

# Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods

Oana-Teodora Chis, Julio R. Banga, Eva Balsa-Canto\*

Bioprocess Engineering Group, IIM-CSIC, Vigo, Spain

## Abstract

Analysing the properties of a biological system through *in silico* experimentation requires a satisfactory mathematical representation of the system including accurate values of the model parameters. Fortunately, modern experimental techniques allow obtaining time-series data of appropriate quality which may then be used to estimate unknown parameters. However, in many cases, a subset of those parameters may not be uniquely estimated, independently of the experimental data available or the numerical techniques used for estimation. This lack of identifiability is related to the structure of the model, i.e. the system dynamics plus the observation function. Despite the interest in knowing *a priori* whether there is any chance of uniquely estimating all model unknown parameters, the structural identifiability analysis for general non-linear dynamic models is still an open question. There is no method amenable to every model, thus at some point we have to face the selection of one of the possibilities. This work presents a critical comparison of the currently available techniques. To this end, we perform the structural identifiability analysis of a collection of biological models. The results reveal that the generating series approach, in combination with identifiability *tableaus*, offers the most advantageous compromise among range of applicability, computational complexity and information provided.

**Citation:** Chis O-T, Banga JR, Balsa-Canto E (2011) Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. PLoS ONE 6(11): e27755. doi:10.1371/journal.pone.0027755

**Editor:** Johannes Jaeger, Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra, Spain

**Received:** April 14, 2011; **Accepted:** October 24, 2011; **Published:** November 22, 2011

**Copyright:** © 2011 Chis et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was financially supported by the Spanish government, MICINN project "MultiSysBio" (ref. DPI2008-06880-C03-02), by Xunta de Galicia project "IDECOP" (ref. 08DPI007402PR) and by CSIC intramural project "BioREDES" (ref. PIE-201170E018). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: ebalsa@iim.csic.es

## Introduction

Modelling and simulation offer the possibility of integrating information, performing *in silico* experiments, generating predictions and novel hypotheses so as to better understand complex biological systems. However, the quality of the results will highly depend on the predictive capabilities of the model at hand. In this regard, the selection of an adequate modelling framework for the system under consideration and for the questions to be addressed is crucial [1] together with the capacity to anchor model sophistication with experimental data [2]. In this respect, parameter estimation by means of data fitting has become a critical step in the model building process [3].

However, and despite the ever increasing availability and quality of biological data, this parameter estimation step still remains a difficult mathematical and computational problem.

It has been argued that such difficulties are often originated in the lack of identifiability, i.e. in the difficulty or (in some cases) impossibility of assigning unique values for the unknown parameters. This has been in fact the case in many examples found in the literature [4–8]. These works report the impossibility to assess unique and meaningful values for the parameters since broad ranges of parameter values result in similar model predictions.

But what is the exact origin of the lack of identifiability? We can distinguish between structural and practical identifiability. Structural identifiability is a theoretical property of the model structure depending only on the system dynamics, the observation and the

stimuli functions [9]. Practical identifiability is intimately related to the experimental data and the experimental noise.

Although the questions seem rather similar, there are several crucial differences. Possibly the most important has to do with the capability to recover identifiability. If some parameters turn out not to be structurally identifiable, numerical approaches will not be able to find reliable values for them. In those situations, the only possibilities for a successful model building will be i) to reformulate the model (reducing the number of states and parameters), ii) to fix some parameter values (for example, those which are less relevant to model predictions) or iii) to design new experiments by adding measured quantities (if technically possible). Lack of practical identifiability will be in general terms solvable, providing the experimental constraints allow designing sufficiently rich experiments. In this regard, recent works suggest the use of model based (optimal) experimental design to iteratively improve the quality of parameter estimates [10–13].

There are, at least, two reasons to assess identifiability. First, most of the model parameters have a biological meaning, and we are interested in knowing whether it is at all possible to determine their values from experimental data. Second, numerical optimisation approaches will find difficulties when trying to estimate the parameters of a non-identifiable model.

In this regard, practical identifiability analysis has received substantial attention in the recent literature. Local analyses are based on the computation of local sensitivities, the Fisher Information Matrix, the covariance matrix, or the Hessian of the least-squares function [14,15]. Hengl et al. [16] proposed the

method of mean optimal transformations to reduce the number of model parameters to improve practical identifiability. Balsa-Canto et al. [10] suggested the use of a bootstrap based approach so as to quantify practical identifiability in terms of eccentricity and pseudo-volume of the robust confidence hyper-ellipsoid. In a more recent work, the same authors suggested the use of the global rank of parameters to assess the relative influence of the parameters in the observables and to anticipate lack of structural or practical identifiability [17].

Despite the importance of knowing *a priori* whether there is any chance of uniquely estimating all model unknowns, the structural identifiability analysis has been ignored in the vast majority of modelling studies in systems biology. Only recently some works have considered the structural identifiability analysis of cell signalling related examples. Balsa-Canto et al. [17] proposed the use of power series based approaches combined with identifiability *tableaus* so as to assess the identifiability of the model of the NFκB module by Lipniacki et al. [4]; Roper et al. [18] considered the analysis of different alternative models of a single phosphorylation-dephosphorylation cycle in the MAPK cascade [19], by means of a differential algebra based approach.

However, the structural identifiability analysis for general non-linear dynamic models in systems biology is still a challenging question. Even though a number of methods exist [20], there is no method amenable to every model, thus at some point we have to face the selection of one of the possibilities.

This work presents a critical comparison of currently available methods so as to evaluate their potential in systems biology. In particular, we will consider the Taylor series method [21], the generating series method [22], both complemented with the identifiability *tableaus* [17], the similarity transformation approach [23], the differential algebra based method [24,25], the direct test method [26,27], a method based on the implicit function theorem [28] and the recently developed test for reaction networks [29–31].

The advantages and disadvantages of all these methods are evaluated on the basis of a collection of examples of increasing size and complexity. The selected models include different types of non-linear terms, such as generalised mass action (GMA), Michaelis-Menten and Hill kinetics, as typically found in systems biology models. The six different examples considered are: the Goodwin oscillator model [32], a pharmacokinetics model that describes the receptor-mediated uptake of glucose oxidase [33], the model of a glycolysis inspired metabolic pathway [34], a high dimensional non-linear model which represents biochemical reaction systems [35], the model of the central clock of *Arabidopsis Thaliana* [36] and the model of the NFκB signalling module [4].

## Methods

### Mathematical model formulation

We will assume a biological system described by:

$$\Sigma(\mathbf{p}) : \begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}) + \sum_{j=1}^{n_u} \mathbf{g}_j(\mathbf{x}, \mathbf{p}) \mathbf{u}_j, \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{p}), \mathbf{x}(t_0) = \mathbf{x}_0(\mathbf{p}) \end{cases} \quad (1)$$

where  $\mathbf{x} = (x_1, \dots, x_{n_x}) \in \mathbf{M} \subset \mathbf{R}^{n_x}$  is the state variable, with  $\mathbf{M}$  a subset of  $\mathbf{R}^{n_x}$  containing the initial state,  $\mathbf{u} = (u_1, \dots, u_{n_u}) \in \mathbf{R}^{n_u}$  a  $n_u$ -dimensional input (control) vector with  $u_1, \dots, u_{n_u}$  smooth functions, and  $\mathbf{y} = (y_1, \dots, y_{n_y}) \in \mathbf{R}^{n_y}$  is the  $n_y$ -dimensional output (experimentally observed quantities). The vector of unknown parameters is denoted by  $\mathbf{p} = (p_1, \dots, p_{n_p}) \in \mathbf{P}$ , and in general is

assumed to belong to an open and connected subset of  $\mathbf{R}^{n_p}$ . The entries of  $\mathbf{f}$ ,  $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_{n_u})$  and  $\mathbf{h}$  are analytic functions of their arguments. These functions and the initial conditions may depend on the parameter vector  $\mathbf{p} \in \mathbf{P}$ .

It should be noted that typical models in systems biology, such as GMA models or those incorporating Michaelis-Menten or Hill type kinetics can be easily drawn in the format of Eqn. (1).

### Structural identifiability definition

Structural identifiability regards the possibility of giving unique values to model unknown parameters from the available observables, assuming perfect experimental data (i.e. noise-free and continuous in time) [9].

- A parameter  $p_i$ ,  $i = 1, \dots, n_p$  is *structurally globally (or uniquely) identifiable* if for almost any  $\mathbf{p}^* \in \mathbf{P}$ ,

$$\Sigma(\mathbf{p}) = \Sigma(\mathbf{p}^*) \Rightarrow p_i = p_i^*, \quad (2)$$

- A parameter  $p_i$ ,  $i = 1, \dots, n_p$  is *structurally locally identifiable* if for almost any  $\mathbf{p}^* \in \mathbf{P}$ , there exists a neighbourhood  $\mathbf{V}(\mathbf{p}^*)$  such that

$$\mathbf{p} \in \mathbf{V}(\mathbf{p}^*) \text{ and } \Sigma(\mathbf{p}) = \Sigma(\mathbf{p}^*) \Rightarrow p_i = p_i^*, \quad (3)$$

- A parameter  $p_i$ ,  $i = 1, \dots, n_p$  is *structurally non-identifiable* if for almost any  $\mathbf{p}^* \in \mathbf{P}$ , there exists no neighbourhood  $\mathbf{V}(\mathbf{p}^*)$  such that

$$\mathbf{p} \in \mathbf{V}(\mathbf{p}^*) \text{ and } \Sigma(\mathbf{p}) = \Sigma(\mathbf{p}^*) \Rightarrow p_i = p_i^*. \quad (4)$$

A vector  $s(\mathbf{p})$  is an *exhaustive summary* of the experiment if it contains only the information about the parameters  $\mathbf{p}$  that can be extracted from knowledge of  $\mathbf{u}(t)$  and  $\mathbf{y}(t, \mathbf{p})$ .

From the previous definitions, structural global ( $\mathbf{p} \in \mathbf{P}$ ) and local ( $\mathbf{p} \in \mathbf{V}(\mathbf{p}^*)$ ) identifiability can be checked by using the exhaustive summary as follows:

$$\mathbf{p}^* \in \mathbf{V}(\mathbf{p}^*) \text{ and } s(\mathbf{p}) = s(\mathbf{p}^*) \Rightarrow \mathbf{p} = \mathbf{p}^*. \quad (5)$$

### Methods for testing structural identifiability

Structural identifiability analysis of linear models is well understood and there are a number of methods to perform such a task. In contrast, there are only a few methods for testing the structural identifiability of non-linear models: the Taylor series method [21], the generating series method [22], the similarity transformation approach [23], the differential algebra based method [24,25], the direct test [26,27], a method based on the implicit function theorem [28] and the recently developed test for reaction networks [29,30].

### Taylor series approach

The Taylor series approach [21] is based on the fact that observations are unique analytic functions of time and so all their derivatives with respect to time should also be unique. It is thus possible to represent the observables by the corresponding Taylor series expansion in the vicinity of the initial state  $t_0$  and the

uniqueness of this representation will guarantee the structural identifiability of the system. The idea is to establish a system of non-linear algebraic equations in the parameters, based on the calculation of the Taylor series coefficients, and to check whether the system has a unique solution.

Let us assume that the state variables  $\mathbf{x} \in \mathbf{M} \subset \mathbf{R}^{n_x}$ , the outputs  $\mathbf{y} \in \mathbf{R}^{n_y}$ , the inputs  $\mathbf{u} \in \mathbf{R}^{n_u}$  and the functions  $\mathbf{f} : \mathbf{M} \rightarrow \mathbf{R}^{n_x}$  and  $\mathbf{g} : \mathbf{M} \rightarrow \mathbf{R}^{n_x} \times \mathbf{R}^{n_u}$  in Eqn. (1) have infinitely many derivatives with respect to time. Let us also assume that  $\mathbf{h} : \mathbf{M} \rightarrow \mathbf{R}^{n_y}$  has infinitely many derivatives with respect to the state vector components and their successive derivatives. The Taylor series expansion of the observation function, in a neighbourhood of the initial state, is then given by

$$y_i(t, \mathbf{p}) = y_i(t_0, \mathbf{p}) + t \dot{y}_i(t_0, \mathbf{p}) + \frac{t^2}{2!} \ddot{y}_i(t_0, \mathbf{p}) + \dots \text{ with } i = 1, \dots, n_y. \quad (6)$$

If we define:

$$a_k^i(\mathbf{p}) := \lim_{t \downarrow t_0^+} \frac{d^k}{dt^k} y_i(t, \mathbf{p}), \quad k = 0, 1, 2, \dots, k_{max}, \quad i = 1, \dots, n_y, \quad (7)$$

then a sufficient condition for global structural identifiability is given by

$$a_k^i(\mathbf{p}) = a_k^i(\mathbf{p}^*), \quad k = 0, 1, 2, \dots, k_{max}, \quad i = 1, \dots, n_y \Rightarrow \mathbf{p} = \mathbf{p}^*, \quad (8)$$

where  $k_{max}$  is the smallest positive integer, such that the symbolic computations give the solution of the parameters.

Possibly the major disadvantage of this method is related to the impossibility to define *a priori* the value of  $k_{max}$ , thus, in general, it will not be possible to talk about a “complete” resolvability for the cases where  $k < k_{max}$ . Some bounds have been established for particular types of models. For example, for a linear model the upper bound on the number of derivatives should be  $2n_x - 1$  [37], for bilinear models,  $2^{2n_x} - 1$  and for homogeneous polynomial systems,  $(s^{2n_x} - 1)/(s - 1)$ , where  $s$  represents the degree of the polynomials [38]. For a single output model, Margaria et al. [39] showed that  $n_x + n_p$  derivatives are sufficient to determine the structural identifiability using the Taylor series method. These bounds could be higher for real problems, particularly when the germ is not informative, i.e. when the Taylor coefficients become zero at the initial conditions.

Another important disadvantage of this method is that the usual complexity of the resulting algebraic parametric relations makes the analysis difficult, allowing, in many cases, only for local identifiability results [40]. This is particularly true when the number of required derivatives is large. This explains why, despite its conceptual simplicity and that computations may be simplified when the initial conditions are known, this approach has not become popular in practice [41].

### Generating series approach

Conceptually similar to the Taylor method, in the generating series approach [22] the observables can be expanded in series with respect to time and inputs in such a way that the coefficients of this series are the output functions  $\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$ , and their successive Lie derivatives along the vector fields  $\mathbf{f}$  and  $\mathbf{g}$  ( $L_{\mathbf{f}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$ ,  $L_{\mathbf{g}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$ ,  $L_{\mathbf{f}}L_{\mathbf{f}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$ ,  $L_{\mathbf{f}}L_{\mathbf{g}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$ ,  $L_{\mathbf{g}}L_{\mathbf{f}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$ ,  $L_{\mathbf{g}}L_{\mathbf{g}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t_0)$  and so on).

The Lie derivative of  $\mathbf{h}$  along the vector field  $\mathbf{f}$ , is given by:

$$L_{\mathbf{f}}\mathbf{h}(\mathbf{x}, \mathbf{p}, t) = \sum_{i=1}^{n_x} \mathbf{f}_i(\mathbf{x}, \mathbf{p}, t) \frac{\partial \mathbf{h}(\mathbf{x}, \mathbf{p}, t)}{\partial x_i} \quad (9)$$

with  $\mathbf{f}_i$  the  $i^{th}$  component of  $\mathbf{f}$ , where  $i, i = 1, \dots, n_x$ .

The exhaustive summary contains the coefficients of  $\mathbf{h}(\mathbf{x}_0(\mathbf{p}))$ , and the successive Lie derivatives along  $\mathbf{g}$  and/or  $\mathbf{f}$ , evaluated at the initial conditions  $\mathbf{x}_0(\mathbf{p})$ . The model (1) is structurally globally identifiable if the exhaustive summary is unique.

As in the case of the Taylor approach, the major disadvantage of the generating series approach is that the minimum number of required Lie derivatives is unknown. The lack of such a bound offers only sufficient, but not necessary, conditions for identifiability. The advantage is that the mathematical expressions obtained with the generating series method are usually simpler than those obtained with the Taylor series approach [42].

It should be remarked at this point that both power series based methods may be applied to arbitrary non-linear functions  $\mathbf{f}$ ,  $\mathbf{g}$  and  $\mathbf{h}$  in the model (1), thus being excellent candidates to perform the analysis for the models in systems biology. However, the solution of the resultant set of non-linear algebraic equations in the parameters may be challenging (or impossible) even with the aid of symbolic manipulation software. In this concern, the systematic computation of so called identifiability *tableaus* [17] is introduced here as a way to easily visualise the possible structural identifiability problems and to systematise the solution of the resulting algebraic system of equations on the parameters.

### Identifiability *tableaus*

The *tableau* represents the non-zero elements of the Jacobian of the series coefficients with respect to the parameters. It consists of a table with as many columns as parameters and with as many rows as non-zero series coefficients (in principle, infinite).

If the Jacobian is rank deficient, i.e. the *tableau* presents empty columns, the corresponding parameters may be non-identifiable. Note that since the number of series coefficients may be infinite, structural non-identifiability may not be fully guaranteed unless higher order series coefficients are demonstrated to be zero.

If the rank of the Jacobian coincides with the number of parameters, then it will be possible to, at least, locally identify the parameters. In this situation a careful inspection of the *tableau* will help to decide on an iterative procedure for solving the system of equations, as follows:

- The number of non-zero coefficients is usually much larger than the number of parameters. In practice this means that we should select the first  $n_p$  rows that guarantee the Jacobian rank condition. The *tableau* helps to easily detect the necessary coefficients and to generate a “minimum” *tableau*.
- A unique non-zero element in a given row of the minimum *tableau* means that the corresponding parameter is structurally identifiable. If the parameters in this situation can be computed as functions of the power series coefficients, they can be then eliminated from the “minimum” *tableau* to generate a “reduced” *tableau*. Subsequent reductions may lead to the appearance of new unique non-zero elements, and so on. Thus, all possible “reduced” *tableaus* should be built in sequence first.
- Once no more reductions are possible, one should try to solve the remaining equations. Since it is often the case that not all remaining power series coefficients depend on all parameters,

the *tableau* will help to decide on how to select the equations to solve for particular parameters.

- If several meaningful solutions exist for a given set of parameters, then the model is said to be structurally locally identifiable.

### Similarity transformation approach

The similarity transformation approach [23] is based on the local state isomorphism theorem. The model should be locally reduced, i.e. controllability and observability conditions must be fulfilled at  $t_0$  and it is assumed that the entire class of bounded and measurable functions is available for stimulus. The method seeks state variable transformations that leave invariant the stimuli-observables map and the structure of the system.

The local state isomorphism is used to establish a set of first order linear inhomogeneous partial differential equations which is used to construct the functional form of such transformations. Unfortunately, the solution of the partial differential equations may be complex, and the need to test controllability and observability conditions poses additional problems to the application of this methodology for general non-linear systems.

An alternative was proposed by Denis-Vidal and Joly-Blanchard [43] that allows to obtain direct relations of the components of the isomorphism.

The identifiability of the parameters of the model (1) can be obtained by using the local state isomorphism theorem as follows:

**Theorem 1.** [40] Let us consider the parameter values  $\mathbf{p}, \mathbf{p}^* \in \mathbf{p}$  such that the model (1) is locally reduced at the initial states  $\mathbf{x}_0(\mathbf{p})$ , respectively  $\mathbf{x}_0(\mathbf{p}^*)$  (observability and controllability rank conditions are satisfied at  $\mathbf{x}_0(\mathbf{p})$ , respectively  $\mathbf{x}_0(\mathbf{p}^*)$ ),  $V \subset \mathbf{R}^{n_x}$  is an open neighbourhood of  $\mathbf{x}_0(\mathbf{p})$ , and there exists an analytical mapping  $\lambda : V \rightarrow \mathbf{R}^{n_x}$  with the following properties:

(i)

$$\text{rank} \frac{\partial \lambda(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}^*} = n_x, \tag{10}$$

(ii)

$$\lambda(\mathbf{x}_0(\mathbf{p}^*)) = \mathbf{x}_0(\mathbf{p}), \tag{11}$$

(iii)

$$\mathbf{f}(\lambda(\mathbf{x}^*), \mathbf{p}) = \frac{\partial \lambda(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}^*} \mathbf{f}(\mathbf{x}^*, \mathbf{p}^*), \tag{12}$$

$$\mathbf{g}(\lambda(\mathbf{x}^*), \mathbf{p}) = \frac{\partial \lambda(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}^*} \mathbf{g}(\mathbf{x}^*, \mathbf{p}^*), \tag{13}$$

$$\mathbf{h}(\lambda(\mathbf{x}^*), \mathbf{p}) = \mathbf{h}(\mathbf{x}^*, \mathbf{p}^*), \tag{14}$$

for all  $\mathbf{x}^* \in V$ . Then (1) is globally identifiable at  $\mathbf{p}$  if and only if conditions (10)–(14) imply  $\mathbf{p} = \mathbf{p}^*$ .

The claim of [44] is that the local state isomorphism between two state space systems corresponding to  $\mathbf{p}$  and  $\mathbf{p}^*$  must be linear. This restriction comes from the assumption that the observability rank condition must be satisfied. Further details may be found in the recent work by Peeters and Hanzon [45]. Note that Denis-Vidal and Joly-Blanchard [43] eliminate the assumption of linearity.

The major disadvantages of this method are related to the difficulty of assessing the observability condition and the complexity to solve the differential equations (12) for general non-linear dynamic systems. Even the modifications proposed by Denis-Vidal and Joly-Blanchard [43] may not be enough for large scale highly non-linear models.

### Direct test

The conceptually simplest approach to test structural identifiability is the so called *direct test* [46], applicable to uncontrolled and autonomous systems.

This method consists basically on trying to solve directly the equality  $\mathbf{f}(\mathbf{p}) = \mathbf{f}(\mathbf{p}^*) \Rightarrow \mathbf{p} = \mathbf{p}^*$ , for getting local or global identifiability of the generic model (1). In general, reaching a conclusion may require excessively complicated formal manipulations or the equations to be solved may be too complicated for an analytic expression to exist, which then imposes the use of numerical methods, thus loosing the formal nature of the solution.

### Differential algebra approach

The differential algebra methods [24] are based on replacing the stimuli-observables behaviour of the system by some polynomial or rational mapping. Non-observable differential state variables are eliminated in order to get differential relations among inputs, outputs and parameters, that result from these differential relations, using Ollivier’s method [47]. The exhaustive summary can be obtained and solved using algebraic methods, such as the Buchberger algorithm [48]. The algorithm is rigorous, as it converges in a finite number of steps [24].

Different strategies using the differential algebra approach have been proposed for models described by linear/non-linear differential equations, in terms of polynomial or rational functions, with or without known initial conditions.

Let us consider the general model given by (1), with  $\mathbf{f} : \mathbf{M} \rightarrow \mathbf{R}^{n_x}$ ,  $\mathbf{g} : \mathbf{M} \rightarrow \mathbf{R}^{n_x} \times \mathbf{R}^{n_u}$ ,  $\mathbf{h} : \mathbf{M} \rightarrow \mathbf{R}^{n_y}$  polynomial or rational functions of their arguments and the  $n_u$  – dimensional differentiable input  $\mathbf{u}$ . The second assumption is that the system is accessible from its initial conditions (equivalent to a “generic controllability”) [25]. The model  $\Sigma(\mathbf{p})$  can be written as differential polynomials

$$\Sigma'(\mathbf{p}) : \begin{cases} \mathbf{x} - \mathbf{f}(\mathbf{x}, \mathbf{p}) - \sum_{j=1}^{n_u} \mathbf{g}_j(\mathbf{x}, \mathbf{p}) \mathbf{u}_j, \\ \mathbf{y} - \mathbf{h}(\mathbf{x}, \mathbf{p}). \end{cases} \tag{15}$$

Rational systems of differential equations are reduced to the same denominator, or to a pure polynomial form.

The differential algebra approach proceeds as follows:

- $\Sigma'(\mathbf{p})$  represents the set of *differential polynomials* denoted by  $\mathbf{F}(\mathbf{u}, \mathbf{y}, \mathbf{x}, t)$ .
- The *differential polynomial ring*  $(\mathbf{R}[\mathbf{u}, \mathbf{y}, \mathbf{x}])$  is made of polynomials of the indeterminate variables  $x_1, \dots, x_{n_x}$  and their derivatives, the inputs  $u_1, \dots, u_{n_u}$  and outputs  $y_1, \dots, y_{n_y}$  and their derivatives.
- $\mathcal{I} \subset \mathbf{R}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$  is the ideal generated by the polynomials  $\mathbf{F}(\mathbf{u}, \mathbf{y}, \mathbf{x}, t)$  and consists of all differential polynomials that



can be obtained by using addition, multiplication and differentiation. A differential ideal is called *prime* if  $P_1 P_2 \in \mathcal{I} \Rightarrow (P_1 \in \mathcal{I} \text{ or } P_2 \in \mathcal{I})$ .

- The differential ideal is represented by a finite basis computed by applying a set “ordering” of the variables and their derivatives, called *ranking*. In literature, the ranking is given by the inputs, as lowest ranked, outputs, and the highest rank is attributed to the state variables [24]:

$$\begin{aligned} u_1 < \dots < u_{n_u} < \dot{u}_1 < \dots < \dot{u}_{n_u} < \ddot{u}_1 < \dots < y_1 < \dots < y_{n_y} < \\ \dot{y}_1 < \dots < \dot{y}_{n_y} < \ddot{y}_1 < \dots < x_1 < \dots < x_{n_x} < \dot{x}_1 < \dots < \dot{x}_{n_x} < \ddot{x}_1 < \dots \end{aligned} \quad (16)$$

The *leader* of a polynomial is the highest ranking derivative of the polynomial, and the corresponding variable is called *leading variable* [24]. The results usually change if the ranking is changed. So, we can say that differential algebra methods are rank dependent. This ranking is used to obtain an observable representation of the model, by eliminating the unmeasured state variables.

- Ritt’s algorithm [49] computes the characteristic set, using the set of differential polynomials and differential ideals. With the ranking (16), the differential ideal has the characteristic set made of differential polynomials of the form

$$\begin{aligned} A_1(\mathbf{u}, \mathbf{y}), \dots, A_{n_y}(\mathbf{u}, \mathbf{y}), \\ A_{n_y+1}(\mathbf{u}, \mathbf{y}, x_1), \dots, A_{n_y+n_x}(\mathbf{u}, \mathbf{y}, x_1, \dots, x_{n_x}), \end{aligned} \quad (17)$$

where  $A_1, \dots, A_{n_y}, \dots, A_{n_y+n_x}$  are differential polynomials, with the leaders of  $A_i, i=1, \dots, n_y$  the derivatives of  $y_i$ . The relations (17) represent the characteristic set associated to the generic model (1) [24,27]. The characteristic set may also be computed using the (improved) Ritt-Kolchin algorithm [50] or Rosenfeld-Gröbner algorithm [26]. All these algorithms eliminate the highest ranking variable, such that differential polynomials in  $\mathbf{u}, \mathbf{y}, \mathbf{p}$  are obtained using symbolic computations. The eliminating process is called *pseudo-division*.

- Normalising the differential polynomial in  $\mathbf{u}, \mathbf{y}, \mathbf{p}$  the exhaustive summary of the model is obtained. It is made of the coefficients  $c_i$  of each polynomial  $A_i(\mathbf{u}, \mathbf{y}), i=1, \dots, n_y$ , denoted by  $c: \mathbf{P} \rightarrow \mathbf{R}^l, l=l_1 + \dots + l_{n_y}$ , defined by  $c(\mathbf{p}) = c_{ij}(\mathbf{p}), i=1, \dots, n_y, j=1, \dots, l_i$ , where  $l_i$  is the number of coefficients in each  $A_i(\mathbf{u}, \mathbf{y}), i=1, \dots, n_y$ . The structural identifiability is equivalent to checking the injectivity of the map  $c: \mathbf{P} \rightarrow \mathbf{R}^l$ . This is equivalent to solving the system of equations  $c_{ij}(\mathbf{p}) = c_{ij}(\mathbf{p}^*), i=1, \dots, n_y, j=1, \dots, l_i$  [39]. In this concern, algorithms based on the Gröbner basis may give information about the nature of the solution. Note that, in some occasions solving that system of non-linear algebraic equations may be complicated, if not impossible; for these situations it is possible to use pseudo-randomly generated numerical values instead of symbolic  $c_{ij}(\mathbf{p}^*)$  [25].

The advantage of these differential algebraic methods is that the solution of the associated algebraic equations gives precise information about the identifiability or non-identifiability of the parameters, but the disadvantage is the great computational requirements when a complex model is considered.

### Implicit Function Theorem

Proposed by Xia and Moog [28], this method is based on computing the derivatives of the observables with respect to independent variables (time) to eliminate unobserved states. A differential system is obtained, depending only on known system

inputs, observable outputs and unknown parameters [41]. An identification matrix is defined, consisting of the partial derivatives of the differential equations with respect to unknown parameters. If the identification matrix is not singular, the system is said identifiable. The identifiability theory is based on the following theorem:

**Theorem 2.** [28] Let  $\Psi: \mathbf{R}^{n_y+n_u+n_p} \rightarrow \mathbf{R}^{n_p}$  denote the function of model parameter  $\mathbf{p} \in \mathbf{R}^{n_p}$ , system input  $\mathbf{u} \in \mathbf{R}^{n_u}$ , system output  $\mathbf{y} \in \mathbf{R}^{n_y}$ , and their derivatives:

$$\Psi = \Psi(\mathbf{p}, \mathbf{u}, \mathbf{u}, \dots, \mathbf{u}^{(k)}, \mathbf{y}, \mathbf{y}, \dots, \mathbf{y}^{(k)}),$$

where  $k$  is a non-negative integer. Assume that  $\Psi$  has continuous partial derivatives with respect to  $\mathbf{p}$ . Then the generic model (1) is locally identifiable at  $\mathbf{p}^* \in \mathbf{P}$  if there exists a point

$$(\mathbf{p}^*, \mathbf{u}^*, \mathbf{u}^*, \dots, \mathbf{u}^{*(k)}, \mathbf{y}^*, \mathbf{y}^*, \dots, \mathbf{y}^{*(k)}) \in \mathbf{R}^{n_p+n_u+n_y}$$

such that

$$\Psi(\mathbf{p}^*, \mathbf{u}^*, \mathbf{u}^*, \dots, \mathbf{u}^{*(k)}, \mathbf{y}^*, \mathbf{y}^*, \dots, \mathbf{y}^{*(k)}) = 0, \det\left(\frac{\partial \Psi}{\partial \mathbf{p}^*}\right) \neq 0. \quad (18)$$

The relations in (18) are equivalent to checking structural identifiability, by examining differential polynomials  $\Psi$  in the characteristic set, that can give us information if the model is identifiable or not, and which parameters are identifiable/non-identifiable.

This method becomes more and more complicated as the number of parameters increases due to the complexity of deriving the matrix  $\frac{\partial \Psi}{\partial \mathbf{p}^*}$ . Wu et al. [41] proposed an alternative, *the multiple time points method*, that may be helpful for large scale systems. This method relies on the computation of the derivatives at a number of sampling times  $t_1, \dots, t_j$ . Note however, that this requires preliminary information about the observables at those sampling times.

This method offers the possibility of detecting the minimum number of observables needed to compute all parameters [28], as the computations may be performed independently for each observable.

### Identifiability analysis for dynamic reaction networks

For the case of chemical reaction networks written as in the chemical reaction network theory (CRNT) [29,30] the structural identifiability may be checked in two steps [30]: the *reaction rate identifiability* and the *structural rate identifiability*.

The idea is to determine the structurally identifiable reaction rates, using the stoichiometric matrix, and then parameter identifiability may be computed for the considered reaction rates, using one of the above mentioned methods. In their work, Davidescu and Jorgensen make use of the generating series approach.

We consider the following facts and notations, as presented in [51]:

- $N \in \mathbf{R}^{n_R \times n_S}$ , with  $n_R$ , the number of reactions and  $n_S$  the number of species, regards the stoichiometric matrix.
- $N_m \in \mathbf{R}^{n_R \times n_{S_m}}, N_{um} \in \mathbf{R}^{n_R \times n_{S_u}}$ , where the index  $m$  stands for measured chemical species and  $um$  for unmeasured ones, regard the stoichiometric sub-matrix corresponding to the observed species and the stoichiometric sub-matrix corresponding to the unobserved species, respectively;

- if  $\text{rank}(N_m) = n_R$ , then all reactions are identifiable;
- if  $\text{rank}(N_m) < n_R$ , an identifiability criterion was introduced by [51], based on the difference between  $N_m N_m^+$  and  $I$ , where  $N_m^+ = N_m^T (N_m N_m^T)^{-1}$  is the Moore Penrose inverse, and  $I$  is the identity matrix.

A reaction rate is called *structurally identifiable* if the corresponding column in the matrix

$$[N_m N_m^+ - I] \quad (19)$$

is represented by the null vector [30].

## Implementation of methods

To the authors knowledge, currently there are only two software tools available that can be used for structural identifiability analysis of non-linear models: DAISY [25] and the recently developed GenSSI toolbox [52].

DAISY implements the differential algebra based approach by using REDUCE. In principle, it is suited for any non-linear dynamic system with known numeric or symbolic non-rational initial conditions. It offers the advantage that non-expert users may perform the structural identifiability analysis even for rational models that be automatically transformed into polynomial forms. The major disadvantage is that no intermediate results may be obtained, i.e. unless the computation is completed no results will be displayed.

To surmount this difficulty, we made an implementation of the method by using the *Epsilon*, *linalg* and *Gröbner* packages, available in MAPLE, for calculations of Gröbner bases and related operations for ideals in polynomial rings. The computation of the characteristic sets has the disadvantage that one should have knowledge about the implementation and theory, and the algorithm needs to be adapted by hand, for example for rational models.

GenSSI implements the combination of the generating series approach with the identifiability *tableaus* [17]. It is also suited for non-linear dynamic models provided they are linear in the control variables (as in Eqn. (1)). It offers several advantages to non-expert users such as the possibility of handling any type of non-linear terms with transforming the models to polynomial form and the possibility of automatically incorporating known symbolic or numeric initial conditions. In addition, intermediate results on the structural identifiability of a sub-set of parameters are provided throughout the process.

The rest of the methods considered here were implemented by using suitable packages available in symbolic manipulation software tools, such as MATHEMATICA, MAPLE and MATLAB.

## Results

As mentioned before, there is no single method amenable to all types of problems for testing structural identifiability. In order to perform a critical comparison of the different possibilities in the context of systems biology, we have considered the structural identifiability analysis of the following models: the Goodwin oscillator model [32], a pharmacokinetics model that describes the receptor-mediated uptake of glucose oxidase [33], the model of a glycolysis inspired metabolic pathway [34], a high dimensional non-linear model which represents biochemical reaction systems [35], the model of the central clock of *Arabidopsis Thaliana* [36] and the model of the NFκB signalling module [4].

## Case study 1: Goodwin's model

The model describes the oscillations in enzyme kinetics [53]. The state variable  $x_1$  represents an enzyme concentration whose rate of synthesis is regulated by feedback control via a metabolite  $x_3$ , and  $x_2$  regulates the synthesis of  $x_3$ . It is characterised by a rational kinetics consisting of a Hill-like term, and it is given by:

$$\begin{cases} \dot{x}_1 = -bx_1 + \frac{a}{A+x_3^\sigma}, \\ \dot{x}_2 = \alpha x_1 - \beta x_2, \\ \dot{x}_3 = \gamma x_2 - \delta x_3, \\ x_1(0) = 0.3617, x_2(0) = 0.9137, x_3(0) = 1.3934 \end{cases} \quad (20)$$

Two scenarios will be considered, (a) the typical case when only  $x_1$  can be measured ( $y_1 = x_1$ ) and (b) a hypothetical situation for which all states can be measured ( $y_1 = x_1, y_2 = x_2, y_3 = x_3$ ).

For the case of one observable, the *power series based methods* (Taylor and generating series) were not able to compute a full rank *tableau*, because only 6 iterative derivatives could be computed. In contrast, for the case of full observation the *power series based methods* ended up in a full rank *tableau* as shown in Figure 1.(a). However, the symbolic manipulation tools were not able to solve the non-linear system of equations on the parameters, so only structural local identifiability may be assessed.

The *similarity transformation approach* could not be applied since the controllability condition is not fulfilled for this system.

The *direct test method* indicated the identifiability of  $\alpha$ , but no information was reported for the remaining parameters due to the complexity of the algebraic manipulations.

The *method based on the implicit function theorem* was only applicable for the case of full observation concluding that the remaining parameters are structurally locally identifiable provided  $\sigma > 2$ .

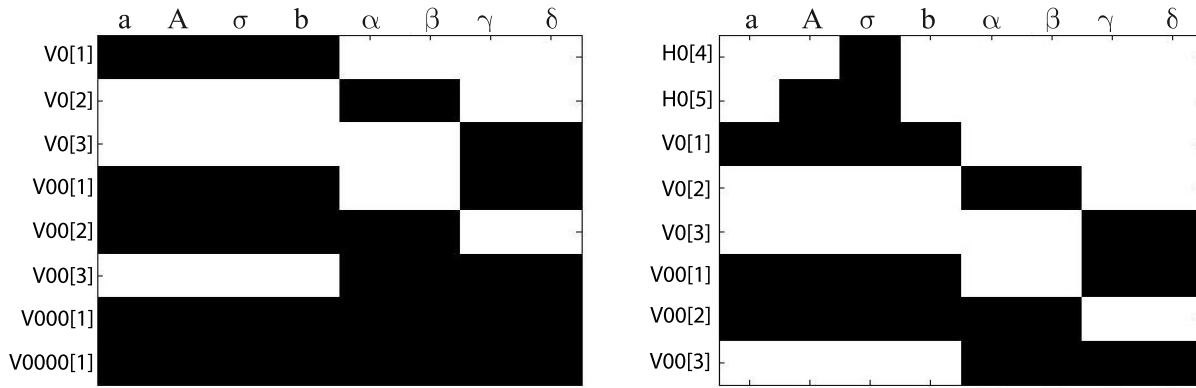
Similarly, to apply the identifiability analysis for dynamic reaction networks we had to fix both  $\sigma$  and  $A$  this allowing to derive the structural local identifiability of the remaining parameters.

The *differential algebra approach*, as implemented in DAISY, results in the non-identifiability of the model when 1 or 3 observables are considered. No results about local identifiability were reported, thus we decided to transform the original model into a full polynomial form of the model, as follows:

$$\begin{cases} \dot{x}_1 = -bx_1 + ax_5, \\ \dot{x}_2 = \alpha x_1 - \beta x_2, \\ \dot{x}_3 = \gamma x_2 - \delta x_3, \\ \dot{x}_4 = \sigma x_4 x_6 (\gamma x_2 - \delta x_3), \\ \dot{x}_5 = -\sigma x_4 x_5^2 x_6 (\gamma x_2 - \delta x_3), \\ \dot{x}_6 = -x_6^2 (\gamma x_2 - \delta x_3), x_1(0) = 0.3617, x_2(0) = 0.9137, \\ x_3(0) = 1.3934, \\ x_4(0) = x_3(0)^\sigma, x_5(0) = \frac{1}{A+x_3(0)^\sigma}, x_6(0) = \frac{1}{x_3(0)}. \end{cases} \quad (21)$$

to check whether further results could be achieved.

Since algebraic operations were much simpler for this model reformulation, the power series based approaches were now able to conclude that the model (21) is structurally globally identifiable



**Figure 1. Goodwin oscillator: Identifiability tableaux.** (a) Identifiability tableau obtained by means of the power series methods for the case of full observation, (b) Identifiability tableau obtained by means of the power series methods for the case of pure polynomial form and full observation.  $H[j]$  and  $V[j]$  regard the different generating series coefficients, H is used for zero order coefficients whereas V correspond to the successive Lie derivatives of  $\mathbf{h}_j$  along  $\mathbf{f}$ , for example,  $V000[j] = L_f L_f L_f h_j, j = 1, \dots, n_y$ . A black square in the coordinates  $(i, k)$  indicates that the corresponding non-zero generating series coefficient  $i$  depends on the parameter  $p_k$ . doi:10.1371/journal.pone.0027755.g001

for all parameters, for the full observation case. However, the DAISY software found the model structurally non-identifiable (initial conditions not used), and was not able to finish the computations reporting errors at the time of introducing the initial conditions.

To sum up, this example illustrates how the structural identifiability analysis may contribute to the design of experiments by providing information on what to be observed so as to guarantee the structural identifiability of a given mathematical model. In addition, results also show how rational terms and Hill coefficients may pose problems to some of the methods and how pure polynomial forms may be useful so as to simplify the analysis.

For illustrative purposes, a detailed explanation of the application of the different methods to this example may be found in Supporting Information S1.

**Case study 2: Pharmacokinetics model**

The pharmacokinetics model [33] is a two compartment model that embodies the ligands of the macrophage mannose receptor, and it is represented mathematically as a system of differential equations of the form:

$$\begin{cases} \dot{x}_1 = \alpha_1(x_2 - x_1) - k_a v_m x_1 / (k_c k_a + k_c x_3 + k_a x_1), \\ \dot{x}_2 = \alpha_2(x_1 - x_2), \\ \dot{x}_3 = \beta_1(x_4 - x_3) - k_c v_m x_3 / (k_c k_a + k_c x_3 + k_a x_1), \\ \dot{x}_4 = \beta_2(x_3 - x_4), \quad x_1(0) = c_0, x_2(0) = 0, x_3(0) = \gamma c_0, x_4(0) = 0, \end{cases} \quad (22)$$

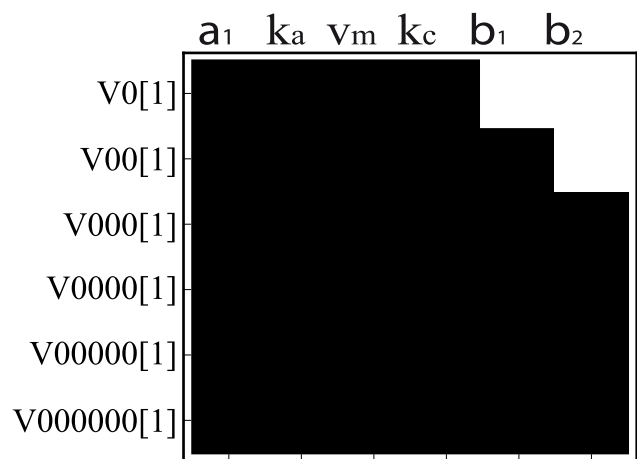
where  $x_1$  represents the enzyme concentration in plasma,  $x_2$  its concentration in compartment 2,  $x_3$  is the plasma concentration of the mannosylated polymer that acts as a competitor of glucose oxidase for the mannose receptor of macrophages, and  $x_4$  is the concentration of the same competitor in the part of the extravascular fluid of the organs accessible to this macromolecule [33]. This example is often used as a benchmark for structural identifiability methods. Two scenarios are considered (a) the case where the measured state corresponds to  $x_1$  ( $y_1 = x_1$ ), (b) the case where “an artificial output”  $y_2 = x_2$  is added [54], to do so  $\alpha_2$  is assumed to be known [33,35].

The model (22) is autonomous and has no control function, so in this case the *Taylor series approach* and *generating series approach* coincide. The corresponding reduced identifiability tableaux are presented in Figure 2. The identifiability tableaux for both scenarios have full rank, thus guaranteeing, at least, structural local identifiability, even for the realistic scenario with one observable.

The introduction of a fictitious control in the model so as to fulfil the controllability condition enabled the application of the *local state isomorphism theorem* to assess local structural identifiability for the case with two observables [55]. However, the presence of a control variable does not correspond to reality, therefore the *similarity transformation approach* can not be directly applied.

The application of the *direct test method* generated two solutions for the parameters. Only for parameter  $\beta_2$  global structural identifiability was confirmed.

Saccomani et al. [35] considered the use of DAISY for the analysis of this model concluding that for the scenario with two observables the six parameters considered are structurally globally identifiable (with known  $\alpha_2$ ). Note however that no results could be obtained for the case with one observable (with unknown  $\alpha_2$ ), generating the computational error “heap space low”.



**Figure 2. Pharmacokinetics model [33].** Identifiability tableau obtained by means of the Taylor/generating series method doi:10.1371/journal.pone.0027755.g002

For the case of the application of the *implicit function theorem* it was possible to obtain the characteristic set independent of the unobserved states. However, manually generating the identifiability Jacobian matrix was too complicated. Therefore, the analysis could not be finished.

In order to apply the method for reaction networks we need to devise the network that gives rise to the model (22). For this particular example a stoichiometric matrix  $N \in \mathbf{R}^{6 \times 4}$  can be obtained, with the matrix of measured states  $N_m$  of rank 4. Final results assess the local identifiability of  $k_a, k_c$  and  $V_m$ . It should be noted that this may be rather complicated since the solution may not be unique [56].

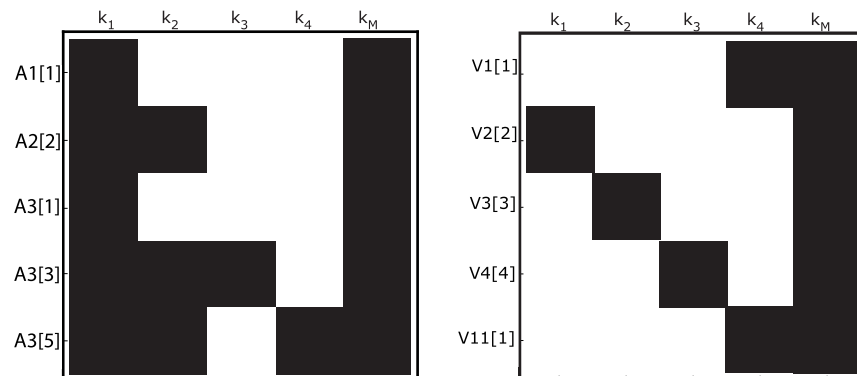
From the results can then be concluded that the model is at least structurally locally identifiable for the realistic case with one observable as reported by the series based methods.

### Case study 3: Glycolysis inspired metabolic pathway

This model represents a glycolysis inspired pathway (the upper part of the glycolysis) with different physiological constraints on enzyme synthesis as described in Bartl et al. [34]. A specific enzyme, here denoted by  $\mathbf{u}$ , usually catalyses a metabolic reaction, expressed in terms of the stoichiometric matrix and the metabolites, here denoted by  $\mathbf{x}$ . The dynamical model can be written as a system of differential equations

$$\begin{cases} \dot{x}_1 = -\frac{k_1 x_1}{x_1 + k_M} u_1, \\ \dot{x}_2 = \frac{k_1 x_1}{x_1 + k_M} u_1 - \frac{k_2 x_2}{x_2 + k_M} u_2, \\ \dot{x}_3 = \frac{k_2 x_2}{x_2 + k_M} u_2 - \frac{k_3 x_3}{x_3 + k_M} u_3, \\ \dot{x}_4 = \frac{k_2 x_2}{x_2 + k_M} u_2 + \frac{k_3 x_3}{x_3 + k_M} u_3 - \frac{k_4 x_4}{x_4 + k_M} u_4, \\ \dot{x}_5 = \frac{k_4 x_4}{x_4 + k_M} u_4. \\ x_1(0) = S_1, x_2(0) = S_2, x_3(0) = S_3, x_4(0) = S_4, x_5(0) = S_5. \end{cases} \quad (23)$$

The model is considered to be fully observed,  $\{y_1 = x_1, y_2 = x_2, y_3 = x_3, y_4 = x_4, y_5 = x_5\}$ , and  $u_1, u_2, u_3, u_4$  independent variables.



**Figure 3. Glycolysis metabolic pathway: Identifiability tableaux.** (a) Identifiability tableau obtained by means of the Taylor series method ( $Ai[j]$ , regards the  $j^{\text{th}}$  component of the  $i^{\text{th}}$  order coefficients of the Taylor series), (b) Identifiability tableau obtained by means of the generating series method.

doi:10.1371/journal.pone.0027755.g003

The *Taylor series approach* produced an identifiability tableau of rank 5 as given in Figure 3.(a). Also, the solutions of the parameters were given: unique solution for  $k_1, k_2$  and  $k_M$ , double solution for  $k_3$  and four solutions for  $k_4$ . However multiple solutions were found and due to their complexity it was impossible to assess their uniqueness for the case of real positive values.

The application of the *generating series approach* indicated the global identifiability of the model. The computational cost was significantly lower as compared to the Taylor series approach. In addition, the identifiability tableau was not as dense, thus the solution of the system of non-linear equations on the parameters was simpler, finally resulting in an unique solution for all parameters.

The *similarity transformation approach* could not be used for this example since the observability condition is not fulfilled. The *direct test method* was also not applicable since the system is autonomous and controlled.

The *method based on the implicit function theorem* could be applied by considering the following 3 relations

$$f_1 = \dot{y}_1 (y_1 + k_M) + k_1 y_1 u_1,$$

$$f_2 = \dot{y}_2 (y_1 + k_M)(y_2 + k_M) - k_1 y_1 (y_2 + k_M) u_1 + k_2 y_2 (y_1 + k_M) u_2,$$

$$f_3 = \dot{y}_3 (y_2 + k_M)(y_3 + k_M) - k_2 y_2 (y_3 + k_M) u_2 + k_3 y_3 (y_2 + k_M) u_3.$$

From the first equation and its derivative, the parameters  $k_1$  and  $k_M$  were found. Using the second one and  $f_2$ , the determinant with respect to  $k_2$  and  $k_3$  was shown to have rank 2, and from the last equation the parameter  $k_4$  could be found. By applying Theorem 2, local identifiability was guaranteed.

Both differential algebra method implementations found the model to be globally identifiable (computation performed without the use of initial conditions).

It should be noted that the metabolic network (23) can be written in terms of stoichiometric matrix and reaction rates. The stoichiometric matrix has rank equal to 5. By choosing one matrix corresponding to the reaction rates 1, 2, 3 and 4, and then the reaction rates 1, 2, 3 and 5, and for each case applying the generating series approach, the identifiability is assessed.



Several methods (the generating series method, differential algebra and the method for reaction networks) were successful in concluding that the model is structurally globally identifiable.

**Case study 4: high dimensional non-linear model [35]**

The system, that could describe a biochemical reaction network, is represented by twenty differential equations, twenty-two parameters, and all the states are assumed to be measured [35]:

$$\left\{ \begin{aligned} \dot{x}_1 &= -v_{max}x_1/(k_m + x_1) - p_1x_1 + u(t), \\ \dot{x}_2 &= p_1x_1 - p_2x_2, \\ \dot{x}_3 &= p_2x_2 - p_3x_3, \\ \dot{x}_4 &= p_3x_3 - p_4x_4, \\ \dot{x}_5 &= p_4x_4 - p_5x_5, \\ \dot{x}_6 &= p_5x_5 - p_6x_6, \\ \dot{x}_7 &= p_6x_6 - p_7x_7, \\ \dot{x}_8 &= p_7x_7 - p_8x_8, \\ \dot{x}_9 &= p_8x_8 - p_9x_9, \\ \dot{x}_{10} &= p_9x_9 - p_{10}x_{10}, \\ \dot{x}_{11} &= p_{10}x_{10} - p_{11}x_{11}, \\ \dot{x}_{12} &= p_{11}x_{11} - p_{12}x_{12}, \\ \dot{x}_{13} &= p_{12}x_{12} - p_{13}x_{13}, \\ \dot{x}_{14} &= p_{13}x_{13} - p_{14}x_{14}, \\ \dot{x}_{15} &= p_{14}x_{14} - p_{15}x_{15}, \\ \dot{x}_{16} &= p_{15}x_{15} - p_{16}x_{16}, \\ \dot{x}_{17} &= p_{16}x_{16} - p_{17}x_{17}, \\ \dot{x}_{18} &= p_{17}x_{17} - p_{18}x_{18}, \\ \dot{x}_{19} &= p_{18}x_{18} - p_{19}x_{19}, \\ \dot{x}_{20} &= p_{19}x_{19} - p_{20}x_{20}. \end{aligned} \right. \quad (24)$$

Saccomani et al. [35] considered the analysis of this system by means of the *differential algebra approach* using DAISY software. They concluded that the model is structurally globally identifiable after 150 min in a computer of 3.13 GHz and 3GB RAM.

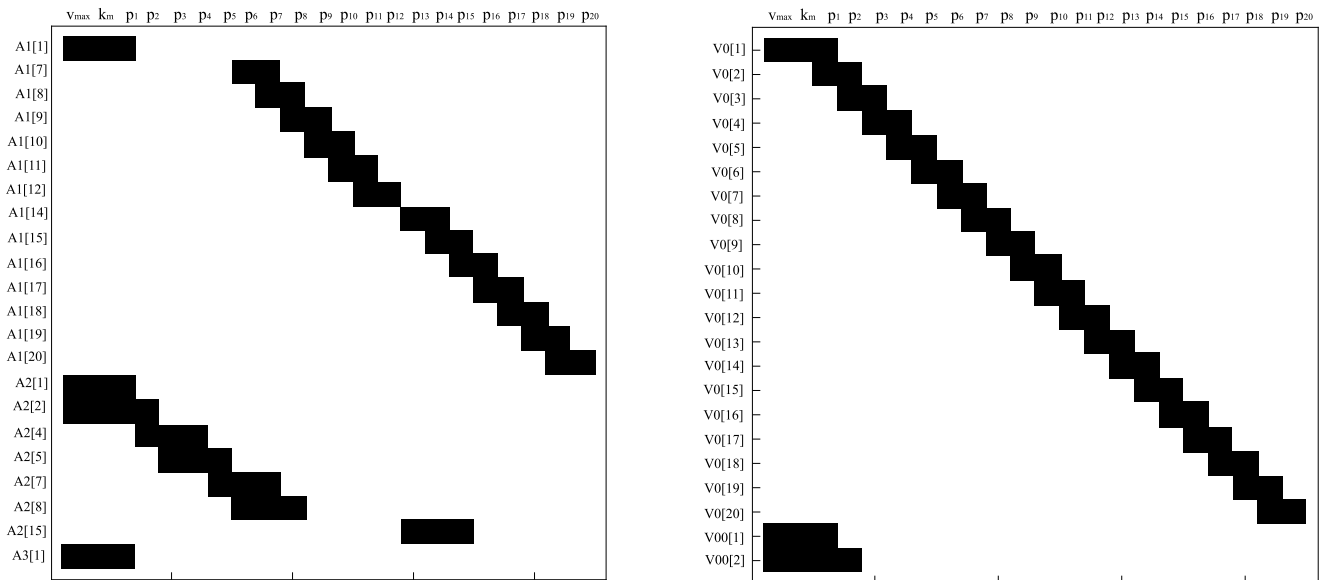
The application of the *Taylor series approach* in combination with the identifiability *tableaus* resulted in structural global identifiability of the model in a few seconds. The reduced identifiability tableau (Figure 4.(a)) needed only 3 derivatives to achieve the maximum rank 22. The solution of the algebraic system was given by considering the following groups of parameters:  $v_{max}, k_m, p_1$ , then,  $p_2$  can be calculated individually. Knowing the solution of these parameters, the next group to be computed is given by  $p_{13}, p_{14}, p_{15}, p_{16}, p_{17}, p_{18}, p_{19}, p_{20}$ , and  $p_6, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}$ . The fourth group of parameters is  $p_3, p_4, p_5$ . All 22 parameters have unique solution, so the model (24) is structurally globally identifiable.

The *generating series approach* in combination with the identifiability *tableaus* also concludes that the model is structurally globally identifiable. The corresponding identifiability *tableau* is represented in Figure 4.(b). All the results were computed in approximately 4s on a computer of 2.66 GHz and 8GB RAM.

The *similarity transformation method* requires observability and controllability rank conditions. To prove the observability rank condition we should calculate the rank of the subspace generated by consecutive differentials of  $\mathbf{h}(\mathbf{x})$  and  $\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}$ . The rank 22 was obtained in MATLAB, in a few minutes, after five iterations. Unfortunately, the controllability condition could not be assessed due to computational requirements.

The *direct test* did not provide conclusive information about the identifiability of the parameters. A unique solution was obtained, but it does not comply with the structural identifiability rules, in the sense that from  $\mathbf{f}(\mathbf{x}, \mathbf{p}) = \mathbf{f}(\mathbf{x}, \mathbf{p}^*)$ , we could not find a solution  $\mathbf{p} = \mathbf{p}^*$ , as required.

The *implicit function theorem* was successfully applied to the problem. The computations were rather simple in this case since all the state variables were measured. With an extra derivative of the corresponding output, the rank condition of the identifiability Jacobian matrix was fulfilled, and so the structural local identifiability was confirmed.



**Figure 4. High dimensional nonlinear model: Identifiability tableaus.** (a) Identifiability *tableau* obtained by means of the Taylor series method, (b) Identifiability *tableau* obtained by means of the generating series method. doi:10.1371/journal.pone.0027755.g004

For this example, it is possible to apply the *identifiability analysis for dynamic reaction networks approach* by defining the corresponding stoichiometric matrix  $N \in \mathbb{R}^{20 \times 21}$ , with the matrix of measured states  $N_m$  of rank 20. Since  $\text{rank}(N_m) = n_R$  then the reaction rate identifiability is satisfied and we can directly apply the generating series approach for all reaction rates. Results coincide with the direct application of the generating series, i.e. the model is structurally globally identifiable.

The first matrix indicated the identifiability of  $k_3, k_{prod}, k_{deg}, i_1$ . The second matrix showed the identifiability of  $t_1, k_1, k_2$ ; the third,  $t_2, c_{4a}, e_{2a}$ ; the fourth,  $c_5, i_{1a}$  and the fifth,  $c_{3a}$ .

Results obtained in this case reveal that nearly linear models with full observation are tractable for most of the methods considered. Major differences rely on the computational cost which ranges from a few seconds (GenSSI) to a couple of hours (DAISY).

### Case study 5: Arabidopsis Thaliana model

The model describes the first multi-gene loop identified in the Arabidopsis circadian clock [36] that comprises a negative feedback loop, in which two partially redundant genes, Late Elongated Hypocotyl (LHY) and Circadian Clock Associated 1 (CCA1), repress the expression of their activator, Timing of CAB Expression 1 (TOC1). A minimal mathematical representation of the system requires 7 coupled differential equations and 29 parameters. The differential equations involve Michaelis-Menten kinetics that describe enzyme-mediated protein degradation, and Hill functions that describe some transcriptional activation terms. The model is given by [36]:

$$\begin{cases} \dot{x}_1 = n_1 \frac{x_6}{g_1 + x_6} - m_1 \frac{x_1}{k_1 + x_1} + q_1 x_7 u(t), \\ \dot{x}_2 = p_1 x_1 - r_1 x_2 + r_2 x_3 - m_2 \frac{x_2}{k_2 + x_2}, \\ \dot{x}_3 = r_1 x_2 - r_2 x_3 - m_3 \frac{x_3}{k_3 + x_3}, \\ \dot{x}_4 = n_2 \frac{g_2^2}{g_2^2 + x_3^2} - m_4 \frac{x_4}{k_4 + x_4}, \\ \dot{x}_5 = p_2 x_4 - r_3 x_5 + r_4 x_6 - m_5 \frac{x_5}{k_5 + x_5}, \\ \dot{x}_6 = r_3 x_5 - r_4 x_6 - m_6 \frac{x_6}{k_6 + x_6}, \\ \dot{x}_7 = p_3 - m_7 \frac{x_7}{k_7 + x_7} - (p_3 + q_2 x_7) u(t), \\ x_i(0) = 0, i = 1, \dots, 7. \end{cases} \quad (25)$$

The observations correspond to the luminescence and the mRNA:  $y_1 = x_1, y_2 = x_4$  [36]. In order to analyse the role of the control variable related to the light intensity we considered the situation for which light intensity is kept constant to its maximum ( $\mathbf{u} = 1$ ) and the case corresponding to a pulse-wise light stimulation.

Results reveal that the model is not structurally globally identifiable for the case with  $\mathbf{u} = 1$  not even structurally locally identifiable since a subset of model parameters are not identifiable ( $m_7, k_7, p_3, q_1$  and  $q_2$ ).

Under the pulse-wise stimulation the *Taylor series approach*, implemented in MATHEMATICA, reached 10 derivatives. Note that this means having only 20 Taylor coefficients that result into a rank 14 identifiability *tableau*. From the parameters appearing in the *tableau* ( $n_1, n_2, m_1, m_4, m_5, m_6, m_7, k_1, k_4, k_5, k_7, p_2, p_3, q_2$ ) only  $p_3$  and  $n_2$  could be regarded as globally

identifiable, since it was not possible to solve the system of equations for the remaining parameters. More derivatives would be required to get further results. However the task was computationally too demanding.

The *generating series approach* was able to reach the 11<sup>th</sup> derivative resulting in an identifiability *tableau* of rank 16. In this case a unique solution could be computed for  $p_3, n_2, q_2, m_1, k_1$ . Similarly to what happened with the Taylor method, further derivatives would be required, but the task is too demanding from the computational point of view.

The *similarity transformation method* could not be applied to this example since the observability condition is not satisfied.

The *direct test method* was also not applicable since the model is controlled.

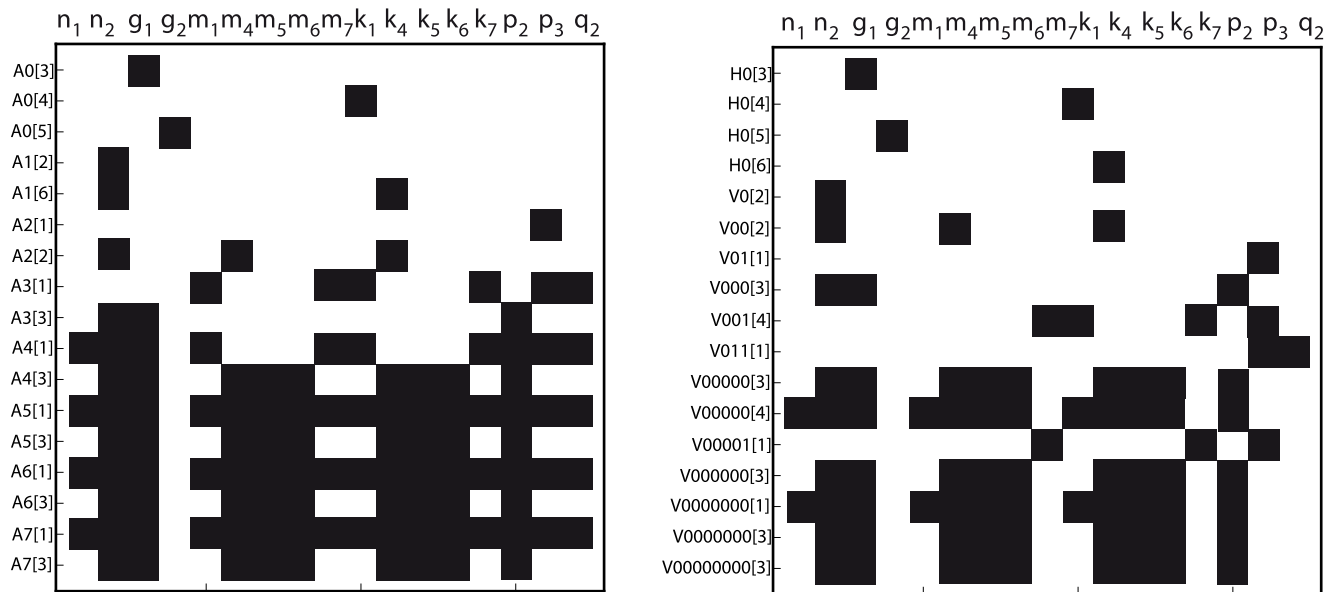
The *differential algebra approach* was not successful in providing results for this example. Both the MAPLE and DAISY implementations reported computational errors due to lack of memory.

As in previous examples, we also resorted to rewrite the model (25) in a pure polynomial form, as a system of 16 differential equations, given below:

$$\begin{cases} \dot{x}_1 = n_1 x_6 x_8 - m_1 x_1 x_9 + q_1 x_7 u(t), \\ \dot{x}_2 = p_1 x_1 - r_1 x_2 + r_2 x_3 - m_2 x_2 x_{10}, \\ \dot{x}_3 = r_1 x_2 - r_2 x_3 - m_3 x_3 x_{11}, \\ \dot{x}_4 = n_2 g_2^2 x_{12} - m_4 x_4 x_{13}, \\ \dot{x}_5 = p_2 x_4 - r_3 x_5 + r_4 x_6 - m_5 x_5 x_{14}, \\ \dot{x}_6 = r_3 x_5 - r_4 x_6 - m_6 x_6 x_{15}, \\ \dot{x}_7 = p_3 - m_7 x_7 x_{16} - (p_3 + q_2 x_7) u(t), \\ \dot{x}_8 = -\dot{x}_6 x_8^2, \\ \dot{x}_9 = -\dot{x}_1 x_9^2, \\ \dot{x}_{10} = -\dot{x}_2 x_{10}^2, \\ \dot{x}_{11} = -\dot{x}_3 x_{11}^2, \\ \dot{x}_{12} = -2x_3 \dot{x}_3 x_{12}^2, \\ \dot{x}_{13} = -\dot{x}_4 x_{13}^2, \\ \dot{x}_{14} = -\dot{x}_5 x_{14}^2, \\ \dot{x}_{15} = -\dot{x}_6 x_{15}^2, \\ \dot{x}_{16} = -\dot{x}_7 x_{16}^2, \\ x_0(p) = \left(0, 0, 0, 0, 0, 0, 0, \frac{1}{g_1}, \frac{1}{k_1}, \frac{1}{k_2}, \frac{1}{k_3}, \frac{1}{g_1^2}, \frac{1}{k_4}, \frac{1}{k_5}, \frac{1}{k_6}, \frac{1}{k_7}\right). \end{cases} \quad (26)$$

Using this pure polynomial form, and the corresponding observable states  $y_1 = x_1, y_2 = x_4, y_3 = x_8, y_4 = x_9, y_5 = x_{12}, y_6 = x_{13}$ , it was possible to extract more information about model identifiability. Using the Taylor series approach, we found an identifiability *tableau* of rank 17, using 7 derivatives. So, at least local identifiability could be checked for the corresponding subset of parameters, as represented in Figure 5(a). For this model formulation, uniqueness of solution was obtained for  $g_1, k_1, g_2, n_2, k_4, p_3, m_4$ .

Additional information could also be obtained using the generating series approach. The corresponding identifiability *tableau* for this method had rank 17, using 8 derivatives (see the corresponding reduced *tableau* in Figure 5(b)). For this model formulation it was possible to compute unique solutions for  $g_1, k_1, g_2, k_4, n_2, m_4, p_3, p_2, q_2, k_7, m_7$ . Therefore, even though pure



**Figure 5. Arabidopsis Thaliana model: Reduced identifiability tableaux.** Reduced identifiability tableau obtained by means of the (a) Taylor series and (b) generating series methods applied to the polynomial form of the model. doi:10.1371/journal.pone.0027755.g005

polynomial forms result in greater computational costs, they usually provide more informative results.

It should be noted that some parameters ( $m_2, m_3, k_2, k_3, p_1, r_1$  and  $r_2$ ) did not appear in the identifiability tableaux despite the large number of coefficients used in both Taylor and generating series approaches (42 and 96, respectively). In addition, higher order coefficients were always dependent on the same parameters, as it was shown by the patterns appearing in the last rows of both tableaux. To further illustrate this point, the complete identifiability tableau obtained by means of the generating series approach is presented in Figure 6.

These results can be complemented with a global sensitivity analysis as proposed in [17]. For this example, the analysis was performed under a pulse-wise experimental scheme and the results revealed that those parameters are in fact slightly influencing the model output, thus they are expected to be structurally locally identifiable even though poorly practically identifiable.

The application of the differential algebra approach resulted in computational errors when trying to apply the initial conditions.

In order to apply the method for reaction networks the control  $u(t)$  should be constant. This allows to derive a stoichiometric matrix  $N \in \mathbf{R}^{16 \times 7}$  with the matrix of measured states  $N_m$  of rank 7. Five stoichiometric matrices of rank 4 could be achieved provided we impose the condition  $q_1 = 0$ . By using the generating series it is then possible to confirm the global identifiability of  $m_1, m_2, m_3, k_1, k_2, k_3, k_6$  and the local identifiability of  $m_5$  and  $k_5$ . It should be noted that the method fails when trying to use the initial conditions.

The results for this case study reflect that a reduced number of observables as compared to the number of parameters poses serious problems for all methods. This will lead, in the best case, to partial solutions related to a sub-set of model parameters. In addition, as for the case of Goodwin's model, results help to decide on the type of experiment to be performed, in this case how to stimulate the system, to improve structural identifiability.

#### Case study 6: NF $\kappa$ B model

The model of the NF $\kappa$ B regulatory module, as proposed by [4], is characterised by two compartment kinetics of the activators *IKK* and *NF*- $\kappa$ B, the inhibitors *A20* and *I $\kappa$ B $\alpha$* , and their complexes. The model is described by the differential system:

$$\begin{cases}
 \dot{x}_1 = k_{prod} - k_{deg}x_1 - k_1x_1u(t), \\
 \dot{x}_2 = -k_3x_2 - k_{deg}x_2 - a_2x_2x_{10} + t_1x_4 - a_3x_2x_{13} \\
 \quad + t_2x_5 + (k_1x_1 - k_2x_2x_8)u(t), \\
 \dot{x}_3 = k_3x_2 - k_{deg}x_3 + k_2x_2x_8u(t), \\
 \dot{x}_4 = a_2x_2x_{10} - t_1x_4, \\
 \dot{x}_5 = a_3x_2x_{13} - t_2x_5, \\
 \dot{x}_6 = c_{6a}x_{13} - a_1x_6x_{10} + t_2x_5 - i_1x_6, \\
 \dot{x}_7 = i_1k_vx_6 - a_1x_{11}x_7, \\
 \dot{x}_8 = c_4x_9 - c_5x_8, \\
 \dot{x}_9 = c_2 + c_1x_7 - c_3x_9, \\
 \dot{x}_{10} = -a_2x_2x_{10} - a_1x_{10}x_6 + c_{4a}x_{12} - c_{5a}x_{10} - i_{1a}x_{10} + e_{1a}x_{11}, \\
 \dot{x}_{11} = -a_1x_{11}x_7 + i_{1a}k_vx_{10} - e_{1a}k_vx_{11}, \\
 \dot{x}_{12} = c_{2a} + c_{1a}x_7 - c_{3a}x_{12}, \\
 \dot{x}_{13} = a_1x_{10}x_6 - c_{6a}x_{13} - a_3x_2x_{13} + e_{2a}x_{14}, \\
 \dot{x}_{14} = a_1x_{11}x_7 - e_{2a}k_vx_{14}, \\
 \dot{x}_{15} = c_{2c} + c_{1c}x_7 - c_{3c}x_{15}.
 \end{cases} \quad (27)$$

In their paper, Lipniacki et al. fixed some of the model parameters by using values from the literature. In order to assign values to the following unknown parameters:

$$\mathbf{p} = [t_1, t_2, c_{3a}, c_{4a}, c_5, k_1, k_2, k_3, k_{prod}, k_{deg}, i_1, e_{2a}, i_{1a}]^T. \quad (28)$$



**Figure 6. Arabidopsis Thaliana model: Full identifiability tableau.** Identifiability tableau obtained by means of the generating series method applied to the polynomial form of the model. Despite the large number of terms included in the tableau some parameters are not appearing. The analysis may be complemented with global sensitivity analysis. doi:10.1371/journal.pone.0027755.g006

They used experimental data from previous works by Lee et al. [57] and Hoffmann et al. [58] which corresponded to the observation of  $y_1 = x_7, y_2 = x_{10} + x_{13}, y_3 = x_9, y_4 = x_1 + x_2 + x_3, y_5 = x_2, y_6 = x_{12}$ .

The application of the *Taylor and generating series approaches*, with the help of the identifiability tableaus, to analyse the structural identifiability of the parameters in the vector  $p$  was discussed in Balsa-Canto et al. [17]. These authors found that the complexity

of the equations resulting from the Taylor series approach prevented drawing conclusions on the identifiability of most of the parameters. The application of the generating series approach resulted, as expected, in a simpler system of equations. In fact it was possible to obtain as many coefficients as necessary to guarantee full rank Jacobian. In addition, the iterative solution of the set of non-linear equations resulted in the structural global identifiability of the parameters in  $p$ .

Since the observability rank condition is not satisfied in this case, the *similarity transformation method* was not applicable. Since the system is controlled, the *direct test method* could not be applied.

The *differential algebra approach* was not successful in providing results for this example. Both implementations of the method, the one based on MAPLE and DAISY, resulted in computational errors (lack of memory problems) and were unable to calculate the characteristic set. The same reason precluded the application of the *implicit function theorem based method*.

For this example, it was possible to apply the *identifiability analysis for dynamic reaction networks approach*. The stoichiometric matrix was formed,  $N \in \mathbb{R}^{32 \times 15}$ , with the matrix of measured states  $N_m$  of rank 7. Five stoichiometric matrices of rank 7 were required to test the identifiability of the parameters in  $\mathbf{p}$ . The first matrix indicated the identifiability of  $k_3, k_{prod}, k_{deg}, i_1$ . The second matrix showed the identifiability of  $t_1, k_1, k_2$ ; the third,  $t_2, c_{4a}, e_{2a}$ ; the fourth,  $c_5, i_{1a}$  and the fifth,  $c_{3a}$ .

As a summary, it can be concluded that the generating series approach, and the chemical reaction network theory combined with the generating series method, are the most suitable methods to handle generalised mass action models, particularly when the number of observables is limited and the number of derivatives required is too large for the Taylor and differential algebra methods (which are computationally not feasible for those cases).

## Discussion

The selected examples include small and medium-size models which incorporate the typical non-linear terms found in systems biology models, such as generalised mass action, Michaelis-Menten or Hill kinetics. The analysis was performed taking into account realistic measured variables (observables) available in experimental labs. For the case of the Goodwin oscillator, a hypothetical situation with full observation was also considered to illustrate how the addition of observables can improve structural identifiability.

The results (summarised in Table 1) reveal some apparent conflicting conclusions regarding the local or global identifiability of the models considered. This may be explained by taking into account that the Taylor and generating series approaches use initial conditions and symbolic quantities to solve the final

algebraic system of equations on the parameters. Local identifiability is concluded when a) several solutions are found for the parameters (in the whole set of real numbers) or b) the system of equations is too complex to be fully solved. Note that in these cases local identifiability could be transformed into global identifiability when knowing the domain of definition of the parameters (for example, positive real numbers).

Differential algebra based methods use randomly generated numerical values to handle complicated systems of equations in the parameters. Thus they may conclude global identifiability in the cases where Taylor or generating series are concluding at least local identifiability. In addition in some cases DAISY does not use initial conditions for the calculations despite their critical role in the analysis [59] being then possible that results may change from local to global. This is clearly the case when some initial conditions are zero.

Regarding a comparison of the performance of the different methods the following criteria have been used: a) range of applicability, b) computational complexity and c) information provided by the method. A general overview of the requirements, advantages and disadvantages of all methods considered is presented in Table 2.

The *Taylor series approach* is probably the most general method since it can be applied to any type of non-linear model. It is also conceptually simple as it relies on the uniqueness of a Taylor expansion of the observables around  $t_0$ . Thus the implementation and the application of the method do not require advanced mathematical knowledge. Its major drawback is that the number of required derivatives is generally unknown and it may become rather large particularly for the cases where the number of observables is small as compared to the number of parameters. In addition, final algebraic symbolic manipulations can become too complicated when solving the resulting systems of equations in the parameters. Even though, this may be partially solved by means of the *identifiability tableaux*, for some particular examples the method may be ultimately unable to provide exact information on the local/global identifiability of the parameters.

The *differential algebra based method* is based on the definition of the observables dynamics as functions of the observables by manipulating the original model. Possibly the major advantage with respect to series based methods is that it is conclusive for

**Table 1.** Summary of results obtained by the different methods.

	T.S.	G.S.	S.T.	D.T.	D.A.	I.F.T.	I.D.R.N.
Goodwin one obs	NR	NR	NA	NC	NR	NA	NA
Goodwin full obs	SLI	SLI	NA	NC	SNI	SLI ( $\sigma > 2$ )	SLI ( $\sigma$ , A fixed)
Goodwin poly. form, 1 obs	SLI	SLI	NA	NC	NR	NA	NA
Goodwin poly. form, full obs	SGL	SGL	NA	NC	SNI no i.c.	SLI no i.c.	NA
Pharma. one obs	SLI	SLI	NA	NC	NR	NR	SLI some pars.
Pharma. two obs	SLI	SLI	NA	NC	SGL	NR	NA
Glycolysis	SLI	SGL	NA	NA	SGL no i.c.	SLI	SGL
High dim. model	SGL	SGL	NR	NC	SGL	SLI	SGL
Arabidopsis clock	SLI 14 pars.	SLI 16 pars.	NA	NA	NR	NA	SLI 12 pars.
NFKB	SLI some pars.	GLI	NA	NA	NR	NR	GLI

T.S.: Taylor series approach; G.S.: generating series approach; S.T.: Similarity transformation approach; D.T.: Direct test; D.A.: differential algebra based approach; I.F.T.: method based on the implicit function theorem; I.D.R.N.: identifiability analysis based on the reaction network theory; SGL: structural global identifiable, SLI: (at least) structural local identifiable, SNI: structural non-identifiable, NA: not applicable, NC: not conclusive and NR: no results were reported due to computational errors or requirements.

doi:10.1371/journal.pone.0027755.t001



**Table 2.** Summary of requirements, advantages and disadvantages for all methods.

<b>T.S.</b>	Requirements	- <b>f; g; h</b> may be non-linear with any dependency on <b>u</b> - <b>x; y; f; g; h</b> allow for infinite derivatives w.r.t. time/states
	Advantages	- conceptually simple - enhanced performance with identifiability <i>tableaus</i>
	Disadvantages	- unknown number of required derivatives - computationally demanding for low number of observables or when the initial conditions are not informative
<b>G.S.</b>	Requirements	- <b>f; g; h</b> may be non-linear but linear dependency on <b>u</b> - <b>x; y; f; g; h</b> allow for infinite derivatives w.r.t. time/states
	Advantages	- conceptually simple - simpler algebra and less computational cost than T.S. - enhanced performance with identifiability <i>tableaus</i> - software available (GenSSI)
	Disadvantages	- unknown number of required derivatives - computationally demanding for low number of observables or when the initial conditions are not informative
<b>S.T.</b>	Requirements	- linear dependence on <b>u</b> that must be bounded and measured - controllability and observability conditions
	Advantages	- software available for part of the analysis
	Disadvantages	- results in a complicated set of partial differential equations - computationally demanding
<b>D.T.</b>	Requirements	- uncontrolled systems
	Advantages	- conceptually simple
	Disadvantages	- requires complicated algebraic manipulations - computationally demanding
<b>D.A.</b>	Requirements	- <b>f; g; h</b> polynomial or rational and <b>u</b> differentiable - generic controllability
	Advantages	- software available (DAISY) - conclusive non-identifiability
	Disadvantages	- rational models are to be reduced to polynomial form - computationally demanding - limited performance when the number of observables is low
<b>I.F.T.</b>	Requirements	- <b>f; g; h</b> non-linear, differentiable and <b>u</b> differentiable
	Advantages	- characteristic set may be obtained with existing software
	Disadvantages	- complicated identifiability matrix - limited performance when the number of observables is low
<b>I.D.R.N.</b>	Requirements	- chemical reaction networks - combined with other methods
	Advantages	- analysis by groups of reaction rates - computationally simple - efficiency in combination with generating series (G.A.)
	Disadvantages	- only suitable for chemical reaction networks - reaction rates needed for identifiability analysis

doi:10.1371/journal.pone.0027755.t002

structurally non-identifiable models. Even though advanced mathematical skills are required so as to understand and implement the method, the recently developed DAISY software [25] enables its application to non-expert users. The major drawbacks appear in the analysis of models incorporating Michaelis-Menten and Hill kinetics, even when transforming the models to pure polynomial forms as suggested by Margaria and coworkers [39]. In addition, the method presents serious difficulties when the number of observables is low as compared

to the number of parameters and the computation of the characteristic polynomial requires high order derivatives.

The applicability of the *similarity transformation approach* relies on the verification of the observability and controllability conditions and the local state isomorphism theorem. Despite many mathematical packages incorporate functions to check the observability and controllability of a given model, in home implementations are required to verify the local state isomorphism conditions. In addition, in many cases, such as most of the

examples considered in this contribution, the observability condition may not be fulfilled or the associated computational burden may be too large thus precluding its application. Additional difficulties might arise when trying to analytically solve the differential equations (10)–(14).

The *direct test method* is only applicable to autonomous and uncontrolled systems. Although it is conceptually the simplest approach, for the examples considered, no reliable results could be achieved due to the complexity of the associated algebraic manipulations.

The *implicit function theorem based method* is, in principle, applicable to any differentiable. As for the case of the differential algebra approach, the method relies on the derivation of the characteristic polynomial. Thus, its complexity grows rapidly when the number of observables is low as compared to the number of parameters. In addition, it only provides information about local identifiability.

The *CRNT based method* is applicable to models that can be written in the CRNT form. This may be difficult for some particular cases with Michaelis-Menten or Hill kinetics or when the corresponding reaction network is unknown (as in some examples considered here). Results rely on the application of another identifiability analysis method, in particular the use of the generating series approach enhances the overall efficiency of the method.

The *generating series approach* in combination with the identifiability *tableaus* offers the most advantageous compromise regarding applicability, computational complexity and information provided. Its computational requirements are significantly lower than the Taylor or the differential algebra approaches, and the information provided is often more precise. This is mainly due to the following facts: i) the required number of derivatives is usually lower than for the other methods and ii) the identifiability *tableaus* are sparser, meaning that the system of non-linear equations on the parameters is simpler, thus providing more information to distinguish between local and global identifiability. The recently developed toolbox GenSSI [52] eases the application of this methodology, offering access to intermediate results throughout the process and allowing for the easy incorporation of known numeric or symbolic initial conditions to the analysis.

Since the structural identifiability analysis will be embedded in a larger systems biology work flow, the selection of the most adequate approach for the model under consideration will be

critical. In this concern, we would suggest the use of the generating series approach in combination with the identifiability *tableaus* as implemented in GenSSI [52] exploiting the CRNT structure when possible. To get conclusive results on the possible structural non-identifiability of a sub-set of parameters for a given model the use of DAISY is suggested. The use of the Taylor approach is only recommended for those rare cases where control dependence is non-linear. Unfortunately remaining methods seem not be adequate to handle typical systems biology models.

## Conclusions

The unique identification of parameters in systems biology models is a very challenging task. The problem becomes especially hard in the case of large and highly non-linear models. In fact, in some cases it will be impossible to compute a unique value for the parameters independently of the available experimental data. This is particularly true for models where the ratio between the number of observables and the number of parameters is low, or when complicated non-linear terms, such as Michaelis-Menten or Hill kinetics, are present. This frequently results in a lack of structural identifiability, which is therefore a key property of these models.

In this work, we have presented a critical comparison of the available techniques for the analysis of structural identifiability of non-linear dynamic models by means of a collection of models related to biological systems of increasing size and complexity.

Results reveal that the combination of the generating series approach with identifiability *tableaus* [17] offers the best compromise between range of applicability, computational complexity and information provided.

## Supporting Information

### Supporting Information S1 Details on the application of the structural identifiability methods for Goodwin's model.

(PDF)

## Author Contributions

Conceived and designed the experiments: EBC JRB. Performed the experiments: OTC EBC. Analyzed the data: EBC OTC JRB. Contributed reagents/materials/analysis tools: EBC OTC. Wrote the paper: EBC OTC JRB.

## References

1. Wolkenhauer O, Ullah M, Kolch W, Cho K (2004) Modeling and simulation of intracellular dynamics: Choosing an appropriate framework. *IEEE Trans on Nanobioscience* 3(3): 200–207.
2. Janes K, Lauffenburger D (2006) A biological approach to computational models of proteomic networks. *Curr Op Chem Biol* 10: 73–80.
3. Banga JR, Balsa-Canto E (2008) Parameter estimation and optimal experimental design. *Essays in Biochemistry* 45: 195–210.
4. Lipniacki T, Paszek P, Brasier A, Luxon B, Kimmel M (2004) Mathematical model of NF $\kappa$ B regulatory module. *J Theor Biol* 228: 195–215.
5. Brown K, Hill C, Calero G, Myers C, Lee K, et al. (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys Biol* 1: 184–195.
6. Achard P, Schutter ED (2006) Complex parameter landscape for a complex neuron model. *PLOS Computational Biology* 2: 0794–0803.
7. Piazza M, Feng X, Rabinowitz J, Rabitz H (2008) Diverse metabolic model parameters generate similar methionine cycle dynamics. *J Theor Biol* 251: 628–639.
8. Gutenkunst R, Waterfall J, Casey F, Brown K, Myers C, et al. (2007) Universally sloppy parameter sensitivities in systems biology models. *Plos Comput Biol* 3: 1871–1878.
9. Walter E, Pronzato L (1997) Identification of Parametric Models from Experimental Data. Springer, Masson.
10. Balsa-Canto E, Alonso A, Banga J (2008) Computational procedures for optimal experimental design in biological systems. *IET Systems Biology* 2(4): 163–172.
11. Bandara S, Scöder J, Eils R, Bock H, Meyer T (2009) Optimal experimental design for parameter estimation of a cell signaling model. *Plos Comput Biol* 5: 1–12.
12. Kreutz C, Timmer J (2009) Systems biology: experimental design. *FEBS J* 276: 923–942.
13. He F, Brown M, Yue H (2010) Maximin and bayesian robust experimental design for measurement set selection in modelling biochemical regulatory systems. *Int J Robust & Nonlinear Control* 20: 1059–1078.
14. Rodriguez-Fernandez M, Egea JA, Banga J (2006) Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics* 7: 483.
15. Srinath S, Gunawan R (2010) Parameter identifiability of power-law biochemical system models. *J Biotechnol* 149: 132–140.
16. Hengl S, Kreutz D, Timmer J, Maiwald T (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* 23(19): 2612–2618.
17. Balsa-Canto E, Alonso A, Banga J (2010) An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Systems Biology* 4: 11.
18. Roper R, Saccomani M, Vicini P (2010) Cellular signaling identifiability analysis: a case study. *J Theor Biol* 264: 528–537.
19. Kholodenko B (2006) Cell-signalling dynamics in time and space. *Nature Reviews, Molecular Cell Biology* 7: 165–176.
20. Miao H, Xia X, Perelson A, Wu H (2011) On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM Rev Soc Ind Appl Math* 53(1): 3–39.

21. Pohjanpalo H (1978) System identifiability based on power-series expansion of solution. *Math Biosci* 41: 21–33.
22. Walter E, Lecourtier Y (1982) Global approaches to identifiability testing for linear and nonlinear state space models. *Mathematics and Computers in Simulation* 24: 472–482.
23. Vajda S, Godfrey K, Rabitz H (1989) Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Mathematical Biosciences* 93: 217–248.
24. Ljung L, Glad T (1994) On global identifiability of arbitrary model parameterizations. *Automatica* 30: 265–276.
25. Bellu G, Saccomani MP, Audoly S, D'Angio L (2007) DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine* 88: 52–61.
26. Denis-Vidal L, Joly-Blanchard G, Noiret C (2001) Some effective approaches to check the identifiability of uncontrolled nonlinear systems. *Mathematics in Computers and Simulation* 57: 35–44.
27. Walter E, Braems I, Jaulin L, Kieffer M (2004) Guaranteed numerical computation as an alternative to computer algebra for testing models for identifiability. In: *Lecture Notes in Computer Science*, Academic Press, pp 124–131.
28. Xia X, Moog CH (2003) Identifiability of nonlinear systems with applications to hiv/aids models. *IEEE Trans Aut Cont* 48: 330–336.
29. Craciun G, Pantea C (2008) Identifiability of chemical reaction networks. *Journal of Mathematical Chemistry* 44: 244–259.
30. Davidescu F, Jorgensen S (2008) Structural parameter identifiability analysis for dynamic reaction networks. *Chemical Engineering Science* 63: 4754–4762.
31. Szederkenyi G (2009) Comment on “Identifiability of chemical reaction networks” by G. Craciun and C. Pantea. *J Math Chem* 45: 1172–1174.
32. Goodwin A, Defibaugh D, Weber L (1992) The vapor pressure of 1,1,1,2-tetrafluoroethane (r134a) and chlorodifluoromethane (r22). *Int J Thermophys* 13: 837.
33. Domurado M, Domurado D, Vansteenkiste S, Marre AD, ESchacht (1995) Glucose oxidase as a tool to study in vivo the interaction of glycosylated polymers with the mannose receptor of macrophages. *J Contr Rel* 33: 115–123.
34. Bartl M, Kotzing M, Kaleta C, Schuster S, Li P (2010) Just-in-time activation of a glycolysis inspired metabolic network - solution with a dynamic optimization approach. *Proc 55nd International Scientific Colloquium 2010 Ilmenau, Germany*.
35. Saccomani M, Audoly S, Bellu G, D'Angio L (2010) Examples of testing global identifiability of biological and biomedical models with daisy software. *Computers in Biology and Medicine* 40: 402–407.
36. Locke J, Millar A, Turner M (2005) Modelling genetic networks with noisy and varied experimental data: the circadian clock in arabidopsis thaliana. *Journal of Theoretical Biology* 234: 383–393.
37. Vajda S (1984) Structural identifiability of linear, bilinear, polynomial and rational systems. *Proceedings of the 9th IFAC World Congress, Budapest, Hungary*. 107 p.
38. Vajda S (1987) Identifiability of polynomial systems: structural and numerical aspects. *Identifiability of parametric models*, Pergamon, Oxford. 42 p.
39. Margaria G, Riccomagno E, Chappell M, Wynn H (2001) Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Mathematical Biosciences* 174: 1–26.
40. Chappel M, Godfrey K, Vajda S (1990) Global identifiability of the parameters of nonlinear systems with specific input: A comparison of methods. *Mathematical Biosciences* 102: 41–73.
41. Wu H, Zhu H, Miao H, Perelson AS (2008) Parameter identifiability and estimation of hiv/aids dynamic models. *Bulletin of Mathematical Biology* 70(3): 785–799.
42. Walter E, Pronzato L (1996) On the identifiability and distinguishability of nonlinear parametric models. *Math Comput Simulat* 42: 125–26.
43. Denis-Vidal L, Joly-Blanchard G (1996) Identifiability of some nonlinear kinetics. *Proceedings of the Third Workshop on Modelling of Chemical Reaction Systems, Heidelberg*.
44. Vajda S, Rabitz H (1989) Isomorphism approach to global identifiability of nonlinear systems. *IEEE Transactions on Automatic Control* 34: 220–223.
45. Peeters R, Hanzon B (2005) Identifiability of homogeneous systems using the state isomorphism approach. *Automatica* 41: 513–529.
46. Denis-Vidal L, Joly-Blanchard G (2000) An easy to check criterion for (un)identifiability of uncontrolled systems and its applications. *IEEE Transactions on Automatic Control* 45: 768–771.
47. Ollivier F (1990) *Le probleme de l'identifiabilite structurelle globale: etude theorique, methodes effectives et bornes de complexite*. Paris, France: These de Doctorat en Science, Ecole Polytechnique.
48. Buchberger B (1976) A theoretical basis for the reduction of polynomials to canonical forms. *ACM SIGSAM Bulletin* 10(3): 19–29.
49. Ritt J (1950) *Differential algebra*. New York: AMS Colloquium Publications.
50. Kolchin E (1973) *Differential algebra and algebraic groups*. New York: Academic Press.
51. Brendel M, Bonvin D, Marquardt W (2006) Incremental identification of kinetic models for homogeneous reactions systems. *Chemical Engineering Science* 61: 5404–5420.
52. Chis O, Banga J, Balsa-Canto E (2011) GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics*;doi: 10.1093/bioinformatics/btr431.
53. Goodwin B (1965) Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation* 3: 425–428.
54. Verdriere N, Denis-Vidal L, Joly-Blanchard G, Domurado D (2005) Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrohagemannose receptor. *Int J Appl Math Comput Sci* 15: 517–526.
55. Chapman MJ, Godfrey K, Chappell MJ, Evans ND (2003) Structural identifiability of non-linear systems using linear/non-linear splitting. *Control* 76: 209–216.
56. Szederkenyi G, Banga J, Alonso A (2011) Inference of complex biological networks: distinguishability issues and optimization-based solutions. *BMC Systems Biology* in press.
57. Lee E, Boone D, Chai S, Libby S, Chien M, et al. (2000) Failure to regulate TNF-induced NF- $\kappa$ B and cell death responses in A20-deficient mice. *Science* 289: 2350–2354.
58. Hoffmann A, Levchenko A, Scott M, Baltimore D (2002) The I $\kappa$ B-NF- $\kappa$ B signaling module: temporal control and selective gene activation. *Science* 298: 1241–1245.
59. Saccomani M, Audoly S, D'Angio L (2003) Parameter identifiability of nonlinear systems: the role of initial conditions. *Chemical Engineering Science* 39: 619–632.