Received: 27 April 2023 | Accepted: 23 July 2023

DOI: 10.1111/nrm.12382

Natural Resource Modeling WILEY

Computing parameter identifiability and other structural properties for natural resource models

J. D. Stigter

Department of Mathematical and Statistical Methods, Wageningen University, Wageningen, The Netherlands

Correspondence
J. D. Stigter, Department of Mathematical and Statistical Methods, Wageningen University, Wageningen, 6708 PD, The Netherlands.
Email: hans.stigter@wur.nl

Abstract

For dynamic natural resource models a few fundamental questions are often omitted in practice such as (i) “Is it possible to calibrate the model?” and (ii) “Can we reconstruct all unknown dynamic state variables given a certain data record?” Other related questions are (iii) “Does the system allow the state variables to be controlled (or managed) to desired optimal values (e.g., in case of maximum sustainable yield)?” In this paper we highlight and discuss a software tool (the StrucID App) that allows a rapid evaluation of these fundamental and structural model properties that are important to study before a calibration and subsequent simulation of the model.

KEYWORDS

identifiability, model reduction, parameter estimation

1 | INTRODUCTION

For calibration, observation, and model-based management of ecosystems, there are a few fundamental questions that need to be addressed regarding the structural properties of the proposed model. As it turns out, these fundamental questions are difficult to address in practice. For example, if we want to know if a given data record allows a unique reconstruction of the model parameter values, then applying the classical identifiability tests that stem from nonlinear control theory will quickly lead to an *infeasible* problem that cannot be solved. Because of the inherent nonlinear behavior and, more general, the high complexity of the advanced models at hand, such an analysis is often too complicated using text-book methodology only. In this paper we introduce the StrucID App that allows a convenient analysis of large (dynamic) models in a rapid way. The App can be used as a model-evaluator and allows analysis of observability (of state variables), and identifiability (of parameters in the model). Analysis of the controllability of state variables is also possible with the StrucID App, but will not be discussed here.

A few words on the classical approach for the analysis of structural properties of dynamic system models: One of the main reasons such analyses are not performed in practice is that nonlinear control theory is not applicable for the more advanced simulation models used in practice that, typically, may contain some 5+ state variables and some 20+ parameters, or even more (Barabasi & Oltvai, 2004; Liu et al., 2012). The reason for this is that so-called Lie-derivatives (Kwatny & Blankenship, 2000) have to be computed repeatedly for the given output (or measurement) signals. It is not uncommon that five derivatives or more have to be computed with a computer-algebra package. With each higher-order derivative, the obtained results of earlier obtained (lower-order) Lie-derivatives, need to be substituted, leading to several A4 pages of output that quickly becomes intractable and even not computable in a reasonable amount of time. In Chappell's paper (Chappell & Godfrey, 1992) this stumble-block was already noted and identified as the most important bottleneck that hampers any computation on identifiability, observability, or controllability, in practical situations. Of course, over the years computational resources have become better, but since the symbolic computations grow exponentially in terms of complexity, the increased computational speed is not a real remedy for large simulation models. Stigter and Molenaar (2015), however, found an alternative route that combines numerical integration of the parametric output sensitivities of the model with symbolic computations on a computer-algebra system. This allows for a rather dramatic decrease in central processing unit time. Their method underpins the approach taken in the StrucID App.

In the following we will demonstrate the applicability of the StrucID App in natural resource modeling on the basis of a few simple models. Because of the more didactical nature of this paper, we limit ourselves to more academic case studies but, needless to say, this can be extended easily to large simulation models as presented in, for example, Joubert et al. (2020).

2 | METHODOLOGY

The dynamic resource models are represented here in the familiar state-space format that stems for the field of systems and control theory (Kalman et al., 1969):

$$\frac{dx(t)}{dt} = f(x(t), u(t), \theta), \quad x(0) = x_0, \tag{1}$$

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

with $x(t)$ the state vector that includes the state variables (e.g., the population densities of n species that interact with one-another), $u(t)$ the input vector with variables that may be manipulated at will (e.g., harvest rates), and θ a set of parameters whose values are often assumed to be unknown and need to be reconstructed from the calibration data. In addition, both the vector field f (state dynamics) and the vector field h (observation equations) may be nonlinear, thereby introducing difficulties that hamper a simple analysis as given in, for example, Kailath (1980).

For the general nonlinear state space model observability tests have been developed (Isidori, 1989; Kwatny & Blankenship, 2000) that unfortunately can only be applied to small-scale examples because of the simple fact that the symbolic manipulations involved are so complicated that with a growing model size the computation becomes infeasible (Miao et al., 2011). As said earlier, in Stigter and Molenaar (2015) a connection was made between the symbolic observability tests and the dynamics of parametric output sensitivities. These sensitivities can easily be derived from the models (1) and (2) as

$$\frac{dx_\theta(t)}{dt} = \frac{\partial f}{\partial x} x_\theta(t) + \frac{\partial f}{\partial \theta}, \quad x_\theta(0) = 0, \tag{3}$$

$$y_\theta(t) = \frac{\partial h}{\partial x} x_\theta(t) + \frac{\partial h}{\partial \theta}, \tag{4}$$

where $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial \theta}$ are the Jacobi matrices of the vector field f with respect to the state variables and parameters,[1] and $x_\theta(t)$ is a *matrix* that contains the parametric state sensitivities. Once the above equations are solved on a time grid $[t_0, ..., t_N]$, we can build the familiar parametric output sensitivity matrix $S(t_0, ..., t_N, \bar{\theta})$ as

$$S(t_0, ..., t_N, \bar{\theta}) = \begin{pmatrix} \frac{\partial y_1(t_0)}{\partial \theta_1} & \cdots & \frac{\partial y_1(t_0)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial y_m(t_0)}{\partial \theta_1} & \cdots & \frac{\partial y_m(t_0)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial y_1(t_N)}{\partial \theta_1} & \cdots & \frac{\partial y_1(t_N)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial y_m(t_N)}{\partial \theta_1} & \cdots & \frac{\partial y_m(t_N)}{\partial \theta_p} \end{pmatrix}. \tag{5}$$

This matrix can be thought of as a series of snapshots of the parametric output sensitivities that have been stacked on top of one-another. Since a sufficient condition for observability is to have a full rank sensitivity matrix (Miao et al., 2011), we can perform a rank test on the above matrix and, in case of a rank that is lower than p, we can test the results symbolically as explained in Joubert et al. (2020). It is well known that any matrix and, more specifically, the matrix $S(t_0, ..., t_N, \bar{\theta})$ can be written as a sum of equally sized matrices and their associated *singular values* as

$$S(t_0, ..., t_N, \theta) = u_1\sigma_1 {v_1}^T + \cdots + u_p\sigma_p {v_p}^T \tag{6}$$

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

with U and V unitary matrices and Σ a diagonal matrix that contains the ordered singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. If we have the smallest singular value $\sigma_p = 0$, then clearly the sensitivity matrix is rank deficient. The corresponding singular vector v_p will then give us a lead in the question which parameters are exactly involved in the lack of identifiability as will be demonstrated in an example below.

All of the above computations are taken care of in our StrucID App that can run without the need to install additional software (stand-alone version) (Figure 1). The only thing that is needed is a text file that contains the model definition, including the measured outputs (without the need for measured data!). The text file for the simple example below (Lotka–Volterra predator–prey model) is given in Appendix A.

2.1 | Example: Lotka–Volterra predator–prey interaction

As a simple example, consider the well-known Lotka–Volterra model describing the nonlinear interaction of a predator that feeds on a prey:

$$x'(t) = \alpha x(t) - \beta x(t)y(t), \quad x(0) = x_0, \tag{8}$$

$$y'(t) = \delta x(t)y(t) - \gamma y(t), \quad y(0) = y_0 \tag{9}$$

with $x(t)$ the prey-density and $y(t)$ the predator-density, respectively. Now, consider the case where we can only take measurements of the predator-density $y(t)$. If we want to calibrate the

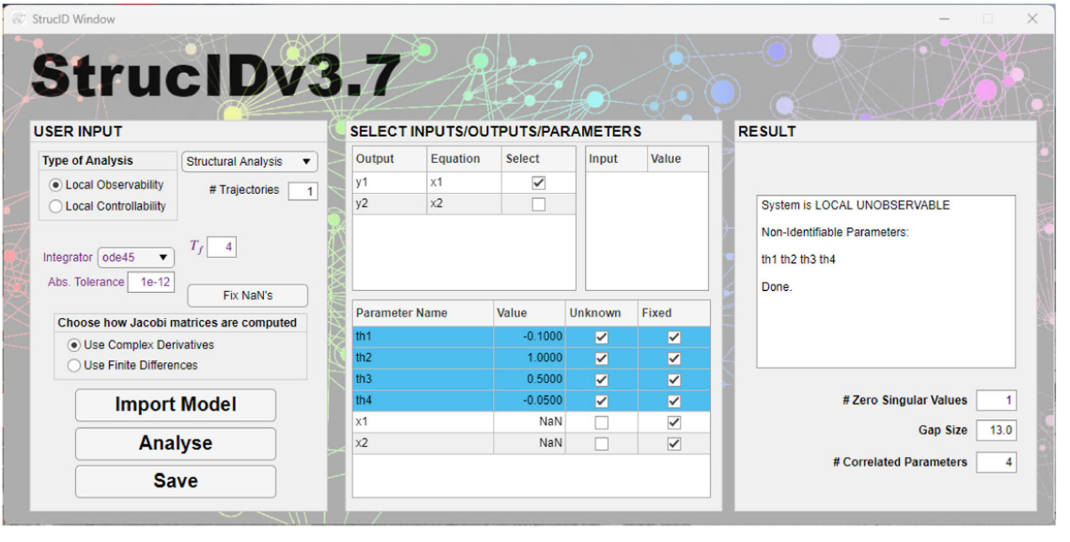


FIGURE 1 The StrucID App main window in which the model is loaded and totally correlated parameters (that are related through an algebraic relation) are highlighted in blue. In this case only sensor y_1 was selected as an output. The states x_1 and x_2 were not included in the analyses, but only the system parameters $\theta_1, \ldots, \theta_4$. Since only one singular zero is detected, there can only be one group of parameters that is totally correlated, namely, $\theta_1, \ldots, \theta_4$. Apparently an algebraic relation exists between these four parameters that may be investigated further with a computer-algebra package.

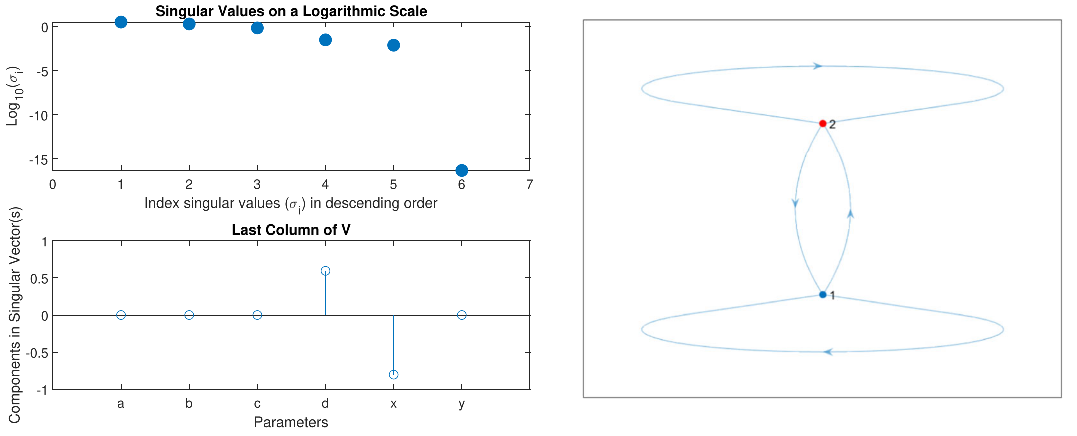


FIGURE 2 Singular values of the sensitivity matrix for the Lotka–Volterra model (left) and directed graph of the information flow in the model (right). The red node indicates that state variable 2, that is, the predator-density $y(t)$, is measured.

above population model, then we need to find values for the initial conditions x_0 and y_0 and, in addition, the population parameters α, β, γ, and δ from measurements of $y(t)$ *only*. Of course, y_0 can be directly observed and does not introduce an identifiability issue. Yet, finding the values of the other five parameters *does* impose a difficulty. If we look at the singular value graph generated by the StrucID App we get Figure 2.

From the singular values graph (one singular value for each parameter in the model) we quickly see that there is a *gap* visible (between the fifth and sixth singular values) that indicates the existence of a zero singular value. In the bottom graph we see that parameters δ and $x(0)$ are the two corresponding parameters that are involved in a total correlation. Apparently there is nonuniqueness or total correlation between these two parameters that hamper a reliable reconstruction of both x and δ. If one would apply a parameter estimation algorithm to find the parameter values from the predator-density measurements, then clearly the prey-density $x(t)$ in combination with the parameter δ cannot be estimated at all! As it turns out, there is a state-transformation $\tilde{x}(t) = \delta\ x(t)$ that puts the model into a format that has one parameter less than the original model while the simulated predator-density $y(t)$ is *exactly the same as the original model prediction*. This transformed model reads

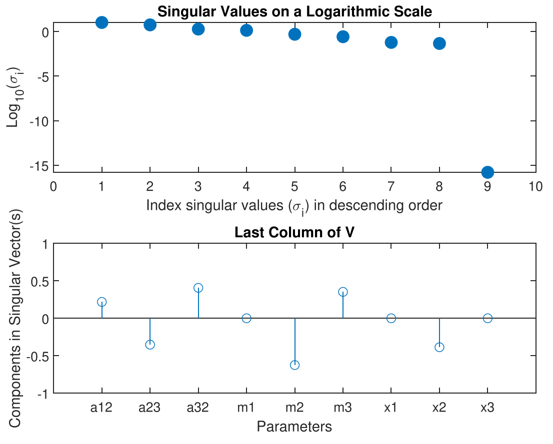
$$\tilde{x}'(t) = \alpha\tilde{x}(t) - \beta\tilde{x}(t)y(t), \quad \tilde{x}(0) = \tilde{x}_0, \tag{10}$$

$$y'(t) = \tilde{x}(t)y(t) - \gamma y(t), \quad y(0) = y_0, \tag{11}$$

where we now have removed the interaction parameter δ completely from the original model.

2.2 | Example: A linear compartmental model of three interacting species

We now consider three interacting species of which two ($x_2(t)$ and $x_3(t)$) stimulate each other's growth, while the first ($x_1(t)$) species preys upon the second ($x_2(t)$), whose density is not



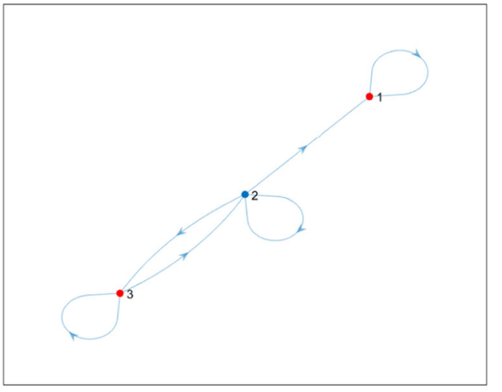


FIGURE 3 Singular values of the sensitivity matrix for the Linear Compartmental model (left) and directed graph of the information flow in the model (right). The red node indicates that state variables 1 and 3 are measured.

measured. In addition, the third species receives an external flow (e.g., sunlight) on which it grows (see Figure 3 for the direction of the flows in this model).

$$x_1{}'(t) = -m_1 x_1(t) + a_{12} x_2(t), \quad x_1(0) = x_{10}, \tag{12}$$

$$x_2{}'(t) = -(a_{12} + a_{32} + m_2) x_2(t) + a_{23} x_3(t), \quad x_2(0) = x_{20}, \tag{13}$$

$$x_3{}'(t) = a_{32} x_2(t) - (a_{23} + m_3) x_3(t) + u_1(t), \quad x_3(0) = x_{30} \tag{14}$$

with m_i, $i = 1, 2, 3$ the mortality rates for each species, and a_{ij} a flow (of energy/matter) from compartment j to i. The input $u_1(t)$ is the external flow input of energy/matter.

The results of an identifiability analysis with the StrucID App when measuring the first and third species densities are presented in Figure 3. It is immediately clear that not all parameters (including the initial conditions) can be estimated—even for this simple model. More specifically, we find that parameters $a_{12}, a_{23}, a_{32}, m_2, m_3$, and state variable $x_2(t)$ cannot be estimated independently from the available measurements of species densities 1 and 3. With the help of a computer-algebra package, after applying the methodology as presented in Joubert et al. (2020) and Stigter and Molenaar (2015), we can find the state transformation that destroys the zero singular value depicted in Figure 3. Without further details of a derivation, that would go beyond the scope of this paper, our findings are that with $\tilde{x}_2(t) = a_{12} x_2(t)$, $\phi_1 = a_{12} a_{23}$, $\phi_2 = \frac{a_{32}}{a_{12}}$, $\phi_3 = a_{12} + a_{32} + m_2$, $\phi_4 = a_{23} + m_3$ the number of system parameters can be reduced from 6 to 5 (while the number of state variables remains the same), yielding the transformed model:

$$x_1'(t) = -m_1 x_1(t) + \tilde{x}_2(t), \quad x_1(0) = x_{10}, \tag{15}$$

$$\tilde{x}_2(t) = -\phi_3 \tilde{x}_2(t) + \phi_1 x_3(t), \quad \tilde{x}_2(0) = \tilde{x}_{20}, \tag{16}$$

$$x_3'(t) = -\phi_2 \tilde{x}_2(t) - \phi_4 x_3(t) + u_1(t), \quad x_3(0) = x_{30}. \tag{17}$$

For clarity, the above-transformed model has exactly the same input–output behavior and, hence, simulates exactly the same species densities in comparison to the original model, but now with one parameter less!

2.3 | Example: Logistic growth and fishing effort

In the celebrated book on renewable resources by Clark (1976) logistic growth models are frequently used for fishery case studies as a basis for model-based optimal control (to maintain, e.g., maximum sustainable yield). For example, fishing of plaice and dab in the North Sea can be represented in a simplified model as

$$x'(t) = rx(t)\left(1 - \frac{x(t)}{K}\right) - q_1 Ex(t), \quad x(0) = x_0, \tag{18}$$

$$y'(t) = sy(t)\left(1 - \frac{y(t)}{L}\right) - q_2 Ey(t), \quad y(0) = y_0 \tag{19}$$

with r, s reproduction rates, K, L carrying capacities, q_i, $i = 1, 2$ are catch-ability quotients, E is fishing effort, and $x(t)$, $y(t)$ are population densities of plaice and dab, respectively. After a first analysis with StrucID, it was found that if all parameters (including catch-ability quotients) need to be estimated from measurements, then it is impossible to find the carrying capacities K and L of the two populations. This can easily be remedied by removing the parameters q_1 and q_2 from the list of unknowns, leaving six unknowns in total, that is, x_0, y_0, r, s, K, L. For that case it was found that *all* population densities and system parameters can, in principle, be estimated.

An interesting side-effect of performing the analysis with StrucID is that from the singular vector that is associated with the *smallest* singular vector (σ_6 in this case) we can clearly see which (combination of) parameter(s) is most difficult to estimate (but, in principle, is identifiable). Hereto, observe Figure 4. It is immediately clear from the entries in the last singular vector v_6 (bottom figure) that the carrying capacity L is the most difficult parameter to estimate from the measurements. Continuing the analysis by ruling out the most difficult parameters one by one, we found that the order of the "estimability" of the parameters for this case study is L, K, x_0, y_0, r, s with L being the most difficult parameter to find from the measurements. In other words, not only structural identifiability can be deduced from a singular value decomposition of the sensitivity matrix $S(t_0, ..., t_N, \bar{\theta})$, but also practical identifiability of the parameters involved. The latter is an important aspect in deducing which of the six parameters needs the most attention when reflecting on an experimental (or input) design that facilitates an informative set of data for a reliable parameter reconstruction of all six parameters in the model.

3 | CONCLUDING REMARKS

For a better insight in some key structural properties of a dynamic resource model it is very useful to perform an "a priori" analysis with the StrucID App. The package will not only show a lack of observability, but will also yield the (lack of) importance of certain combinations of parameters for the prediction of a particular model output. An interesting comment to make is

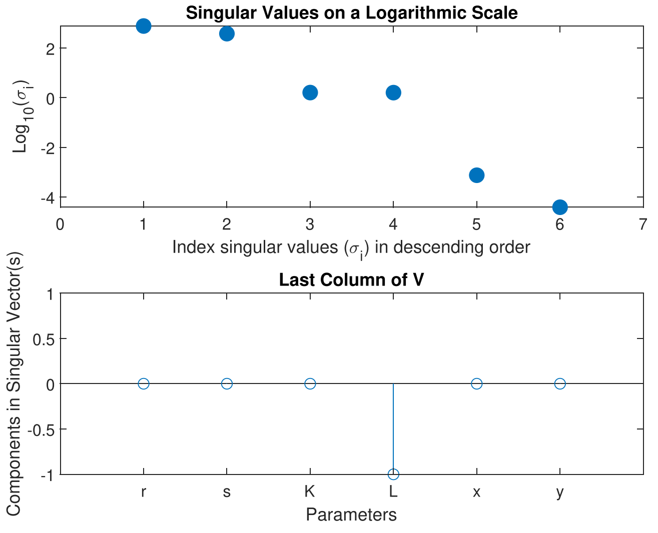


FIGURE 4 Singular values of the sensitivity matrix for the plaice/dab model. Note that, for this case, no zero singular values are detected and so, in principle, the six parameters in the model can be estimated from the measurements.

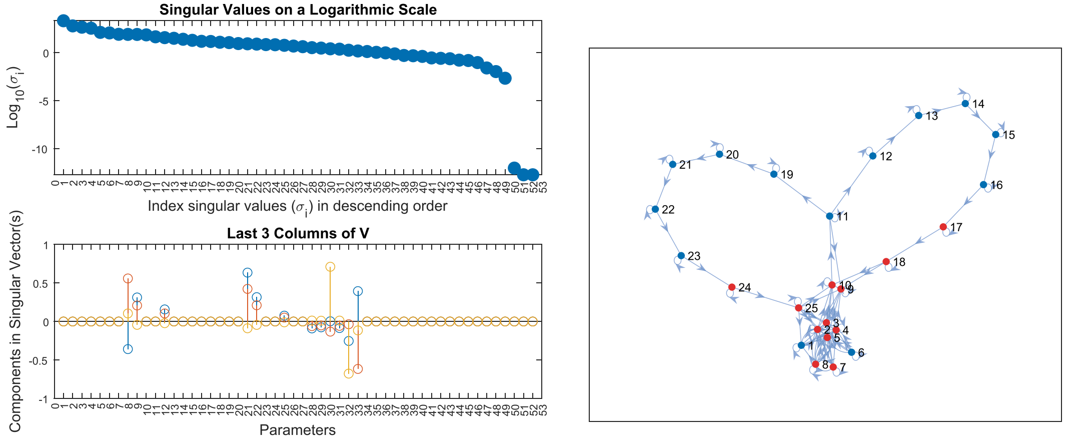


FIGURE 5 Singular values, including the last three singular vectors (corresponding with 3 zero singular values) for the sensitivity matrix for the JAKSTAT-Bachman model, together with a directed graph that shows the interactions between the state variables. The red dots show which of the state variables are included in a measurement signal.

that an analysis of colinearities between the columns of the sensitivity matrix $S(t_0, \ldots, t_N, \bar{\theta})$ does not only detect a lack of observability. It essentially also gives a lead into finding a reduced model that has a smaller number of differential equations than the original one. Current ongoing research has already shown that substantial reduction can be achieved for a large model that yields a better insight into the most important state variables that are dominant in the model prediction. Furthermore, a similar analysis on the controllability of the given model also yields a better insight into the question of which state variables are influenced most by a giving input signal (e.g., a harvest rate). This, obviously, is a very important aspect of the management of ecosystems in general.

Finally, we note that much more complex models than the ones in this paper (that were mainly included for didactic reasons) can be analyzed with the StrucID App. The results of a larger model with 26 system parameters and 25 state variables are presented in Figure 5. This is the JAKSTAT-Bachmann model that simulates a signaling pathway in a cell (Bachman et al., 2011) and is well known in the field of Systems Biology. In a recent paper this model served as a benchmark problem for an identifiability test and 13 software packages were compared on this particular problem. Results of the comparison showed that seven packages were not even capable to compute a result because of the model's complexity (Barreiro & Villaverde, 2023). From the remaining six packages, the fastest performance was 35.7 s, while the worst performance was 40.7 h. The StrucID package (that was not included in the benchmark paper) was capable of computing the results presented in Figure 5 in approximately 2 s.

The StrucID App is available from the author upon request. Both a stand-alone version and a Matlab version of this software package are available. For the stand-alone version the only thing needed is a Windows PC. Matlab versions supported are the ones after version 2021a.

AUTHOR CONTRIBUTIONS

Johannes Daniel Stigter: Software; formal analysis; methodology; writing—original draft.

ACKNOWLEDGMENTS

The author wishes to thank the organizers of the 2022 World Conference on Natural Resource Modeling in Leipzig for the opportunity to present the results in this paper.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

ORCID

J. D. Stigter http://orcid.org/0000-0001-5210-5929

ENDNOTE

[1] Of course, the same for the matrices $\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial \theta}$ and the parametric output sensitivities $y_\theta(t)$ in the measurement equation (2).

REFERENCES

Bachman, J., Raue, A., Schilling, M., Böhm, M. E., Kreutz, C., Kaschek, D., Busch, H., Gretz, N., Lehmann, W. D., Timmer, J., & Klingmüller, U. (2011). Division of labor by dual feedback regulators controls jak2/stat5 signaling over broad ligand range. *Molecular Systems Biology*, *7*, 516. https://doi.org/10.1038/msb.2011.50

Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–115.

Barreiro, X., & Villaverde, A. (2023). Benchmarinking tools for a priori identifiability analysis. *Bioinformatics*, *39*(2), btad065. https://doi.org/10.1093/bioinformatics/btad065

Chappell, M., & Godfrey, K. (1992). Structural identifiability of the parameters of a nonlinear batch reactor model. *Mathematical Biosciences*, *108*, 23.

Clark, C. W. (1976). *Mathematical bioeconomics: The optimal management of renewable resource*. John Wiley.

Isidori, A. (1989). *Nonlinear control systems*. Springer-Verlag.

Joubert, D., Stigter, J., & Molenaar, J. (2020). An efficient procedure to assist in the re-parametrization of structually unidentifiable models. *Mathematical Biosciences*, *323*, 108328. https://doi.org/10.1016/j.mbs.2020.108328

Kailath, T. (1980). *Linear systems*. Prentice Hall.

Kalman, R. E., Falb, P. L., & Arbib, M. A. (1969). *Topics in mathematical system theory*. McGraw-Hill.

Kwatny, H. G., & Blankenship, G. L. (2000). *Nonlinear control and analytical mechanics: A computational approach*. Birkhäuser.

Liu, Y., Slotine, J., & Barabási, A. (2012). Observability of complex systems. *Proceedings of the National Academy of Sciences*, *110*(7), 2460–2465.

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

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

How to cite this article: Stigter, J. D. (2023). Computing parameter identifiability and other structural properties for natural resource models. *Natural Resource Modeling*, e12382. https://doi.org/10.1111/nrm.12382

APPENDIX A: EXAMPLE INPUT TEXT FILE FOR THE StrucID APP

```
Algebraic Rules!
   v1 = x*y % optional for simplification
   ODEs (define the individual ODE equations - 1 per line)!
   dx/dt = a*x - b*v1
   dy/dt = d*v1 - c*y
   Input variables (that can be manipulated)!
   Measured Outputs (define the measured sensors - 1 per line)!
   y1 = y
   Parameter names and values (system parameters - 1 per line,
OPTIONAL - define known parameter values)!
   a = 0.3
   b = 0.5
   c = 0.05
   d = 0.1
   State names and initial values (1 per line, OPTIONAL - define known
initial values)!
   x =
   y =
   Analyze (all unknowns to be included in the analysis)!
   % if empty all defined states/parameters will be included
```