation $w_3(t)$ is a non-scalable transformation! The method clearly has no limitations w.r.t. a particular format of state transformations such as, for example, scaled state variables. **Step 5:** We refer to the supplementary material for the verification of the re-parametrized model structure.

4. Conclusions

In this paper we have introduced an efficient method to assist in the re-parametrization of structurally unidentifiable models. Our hybrid method starts with a numerical identifiability analysis. This analysis is tasked with spotting possible unidentifiable parameters and detecting the number of totally correlated sets. These numerical results are used in subsequent symbolic calculations. Traditionally, these calculations are computationally demanding, sometimes requiring days to obtain relevant results. However, the numerical results allow us to deal with a highly reduced Jacobi matrix. This reduced matrix now only contains columns related to unidentifiable parameters and this enables us to analyze large systems biology models within minutes. Symbolic calculations are used for the verification of numerical results and simultaneously help us to formally characterize the linear dependence between the columns of the Jacobi matrix. After describing these linear dependencies in the form of partial differential equations, solutions to these equations yield suggestions for new parameters that should replace the unidentifiable parameters in the new identifiable model structure. These new parameters describe the algebraic relationships between the unidentifiable parameters. As was shown, the re-parametrization process may require the redefinition of certain model states by means of state transformations. If a state transformation is required, this becomes immediately apparent from the obtained solution to the partial differential equation, due to the fact that our method allows for initial conditions to be incorporated into a model's analysis by regarding them as additional parameters. Alternatively, a required transformation may also become apparent after inspecting the relevant model equations as was the case with the JAK/STAT model.

We showed that not only is our approach capable of reducing models that were already treated in the literature, but it also allows for the re-parametrization of models that are hard to analyze, as was the case with Example 4 in Section 3. The method presented here has the advantage that it can be used to re-parametrize large models in short computing times and is not restricted to rational models or the use of scaled parameters.

Declaration of Competing Interest

The authors declare no conflict of interest.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.mbs.2020.108328

References

[1] D.J. Cole, B.J.T. Morgan, D.M. Titterington, Determining the parametric structure of models, Math. Biosci. 228 (1) (2010) 16–30, https://doi.org/10.1016/j.mbs.2010.08.004.

[2] P. Li, Q.D. Vu, Identification of parameter correlations for parameter estimation in dynamic biological models, BMC Syst. Biol. 7 (91) (2013) 1–12, https://doi.org/10.1186/1752-0509-7-91.

[3] M. Anguelova, J. Karlsson, M. Jirstrand, Minimal output sets for identifiability, Math. Biosci. 239 (1) (2012) 139–153, https://doi.org/10.1016/j.mbs.2012.04.005.

[4] D. Joubert, J. Stigter, J. Molenaar, Determining minimal output sets that ensure structural identifiability, PLoS One 13 (11) (2018) e0207334, https://doi.org/10.1371/journal.pone.0207334.

[5] A.F. Villaverde, J.R. Banga, Dynamical compensation and structural identifiability of biological models: analysis, implications, and reconciliation, PLOS Comput. Biol. 11 (13) (2017) 1–18, https://doi.org/10.1371/journal.pcbi.1005878.

[6] M.P. Saccomani, S. Audoly, L. D'Angió, Parameter identifiability of nonlinear systems: the role of initial conditions, Automatica 39 (4) (2003) 619–632, https://doi.org/10.1016/S0005-1098(02)00302-3.

[7] N. Meshkat, M. Eisenberg, J.J. DiStefano, An algorithm for finding globally identifiable parameter combinations of nonlinear ode models using Grobner Bases, Math. Biosci. 222 (2) (2009) 61–72, https://doi.org/10.1016/j.mbs.2009.08.010.

[8] N. Meshkat, C. Anderson, J.J. DiStefano, Finding identifiable parameter combinations in nonlinear ode models and the rational reparameterization of their input-output equations, Math. Biosci. 233 (1) (2011) 19–31, https://doi.org/10.1016/j.mbs.2011.06.001.

[9] N. Meshkat, S. Sullivant, Identifiable reparametrizations of linear compartment models, J. Symb. Comput. 63 (2014) 46–67, https://doi.org/10.1016/j.jsc.2013.11.002.

[10] M.J. Chappell, R.N. Gunn, A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach, Math. Biosci. 148 (1) (1998) 21–41, https://doi.org/10.1016/S0025-5564(97)10004-9.

[11] N.D. Evans, M.J. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, Math. Biosci. 168 (2) (2000) 137–159, https://doi.org/10.1016/S0025-5564(00)00047-X.

[12] M.C. Eisenberg, M.A. Hayashi, Determining identifiable parameter combinations using subset profiling, Math. Biosci. 256 (2014) 116–126, https://doi.org/10.1016/j.mbs.2014.08.008.

[13] Y. Chu, J. Hahn, Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters, Ind. Eng. Chem. Res. 48 (13) (2009) 6000–6009, https://doi.org/10.1021/ie800432s.

[14] R. Choquet, D. Cole, A hybrid symbolic-numerical method for determining model structure, Math. Biosci. 236 (2) (2012) 117–125, https://doi.org/10.1016/j.mbs.2012.02.002.

[15] M. Henson, D. Seborg, Nonlinear Process Control, New Jersey: Prentice Hall, 1997.

[16] J.D. Stigter, R.L.M. Peeters, On a geometric approach to the structural identifiability problem and its application in a water quality case study, Proceedings of the European Control Conference (ECC), (2007), pp. 3450–3456.

[17] J.D. Stigter, J. Molenaar, A fast algorithm to assess local structural identifiability, Automatica 58 (2015) 118–124, https://doi.org/10.1016/j.automatica.2015.05.004.

[18] J.D. Stigter, D. Joubert, J. Molenaar, Observability of complex systems: finding the gap, Sci. Rep. 7 (16566) (2017) 1–9, https://doi.org/10.1038/s41598-017-16682-x.

[19] J. Reid, Structural identifiability in linear time invariant systems, IEEE Trans. Autom. Control 22 (2) (1977) 242–246, https://doi.org/10.1038/s41598-017-16682-x.

[20] Y. Bard, Nonlinear Parameter Estimation, Academic Press Inc, 1974.

[21] M. Vidyasagar, Nonlinear Systems Analysis, 1, Prentice Hall, 1993.

[22] A. Gábor, A.F. Villaverde, J.R. Banga, Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems, BMC Syst. Biol. 11 (54) (2017) 1–16, https://doi.org/10.1186/s12918-017-0428-y.

[23] G.H. Golub, C.F. Van Loan, Matrix Computations, 4, The Johns Hopkins University Press, 2013.

[24] G. Quintana-Ortí, E.S. Quintana-Ortí, Parallel codes for computing the numerical rank, Linear Algebra Appl. 275–276 (1998) 451–470, https://doi.org/10.1016/S0024-3795(97)10032-5.

[25] D. Döpfer, Recurrent Clinical Escherichia Coli Mastitis in Dairy Cows, Utrecht University, 2000 Ph.D. thesis.

[26] G. Margaria, E. Riccomagno, M.J. Chappell, H.P. Wynn, Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences, Math. Biosci. 174 (2001) 1–26, https://doi.org/10.1016/S0025-5564(01)00079-7.

[27] M. Fliess, Fonctionnelles causales non lineaires et indeterminees non commutatives, Bull. Soc. Math. France 109 (1981) 3–40.

[28] E.T. Tunali, T.J. Tarn, New Results for Identifiability of Nonlinear Systems, 32 (2) (1987) 146–154, https://doi.org/10.1109/TAC.1987.1104544.

[29] E. August, A. Papachristodoulou, A new computational tool for establishing model parameter identifiability, J. Comput. Biol. 6 (16) (2009) 875–885, https://doi.org/10.1089/cmb.2008.0211.

[30] J. Karlsson, M. Anguelova, M. Jirstrand, An efficient method for structural identifiability analysis of large dynamic systems*, IFAC Proc. Vol. 45 (16) (2012) 941–946, https://doi.org/10.3182/20120711-3-BE-2027.00381. 16th IFAC Symposium on System Identification

[31] E. Walter, L. Pronzato, On the identifiability and distinguishability of nonlinear parametric models, Math. Comput. Simul. 42 (1996) 125–134.

[32] A. Iggidr, Controllability, Observability, and Stability of Mathematical Models, in: J.A. Filar (Ed.), Encyclopedia of Life Support Systems (EOLSS), Mathematical Models UNESCO, Eolss Publishers, 2004.

[33] J.D. Stigter, M.B. Beck, J. Molenaar, Assessing local structural identifiability for environmental models, Environ. Model. Softw. 93 (2017) 398–408, https://doi.org/10.1016/j.envsoft.2017.03.006.

[34] A. Holmberg, J. Ranta, Procedures for parameter and state estimation of microbial growth process models, Automatica 18 (2) (1982) 181–193, https://doi.org/10.1016/0005-1098(82)90107-8.

[35] F. Bianconi, E. Baldelli, V. Ludovini, L. Crinó, A. Flacco, P. Valigi, Computational model of egfr and igf1r pathways in lung cancer: a systems biology approach for translational oncology, Biotechnol. Adv. Jan-Feb (1) (2012) 142–153, https://doi.org/10.1016/j.biotechadv.2011.05.010.

[36] F. Bianconi, E. Baldelli, V. Ludovini, L. Crinó, A. Flacco, P. Valigi, Egfr and igf1r Pathway in Lung Cancer (https://www.ebi.ac.uk/biomodels/BIOMD0000000427#History). (2018).

[37] Y. Liu, J. Slotine, B. A.L., Observability of complex systems, Proc. Natl. Acad. Sci. 10 (7) (2013) 2460–2465, https://doi.org/10.1073/pnas.1215508110.

[38] V. Raia, M. Schilling, M. Böhm, B. Hahn, A. Kowarsch, A. Raue, C. Sticht, S. Bohl, M. Saile, P. Möller, N. Gretz, J. Timmer, F. Theis, W.-D. Lehmann, P. Lichter, U. Klingmüller, Dynamic mathematical modeling of IL13-induced signaling in hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets, Cancer Res. 71 (3) (2011) 693–704, https://doi.org/10.1158/0008-5472.CAN-10-2987.

[39] A. Raue, J. Karlsson, M.P. Saccomani, M. Jirstrand, J. Timmer, Comparison of approaches for parameter identifiability analysis of biological systems, Bioinformatics 30 (10) (2014) 1440–1448, https://doi.org/10.1093/bioinformatics/btu006.

[40] D.V. Raman, On the Identifiability of Highly Parameterised Models of Physical Processes, University of Oxford, 2016 Ph.D. thesis.

[41] N. Meshkat, Identifiability of linear compartmental models, Proceedings of the Parameter Estimation Workshop, Raleigh, NC, (2014).