

SCIENTIFIC REPORTS



OPEN

Observability of Complex Systems: Finding the Gap

J. D. Stigter, D. Joubert & J. Molenaar 

For a reconstruction of state and parameter values in a dynamic system model, first the question whether these values can be *uniquely* determined from the data must be answered. This structural model property is known as observability or, in case of parameter calibration only, *identifiability*. Testing a given model for observability is a well studied problem in the systems and control sciences. However, it is increasingly difficult, if not impossible, to address this property for large size models that, nowadays, are frequently used. We demonstrate the application of a recently developed algorithm that overcomes this problem and is remarkably efficient. As an illustration we show how an observability analysis for a Chinese Hamster Ovary Cell model (34 states, 117 parameters), a JAKSTAT signalling model (31 states, 51 parameters), and a MAP Kinase model (100 states, 88 parameters) can be established in a very short time.

Reconstructing the values of certain time varying variables $x(t)$ or time-independent parameters θ for a given dynamic system model from measurement signals touches upon a fundamental question, namely ‘Is it actually possible to find the values of these unknowns, given a data-record of measurement signals?’ In his celebrated paper on mathematical systems theory, Kalman introduced the so-called observability property for models in state-space format. His definition of this structural model property (together with its dual form, i.e. structural controllability) is restricted to so-called linear-time-invariant (LTI) system models¹. An LTI dynamic system model may represent, for example, a linear approximation of a large network of chemical reactions at a certain operation point (\bar{x}) that may be graphically summarized in a directed graph². Given an input-output signal pair $\{u(t), y(t)\}$ (where the input variables $u(t)$ can be thought of as known signals, manipulated by the user, and the output variables $y(t)$ are the dynamic measured response signals), both the unknown parameters θ and the internal states $x(t)$ need to be reconstructed from this signal pair. Think, for example, of concentrations of chemicals that are involved in a reaction scheme: While only a subset of their concentrations can be measured, other concentrations of chemical species have to be deduced from the given data on the basis of a model and a so-called ‘state-observer’³.

Kalman showed that for LTI system models an observability rank test suffices to theoretically address the question whether the state $x(t)$ can be reconstructed from the available data¹. For the observability question for *both* unknown states and parameters, the Kalman observability test immediately poses a difficult problem: Once we have augmented the state vector $x(t)$ with the model parameters θ to form a new state vector $\begin{pmatrix} x(t) \\ \theta \end{pmatrix}$, the system model becomes *non-linear*. This non-linearity introduces substantial difficulties in the observability analysis and the Kalman rank test can not be used straightforwardly. It took until the late 1970’s before this problem was actually understood in a satisfactory way using more advanced tools from the then emerging field of non-linear control theory^{4–6}. The problem with these advanced tools, however, is that the actual computation involves so-called Lie-derivatives and Lie-brackets⁶. Computation of higher order brackets and/or Lie-derivatives is, in general, cumbersome for large system models leading to unacceptable computation times and memory requirements.

To provide context, we start with a simple non-linear example and then extend the approach we take to large non-linear system models. In a review paper⁷ several influenza models are analysed for identifiability of their parameters. The following 4-state model is an example of a target cell-limited influenza model with delayed virus production and reads

$$\frac{dT(t)}{dt} = -\beta T(t) V(t) \quad (1)$$

Biometris, Department of Mathematical and Statistical Methods, Wageningen University and Research, Wageningen, 6708 PD, The Netherlands. Correspondence and requests for materials should be addressed to J.D.S. (email: hans.stigter@wur.nl)

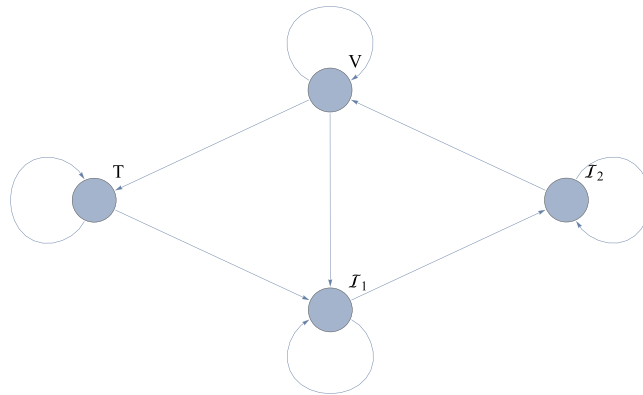


Figure 1. Directed graph summarizing the information flow in the Influenza model.

$$\frac{dI_1(t)}{dt} = \beta T(t) V(t) - k I_1(t) \quad (2)$$

$$\frac{dI_2(t)}{dt} = k I_1(t) - \delta I_2(t) \quad (3)$$

$$\frac{dV(t)}{dt} = p I_2(t) - c V(t) \quad (4)$$

where $T(t)$ is the number of uninfected target (epithelial) cells, $I_1(t)$ the number of latent infected epithelial cells, not producing virus, $I_2(t)$ the number productively infected epithelial cells, and $V(t)$ is the infectious viral titer which is the *only* variable measured. The parameters to be determined from the $V(t)$ readings are β , δ , c , k , p . In addition to these model parameters, we also include the initial conditions $T(0)$, $I_1(0)$, $I_2(0)$, $V(0)$ as unknowns and analyse whether their values can be reconstructed from the $V(t)$ readings. A directed graph of the influenza model (Fig. 1) shows that the information flow between the states $T(t)$, $I_1(t)$, $I_2(t)$, and $V(t)$ is good, meaning that the measured state $V(t)$ receives inflow from all other states, either directly or indirectly, so that one would expect that all parameters should be identifiable. Using the implicit function method Miao *et al.* found that the parameter p in equation (4), however, cannot be identified from the sensor readings⁷. In the Supplemental Material 1.1 we show that this parameter is indeed totally correlated with the initial conditions $T(0)$, $I_1(0)$, and $I_2(0)$, meaning that four (out of nine) parameters are completely undetermined in this specific experimental set-up. Of course, this negative result leads to quite a dramatic conclusion, namely: Whatever the number and quality of the viral titre data $V(t)$ is, it is *impossible* to ascertain all unknowns from these data. Ignoring this insight one would, by applying one of the many available parameter estimation algorithms, arrive at results that are totally unreliable for the correlated unknowns in the model.

An enormous effort has been put into finding algorithms to assess the observability question for non-linear systems⁸. Most of these algorithms are *algebraic* observability tests that suffer from one important drawback: Due to their symbolic/algebraic nature, computer algebra systems must be used to calculate the derivatives of the output equation repeatedly and, while this is in many cases tractable for small scale system such as the influenza model, the symbolic computations become intractable for large scale systems in view of the computation time needed. Chappell *et al.*⁹ already noted this problem in the early nineties and, fairly recently, their conclusion was re-confirmed for the current state-of-the-art algorithms in this field¹⁰. Other, numerically based methods, such as the sensitivity based identifiability test have mainly been used to assess *practical* identifiability¹¹, meaning that these methods are used to find which parameters can be estimated from *real* data that include noise-corrupted observations. If we assume perfect, noiseless data, *theoretical* observability addresses the question whether it is mathematically possible to find unique values for the unknowns in the model.

Liu *et al.*¹² note in their recent and excellent review on the control principles of complex systems that control theory has unfortunately not been very successful in answering the theoretical observability question for complex systems. The currently best available *symbolic* algorithm is due to Sedoglavic¹³ and has been applied to large models by Anguelova¹⁴. In Liu's observability paper it was used to verify a novel graph-theoretical based observability test². This novel test comprises a matching algorithm and it is capable of quickly finding the root-compartments in a directed graph of the dynamic model. From these root compartments sensors must then be selected for the system to become observable. The graph-theoretical approach uses as its only input the adjacency graph of a given network and there is no direct reference to the algebraic relations that connect the state $x(t)$ and the parameters θ . In addition, Liu's question only includes observability of initial *states* in his analysis, while the parameters that are involved in the interactions are assumed known or, better, only deemed relevant for the analysis in the sense that their values are either zero (there is no connection between two nodes i and j) or non-zero (there is a connection between two nodes i and j).

Results

Yet, we claim there is a way out of the cul-de-sac of finding an *algebraic* answer to the structural observability question. In this escape the so-called parametric output sensitivity matrix plays a crucial role. To explain this role, consider the dynamic behaviour of the general non-linear state space model

$$\frac{dx(t)}{dt} = f(x(t), u(t), \theta) \quad (5)$$

$$x(0) = x_0 \quad (6)$$

$$y(t) = h(x(t), u(t), \theta) \quad (7)$$

with (5) and (7) the dynamical model and measurement equations, respectively, $\dim(x) = n$, $\dim(u) = r$, and $\dim(\theta) = p$. If we include the initial conditions x_0 as *additional parameters* in the parameter vector θ , we get a $q = (n + p)$ -dimensional parameter vector θ . For this augmented parameter vector θ the sensitivity dynamics become:

$$\frac{dx_\theta(t)}{dt} = \frac{\partial f}{\partial x} x_\theta(t) + \frac{\partial f}{\partial \theta} \quad (8)$$

$$y_\theta(t) = \frac{\partial h}{\partial x} x_\theta(t) + \frac{\partial h}{\partial \theta} \quad (9)$$

with $x_\theta = (\partial x)/(\partial \theta)$ and $y_\theta = (\partial y)/(\partial \theta)$. The matrix x_θ has dimensions $(n \times q)$. For the extra initial condition parameters x_0 the sensitivities are initialized as $x_\theta(0) = I_n$, i.e. the identity matrix, while $x_\theta(0) = O_{n \times p}$ (a zero-matrix), for the original system parameters. Note that the sensitivity equation (8) constitutes q *linear, time-varying, systems* that run in parallel and whose seed of dynamics originates from the original system model (5). Each of these q sensitivity systems generates its own dynamics for one particular parameter θ_i , $i = 1, \dots, p$, $p + 1, \dots, q = n + p$. The parametric output sensitivities $y_\theta(t)$ can now be calculated by integrating (5) and (8) and substituting the solution into (9). This allows construction of the well known *sensitivity matrix*^{15,16}

$$S(t_0, \dots, t_N, \theta) = \begin{pmatrix} \frac{\partial y_1(t_0)}{\partial \theta_1} & \dots & \frac{\partial y_1(t_0)}{\partial \theta_q} \\ \vdots & & \vdots \\ \frac{\partial y_m(t_0)}{\partial \theta_1} & \dots & \frac{\partial y_m(t_0)}{\partial \theta_q} \\ \vdots & & \vdots \\ \frac{\partial y_1(t_N)}{\partial \theta_1} & \dots & \frac{\partial y_1(t_N)}{\partial \theta_q} \\ \vdots & & \vdots \\ \frac{\partial y_m(t_N)}{\partial \theta_1} & \dots & \frac{\partial y_m(t_N)}{\partial \theta_q} \end{pmatrix} \quad (10)$$

This matrix can be thought of as a series of “snapshots” of the sensitivity dynamics $y_\theta(t)$ that have been stacked vertically, one for each time instant t_i . When moving downwards from top to bottom in the above matrix, we are watching a stroboscopic film, each frame consisting of a snapshot of the sensitivity dynamics at time instant t_i . A *local* (or at-a-point) structural observability test may now be formulated that is based on a rank-test of the sensitivity matrix^{17–19}. A full rank sensitivity matrix is a *sufficient* condition for observability⁷. This fact can be re-phrased into its contrapositive as: If a system is *not* observable, then the matrix $S(t_0, \dots, t_N, \theta)$ *must* be rank deficient.

A well-known tool to assess the rank of a given matrix is the singular value decomposition²⁰ (SVD) that allows the matrix to be written as a sum of equally sized matrices that decrease in dominance (or singular value σ_i) as

$$S(t_0, \dots, t_N, \theta) = u_1 \sigma_1 v_1^T + \dots + u_q \sigma_q v_q^T \quad (11)$$

Having this decomposition of a given sensitivity matrix $S(t_0, \dots, t_N, \theta)$ available for a value of θ , the observability question can be rephrased to one that asks for the nullity of the *last* singular value σ_q . Indeed, if for some $k \leq q$, $\sigma_k = \dots = \sigma_q = 0$, we know that $S(t_0, \dots, t_N, \theta)$ is rank deficient and finding unique values of all parameters in the model is very likely to fail. Hunting down one or several zero singular values is therefore an important mission in answering the question of a lack of observability. In fact, detecting zero singular values can be considered as part 1 of our strategy. In the supplemental material we demonstrate that there is, in addition, a part 2, namely the symbolic verification of a lack of identifiability, once we have detected this on the basis of a SVD of the sensitivity matrix (10).

Recent work^{19,21} shows that observability is decided upon by setting a *threshold* on the eigenvalues of the Fisher Information Matrix (FIM) below which a model is considered not identifiable. A proxy of the FIM that

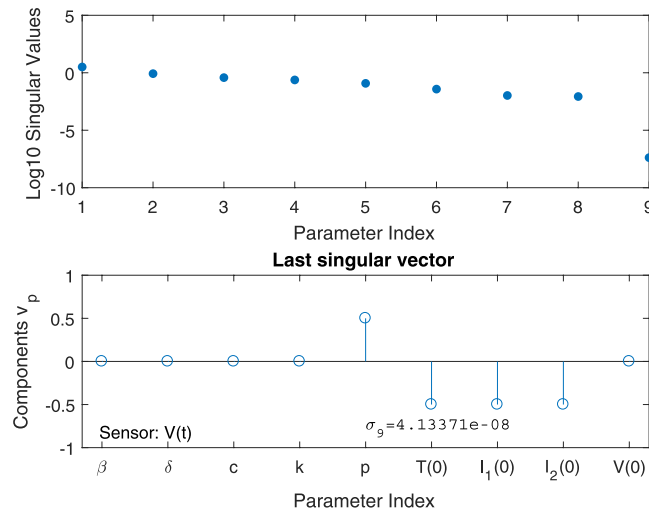


Figure 2. Observability signature for the Influenza model: Since there are nine parameters to be determined, there are nine columns in the matrix $S(t_0, \dots, t_N, \theta)$. An SVD of this matrix then yields nine singular values $\sigma_1, \dots, \sigma_9$. These singular values are presented in the top graph on a logarithmic scale. Note the sudden *gap* between σ_8 and σ_9 of approximately 3.5 decades on the logarithmic scale, indicating a true zero for the last singular value and, hence, a possible lack of observability. The bottom graph clearly shows that the parameters q , $T(0)$, $I_1(0)$, and $I_2(0)$ are involved in this total correlation, suggesting that there is an algebraic relationship between these parameters that deems the model un-identifiable.

has exactly the same rank as (10) is simply calculated by multiplying the transposed of the sensitivity matrix with itself. A known source of inaccuracy is that a rank test on the FIM is not as precise as a rank test on the original sensitivity matrix (10), see also Moore's paper where this is already noted²². Another source of inaccuracy is that only one trajectory (or simulation) of the model is considered, corresponding to one value for the parameters in the vector θ . Here we *combine* several trajectories for different values of the parameter vector θ and we will demonstrate (see Supplementary Information 1.2) that this leads to a dramatic improvement in the accuracy of a zero-singular-value-detection. In addition, a wealth of information can be deduced by first looking at the singular vectors $\{v_i, i = k, \dots, q\}$ for which $\sigma_i = 0$, i.e. the basis of the nullspace of $S(t_0, \dots, t_N, \theta)$ ^{18,23}. We demonstrate that using this information leads to a tremendous simplification and allows a tractable symbolic computation to be performed that further tests the SVD-detected lack of observability and validates these results.

The nullspace basis vector that shows the correlations. Consider the case that one singular value is zero, i.e. $\sigma_q = 0$. The corresponding right singular vector v_q in (11) provides detailed information on the parameters that are correlated and cause the model to be unidentifiable. Indeed, the entries in v_q clearly show which of the columns in the sensitivity matrix are linearly dependent and we therefore only need to inspect the non-zero entries in this vector to find out *which* parameters are correlated. Note that this is not necessarily a pairwise correlation between two parameters but, rather, a complete list of those parameters that are involved in a total correlation. As an example consider, again, the Influenza case study. Calculating a SVD of the sensitivity matrix for random values of θ^i yields what we term as the *observability signature* of the given model and experimental configuration²³. In general, the observability signature is summarized in two graphs, namely (i) a graph of the q singular values $\{\sigma_i, i = 1, \dots, q\}$ of the sensitivity matrix (10), in decreasing order, as a function of their index i , and (ii) a graph of the last right-singular vectors $\{v_i, i = k, \dots, q\}$ in (11) that correspond to zero singular values with, on the horizontal axis, the vector entry index $l = 1, \dots, q$ and on the vertical axis the corresponding values in v_i . In Fig. 2 a graph of the nine entries of the right singular vector v_9 for the influenza model is presented. Its corresponding singular value σ_9 was found in the order $\mathcal{O}(10^{-8})$ and, when comparing this to the other 8 singular values, one can observe a clear *gap* between σ_8 and σ_9 of approximately 3.5 decades on the logarithmic scale (Fig. 2). This is typical for a rank deficient sensitivity matrix. From Fig. 2, bottom panel, we further conclude that *four* out of nine parameters are involved in a total correlation. These are the parameters p , $T(0)$, $I_1(0)$, and $I_2(0)$. Knowing this correlation between these specific parameters, we may re-parametrise the model on the basis of a simplified symbolic computation (see Supplementary Information 1.1). Since the sensitivity of the output with respect to the remaining 5 parameters is *not* present in the SVD-detected nullspace, these parameters can, in principle, be identified from the sensor readings.

Pushing the Limits: Complex Systems. An immediate question is whether the idea of a rank test of $S(t_0, \dots, t_N, \theta)$ can be scaled up to large simulation models with many states and parameters in their structural equations. The results we present in this paper demonstrate that the answer is definitely positive. This success stems from a substantial part from the SVD algorithm in combination with an accurate integration of equations (5–8). To guarantee a precise solution of the sensitivity equations we employed the use of complex derivatives²⁴ that allow a very accurate computation of the Jacobi matrices $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial \theta}$ in (8), thereby decreasing the absolute tolerance on the

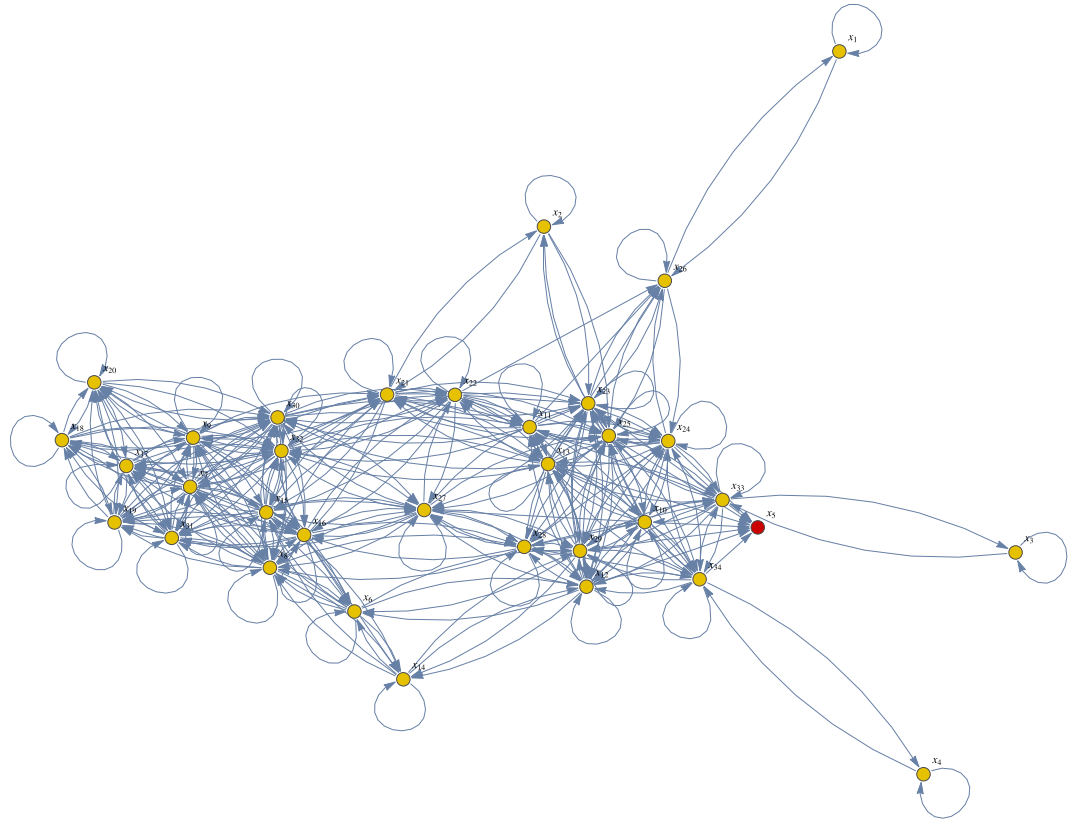


Figure 3. Directed graph Chinese Hamster case: The 34 nodes in this graph correspond to the 34 states $x_i(t)$, $i = 1, \dots, 34$. An arrow from node x_i to node x_j means that the j^{th} differential equation in the model contains a term with the state variable $x_i(t)$ in it. Self-loops show that most nodes also influence their own dynamics. Finally, the connected nodes are coloured in yellow and red so that one can easily see that node x_5 is a sink with no outgoing arrows.

sensitivities $y_{\theta}(t)$ to 10^{-20} . In addition, combining several sensitivity matrices for different parameter values also leads to a substantial improvement of the accuracy on a zero-singular-value detection as demonstrated in Supplementary Information 1.2. We have, until now, not seen a case yet in which the algorithm failed to demonstrate a lack of observability for models that, in the literature, are known to be unobservable. Our findings are substantiated with the following three large case studies, namely (i) The Chinese Hamster ovary cell model, (ii) the JAKSTAT model, and (iii) the MAP-Kinase model. These examples were chosen since the first two are the largest ones available in the current literature on local structural identifiability. Since algebraic/symbolic computations have been performed for these two large case studies, our results can therefore directly be compared. Furthermore, the third case study is a good demonstration of Liu’s conjecture on the observability of complex systems¹².

Chinese Hamster Ovary cell model. The first large-scale example model describes the dynamic behaviour of 34 metabolites in three compartments (fermenter, cytosol, and mitochondria) of a Chinese Hamster Ovary cell. Thirteen of these metabolites can be measured directly, corresponding to 13 measurable states in the output equation (7), namely $x_1, x_2, x_3, x_4, x_5, x_{11}, x_{13}, x_{15}, x_{21}, x_{27}, x_{29}, x_{30}$, and x_{32} . For exact details we refer to the paper²⁵ that includes this example as one of a series of benchmark problems for structural identifiability analysis. This non-linear model has a total of 117 system parameters and 34 states in its structural equations. In Fig. 3 a directed graph is presented that summarizes the exact wiring diagram of these 34 states. Assuming the initial conditions are also unknowns, there are $117 + 34 = 151$ parameters to be identified from the output signals. This identifiability case study is considered “very challenging”, since the symbolic computations to address identifiability for this model are so enormous that it takes many hours (if not days) to complete this task. These symbolic computer algebra results clearly show this model is *non-identifiable*²⁵.

In Fig. 4 the observability signature for this large model is presented in two sub-plots. Generating the observability signature is a *matter of seconds* and only comprises a number of model simulations. Figure 4 immediately shows the typical gap in the singular values, indicating a possible lack of identifiability. There are two singular values detected as zero in this particular case and these correspond to two groups of correlated parameters. The “zerness” of these last two singular values is even more apparent when looking at the nullspace graph of the last singular vectors v_{150} and v_{151} . Clearly there are four parameters involved in the correlations and, since two zero singular values are detected, these four parameters can be classified into two groups of two correlated parameters each (see Supplementary Information 1.2).

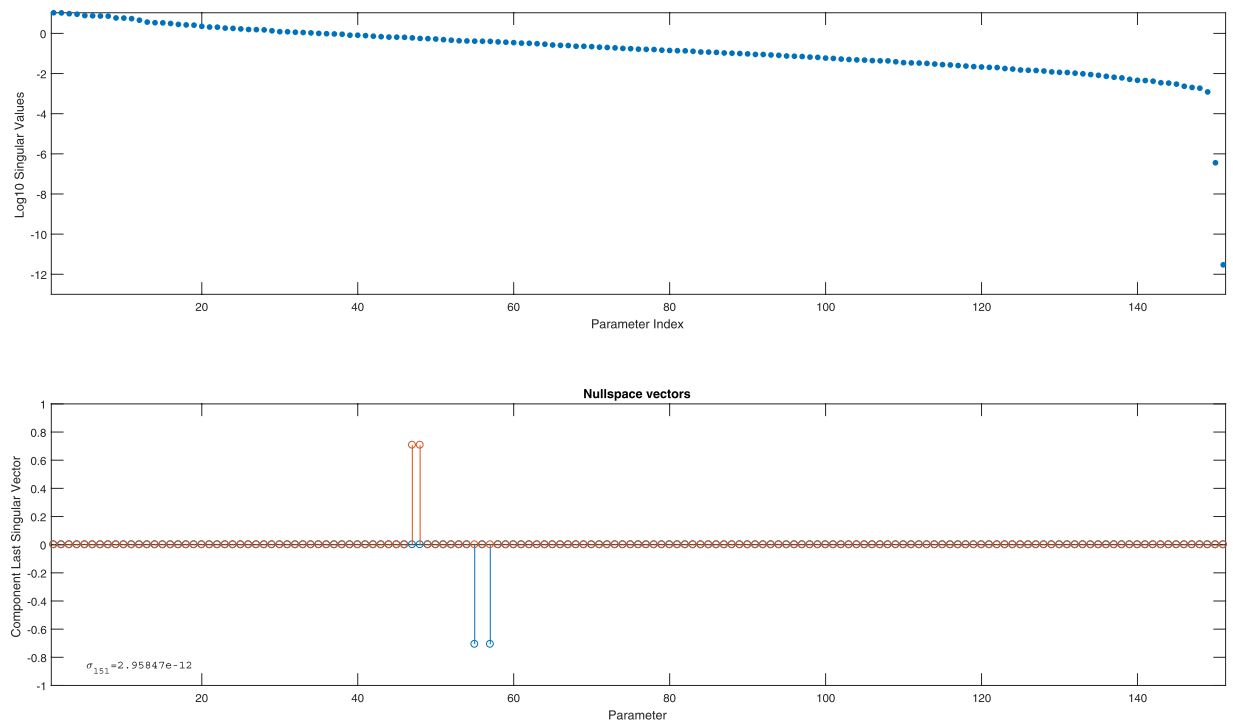


Figure 4. Observability signature Chinese Hamster model. The top graph shows all singular values for the 151 parameters in the model (including the parametrized initial conditions). A gap is clearly visible, indicating a lack of identifiability. The bottom graph shows the nullspace vectors for the sensitivity matrix corresponding to the two smallest singular values (order 10^{-12}).

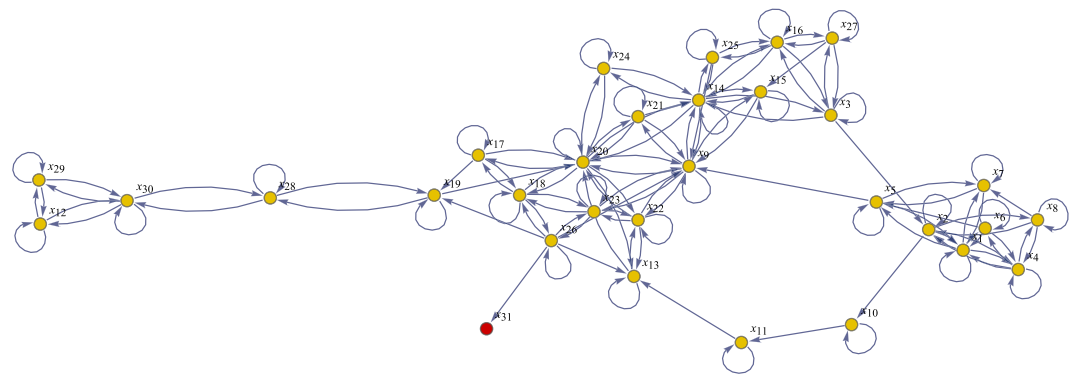


Figure 5. Directed graph JAKSTAT model: The 31 nodes in this graph correspond to the 31 states $x_i(t)$, $i = 1, \dots, 31$. The connected nodes are coloured in yellow and red so that one can easily see that node x_{31} is the only one-node-sink with no outgoing arrows.

JAKSTAT model. In her paper on minimal output sets that allow identification of all parameters in a model, Anguelova checks observability of the JAKSTAT model (summarized in a directed graph in Fig. 5) after first sifting out so-called translation and scaling (Lie-)symmetries. The final observability test is performed with Sedoglavic' algorithm and takes most of the computation time¹⁴. This well-known signalling model²⁶ comprises 31 states and 51 parameters and was shown to be un-identifiable. We computed in roughly 2 seconds the observability signature (using five trial values θ^j) for the JAK-STAT model for a sensor set that does *not* include the states x_{10} and x_{11} (these two states are responsible for a lack of identifiability). The result is summarized in Fig. 6. Again, we observe a clear gap in the singular values, and the bottom panel in this figure immediately shows that parameters θ_{14} , θ_{51} , $x_{10}(0)$, $x_{11}(0)$ are responsible for this. This result verifies Anguelova's findings¹⁴.

MAP Kinase model. In our third example we analyse the well-known MAP Kinase model²⁷ by Schoeberl *et al.*. Details of this model have been widely discussed in the literature. The number of states is 100, while there are 84 system parameters to be identified. Including the initial conditions as unknowns we have 184 parameters to

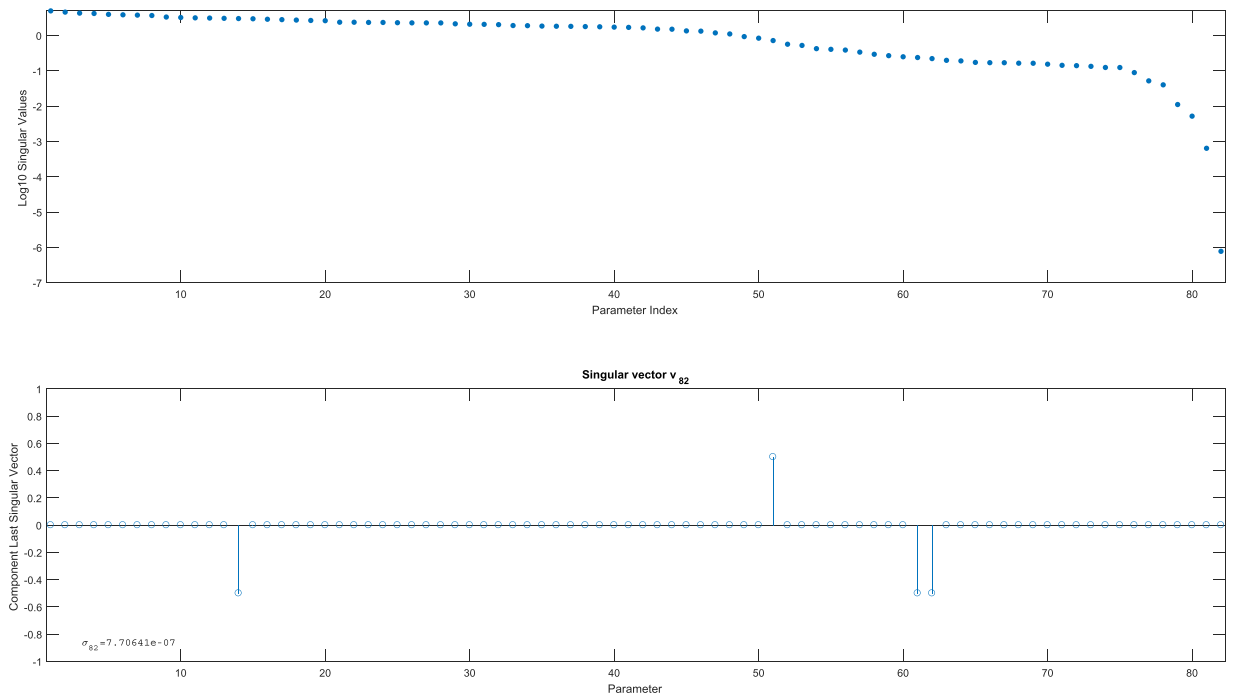


Figure 6. Observability signature JAK-STAT model for a concatenation of five trials for the sensitivity matrix (10) when x_{10} and x_{11} are *not* in the sensor set. The top graph shows all singular values for the 82 parameters in the model (including the parametrized initial conditions). The gap indicates a lack of observability and the bottom graph summarizes the parameters involved, namely θ_{14} , θ_{51} , $x_{10}(0)$, $x_{11}(0)$. A symbolic verification confirmed a lack of observability for this particular sensor set.

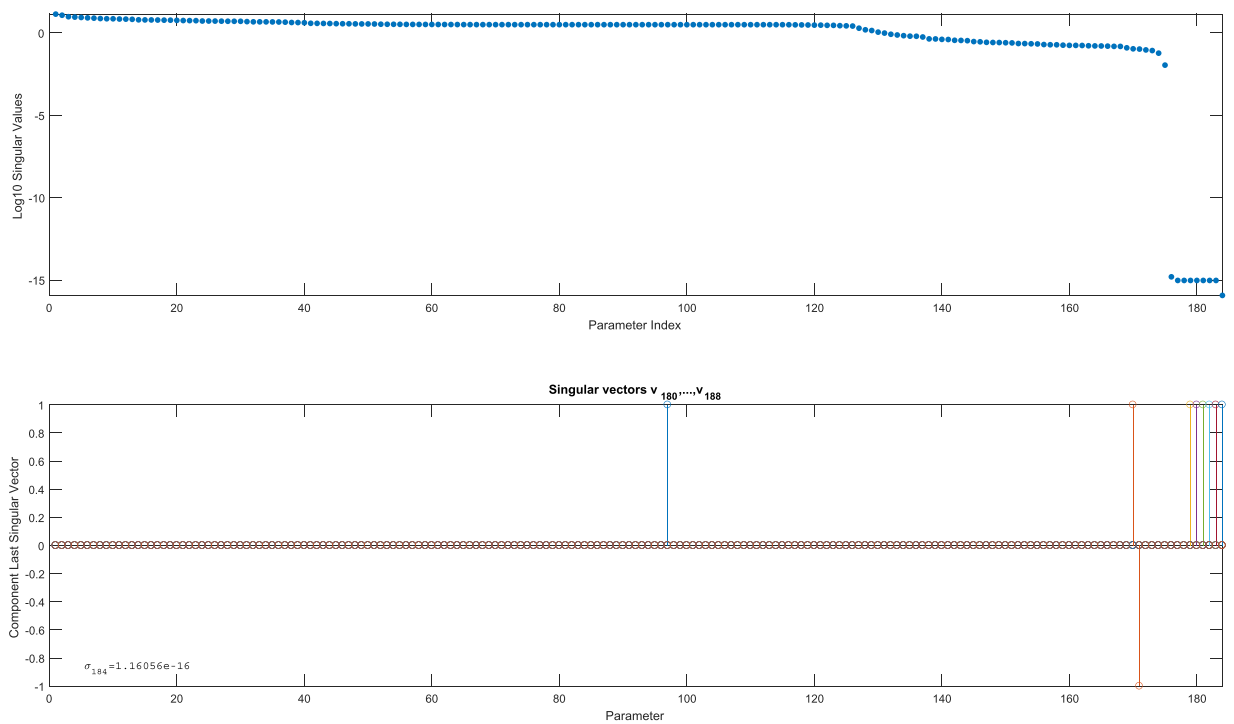


Figure 7. Observability signature Map Kinase model. The top graph shows all singular values for the 184 parameters in the model (including the parametrized initial conditions). Again, a gap is clearly visible, indicating a lack of identifiability if the one-node-sinks x_{13} , x_{86} , x_{87} , x_{95} , x_{96} , x_{97} , x_{98} , x_{99} , and x_{100} are not included in the sensor set. The bottom graph shows the nullspace vector that indicates precisely these nodes.

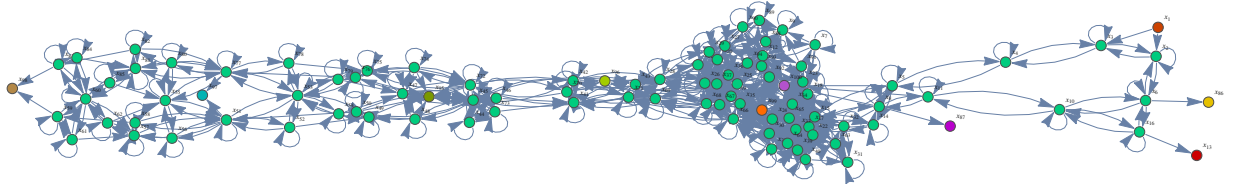


Figure 8. Directed graph MAP Kinase model. Connected components are labelled with colors. There are nine one-node-sinks observable.

be identified from the sensor readings. For this model we first applied our method using various sensor combinations, assuming that the states could be measured directly (which is not realistic in practice). The sensors were assigned as a random permutation of 50 out of 100 possible direct state measurements and 100 randomly chosen sensor sets were then analysed. In all cases gaps were observable, see Fig. 7. These gaps correspond to state-initial-condition-parameters x_0 only. After further inspection, these states were all found to be *sinks* in the network structure that have no outgoing arrows. For the initial condition of such a sink-node to be estimated, one must include it as a sensor because otherwise, of course, no information on its corresponding parameter in the parameter vector x_0 can be inferred from the other sensors. Put differently, our approach results in an alternative method of finding the sinks of a network from which no information of their initial condition can be received on the sensor readings if these nodes are not part of the sensor set. In fact, these results confirm Liu's conjecture² that, in order to have an observable model, one has to include one of the sensors in each root compartment. For the MAP Kinase model this means that at least sensors x_{13} , x_{86} , x_{87} , x_{95} , x_{96} , x_{97} , x_{98} , x_{99} , and x_{100} have to be in the sensor set in order to identify (theoretically) all unknowns in the model. This result is completely in tune with the directed graph presented in Fig. 8 where exactly these nodes are highlighted in colour as one-node-sinks of the directed graph associated with the MAP Kinase model.

Discussion

The results in this paper show that observability of non-linear large scale models can be achieved in a short computation time that is orders of magnitudes smaller than what is currently common-practice using algebraic observability tests. The avenue we followed opens up interesting perspectives for future research. For example, it is well known that structural controllability is the dual version of the structural observability problem discussed here. Extending the analysis to local controllability is therefore within easy reach, and will be the subject of further investigation. We note that preliminary results on this front are promising.

Finally, we have restricted our examples to complex biological systems but, of course, the methodology can be easily used in other applications as well. For example, observability is an important topic in aeronautics²⁸ and helps the engineer to decide on optimal design and sensor placement for position and velocity estimation of space vehicles. Another application is fluid dynamics in which exactly the same type of questions need to be addressed²⁹. These are just two of the many examples that can be analysed in detail with the algorithm we have explained and demonstrated in the above.

References

- Kalman, R. Mathematical description of linear dynamical systems. *SIAM Journal of Control* (1963).
- Liu, Y., Slotine, J. & Barabási, A. Observability of complex systems. *Proceedings of the National Academy of Sciences* (2012).
- Bastin, G. & Dochain, D. *On-Line Estimation and Adaptive Control of Bioreactors* (Elsevier, Amsterdam, 1990).
- Hermann, R. & Krener, A. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control* **AC-22**(5), 728–740 (1977).
- Tunali, T. & Tarn, T. New results for identifiability of nonlinear systems. *IEEE Transactions on Automatic Control* **AC-32**(2), 146–154 (1987).
- Isidori, A. *Nonlinear Control Systems* (Springer-Verlag, Heidelberg, 1989).
- Miao, H., Xia, X., Perelson, A. & Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review* **53**(1), 3–39 (2011).
- Walter, E. *Identification of State Space Models* (Springer, Berlin, 1982).
- Chappell, M., Godfrey, K. & Vajda, S. Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods. *Mathematical Biosciences* **102**, 41–73 (1990).
- Chis, O., Banga, J. & Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE* **6**, (2011).
- Brun, R., Reichert, P. & Künsch, H. Practical identifiability analysis of large environmental simulation models. *Water Resources Research* **37**(4), 1015–1030 (2001).
- Liu, Y.-Y. & Barabasi, A.-L. Control principles of complex systems. *Review of Modern Physics* **88** (2016).
- Sedoglavic, A. A probabilistic algorithm to test local algebraic observability in polynomial time. *Journal of Symbolic Computation* **33**, 735–755 (2002).
- Anguelova, M., Karlsson, J. & Jirstrand, M. Minimal output sets for identifiability. *Mathematical Biosciences* **239**, 139–153 (2012).
- Eykhoﬀ, P. *System Identification: Parameter and State Estimation* (Wiley-Interscience, London - New York, 1974).
- Bard, Y. *Nonlinear Parameter Estimation* (Academic Press, 1974).
- Jacquez, J. Parameter estimation: local identifiability of parameters. *American journal of physiology: endocrinology and metabolism* **258**, 727 (1990).
- Stigter, J. & Molenaar, J. A fast algorithm to assess local structural identifiability. *Automatica* **58**, 118–124 (2015).
- Quaiser, T. & Mönnigmann, M. Systematic identifiability testing for unambiguous mechanistic modeling—application to JAK-STAT, MAP kinase, and NF- κ B signaling pathway models. *BMC Systems Biology* **3** (2009).
- Golub, G. & van Loan, C. *Matrix Computations* (The John Hopkins University Press, 1996).

21. Fink, M. & Noble, D. Markov models for ion channels: Versatility versus identifiability and speed. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **367**, 2161–2179 (2009).
22. Moore, B. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Automat. Control* **AC-26**, 17–32 (1981).
23. Stigter, J., Beck, M. & Molenaar, J. Assessing local structural identifiability for environmental models. *Environmental Modelling & Software* **93**, 398–408 (2017).
24. Martins, J., Sturdza, P. & Alonso, J. The connection between the complex-step derivative approximation and algorithmic differentiation. *American Institute of Aeronautics and Astronautics* (2001).
25. Villaverde, A. F., Barreiro, A. & Papachristodoulou, A. Structural identifiability of dynamic systems biology models. *PLoS Computational Biology* (2016).
26. Yamada, S., Shiono, S., Joo, A. & Yoshimura, A. Control mechanism of jak/stat signal transduction pathway. *FEBS Lett.* 190–196 (2003).
27. Schoeberl, B., Eichler-Jonsson, C., Gilles, E. & Müller, G. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature Biotechnology* **20** (2002).
28. Kaufman, E., Lovell, T. & Lee, T. Nonlinear observability for relative orbit determination with angles-only measurements. *Journal of Astronautical Sciences* (2016).
29. Krener, A. & Ide, K. Measures of unobservability. In *Proceedings of the 48th conference on decision and control* (IEEE, 2009).

Author Contributions

All authors are in the Department of Mathematical and Statistical Methods, Wageningen University, The Netherlands. J.D.S. contributed in the writing of the paper, theoretical analysis, and programming/computations. D.J. contributed in providing case studies, discussion and analysis of results. J.M. contributed in the writing of the paper, discussion and analysis of results. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-16682-x>.

Competing Interests: The authors declare that they have no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017