

Identifiability of Large Nonlinear Biochemical Networks

Alejandro F. Villaverde^{*a,b,c}, Antonio Barreiro^c

^aDepartment of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

^bCentre for Biological Engineering, University of Minho, Braga 4710-057, Portugal

^cDepartment of Systems & Control Engineering, University of Vigo, Vigo 36310, Spain

(Received February 3, 2016)

Abstract

Dynamic models formulated as a set of ordinary differential equations provide a detailed description of the time-evolution of a system. Such models of (bio)chemical reaction networks have contributed to important advances in biotechnology and biomedical applications, and their impact is foreseen to increase in the near future. Hence, the task of dynamic model building has attracted much attention from scientists working at the intersection of biochemistry, systems theory, mathematics, and computer science, among other disciplines—an area sometimes called systems biology. Before a model can be effectively used, the values of its unknown parameters have to be estimated from experimental data. A necessary condition for parameter estimation is *identifiability*, the property that, for a certain output, there exists a unique (or finite) set of parameter values that produces it. Identifiability can be analysed from two complementary points of view: structural (which searches for symmetries in the model equations that may prevent parameters from being uniquely determined) or practical (which focuses on the limitations introduced by the quantity and quality of the data available for parameter estimation). Both types of analyses are often difficult for nonlinear models, and their complexity increases rapidly with the problem size. Hence, assessing the identifiability of realistic dynamic models of biochemical networks remains a challenging task. Despite the fact that many methods have been developed for this purpose, it is still an open problem and an active area of research. Here we review the theory and tools available for the study of identifiability, and discuss some closely related concepts such as sensitivity to parameter perturbations, observability, distinguishability, and optimal experimental design, among others.

*Corresponding author (Email: afernandez@uvigo.es)

1 Introduction: motivating the study of parametric identifiability

Mathematical models play a key role in many scientific areas as well as in engineering practice. They are used not only for representing the available knowledge about a system in an unambiguous, compact form, but also for making informed predictions, among other applications. Models that include information about the system kinetics (*dynamic models*) allow to characterize in detail the evolution of the studied system in time.

Dynamic models are extensively used in chemistry, biology, physics, engineering, and other disciplines. In some applications, such as those involving engineered (designed) systems, it can be relatively straightforward to know the model structure. Their dynamic equations can often be obtained from first principles, and their parameters can be either directly measured or calculated by other means.

In contrast, the mechanisms underlying biological systems—and biochemical networks in particular—are generally known only partially, and must be “reversed engineered” from data. The ability to do so is of high relevance, since dynamic models are increasingly used in biology and medicine, and their use is expected to contribute to important progress in areas such as systems biology [38, 115], metabolic engineering [18, 27, 96], industrial biotechnology [1, 94], or personalized medicine [4, 68], to name a few.

As a motivating example, consider the diagram shown in Figure 1, which represents the network of chemical reactions included in a model of a bioprocess involving Chinese Hamster Ovary cells (CHO) [116]. These cells are used for antibody production by means of a fed-batch fermentation process. Many efforts are currently being made in the biotechnology industry to optimize CHO strains in order to improve process yield and productivity. Such an optimization can be efficiently aided by computer simulations that predict the behaviour of CHO cells after modifications in the enzyme levels of specific reactions, or in different experimental conditions. To perform those simulations, a detailed kinetic model of the metabolic network is required. Dynamic models of CHO cells, consisting of a set of densely parameterized ordinary differential equations (ODEs), have been already used in metabolic engineering [77, 78]. However, the success of their application depends on the accuracy of their predictions, which is limited by the uncertainty that is inherently present in the model.

Biochemical systems of interest such as the cellular network mentioned in this exam-

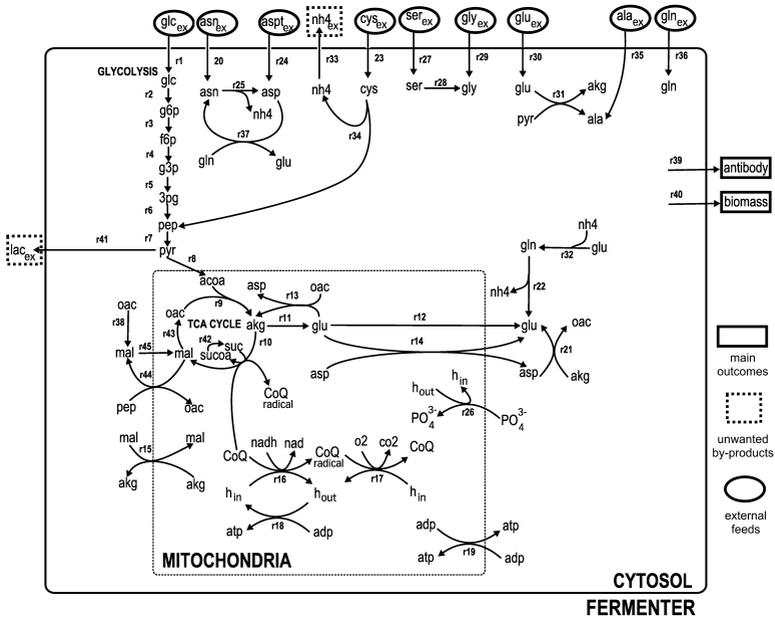


Figure 1: Metabolic network of Chinese Hamster Ovary cells, which are used for antibody production [116]. The concentrations of those metabolites located in the fermenter can be measured; the ones in the cytosol and mitochondria are obtained by model simulation.

ple often have a complex and highly redundant structure, being the result of long term evolution by natural selection [2]. They frequently adopt the form of densely connected networks with many components and abundance of regulatory loops [34, 126]. Even in the best studied of those systems, many details are commonly unknown. Biochemical models are simplifications of such highly complex systems, and as a consequence they usually have a large uncertainty associated to their structure and parameter values. Furthermore, realistic models often have nonlinear dynamics, which further complicates their analysis.

Due to these complexities, it is crucial to be aware of the limitations of current biochemical models in order to avoid creating false expectations and having excessive trust in potentially unreliable results [22]. Limitations may be intrinsic to the model, that is, a direct consequence of its structure, or they may be originated from more external factors such as lack of data necessary for calibration [52]. A number of concepts have been introduced to describe and quantify these limitations, and many methods for calculating

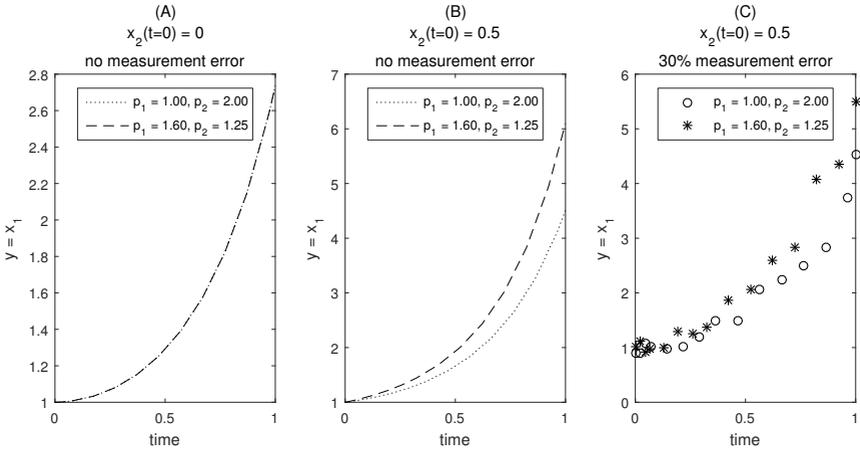


Figure 2: Simple example illustrating structural and practical identifiability issues.

them are available. Terms such as identifiability (structural or practical, *a priori* or *a posteriori*, quantitative or qualitative, theoretical or numerical, geometric, algebraic...), estimability, parameter sloppiness, sensitivity, distinguishability, observability, controllability, reachability, and designability, among others, have been coined.

Simply speaking, a model is identifiable if it is possible to determine, in a unique way, the values of its parameters by observing the model inputs and outputs [125]. Identifiability problems can be illustrated with the following toy example [9]. Let us assume a model consisting of two states, x_1 and x_2 , which represent e.g. the concentrations of two chemical compounds. Assume that their evolution in time is given by the following equations:

$$\begin{aligned} \dot{x}_1 &= p_1 \cdot x_1 \cdot x_2 \\ \dot{x}_2 &= p_2 \cdot u \end{aligned} \tag{1}$$

where u is an external input assumed constant, positive, and known. We can measure the concentration of the first state, that is, $y = x_1$, which initially is $x_1(t = 0) = 1$. The identifiability question is: can the values of the unknown parameters p_1 and p_2 be determined from measurements of $y = x_1$? Figure 2 plots the system output y in several scenarios. In panel (A) the initial condition of x_2 is zero, and two different combinations of parameter values produce the same system output (in fact, any combination of values such that $p_1 \cdot p_2 = 2$ yields the same outcome, so there is an infinite number of possible

solutions for p_1 and p_2). Hence the model is not identifiable. It might be thought that this lack of identifiability is not a problem, because even though the parameters cannot be determined, the output of the model is still correct. However, as panel (B) shows, under a different experimental condition (when $x_2(t = 0) \neq 0$) the output of the model becomes different for the same choice of parameter values. Thus, if we were going to use the model calibrated with data from panel (A) to make predictions in the situation shown in panel (B), we would be almost surely making a wrong prediction (unless by pure chance we had chosen the right combination of values among the infinite possibilities, which is of course highly unlikely). This example illustrates the problems that may appear due to lack of *structural identifiability*.

Let us now assume that the fact that $x_2(t = 0) > 0$ is part of the model specifications, and therefore the output will always be different for different values of p_1 and p_2 , as in panel (B). That is, now we have a structurally identifiable model, and we want to estimate the values of its parameters. In practice, any experimental data will contain some measurement errors, so instead of the situation in panel (B) we will have something like panel (C). The two curves are now hardly distinguishable, and the parameter values estimated from such dataset will have some degree of uncertainty. Quantifying that uncertainty is the aim of *practical identifiability* analysis.

The present paper addresses the study of the identifiability problem, and aims at

1. providing a tutorial introduction to well established concepts and techniques,
2. reviewing recent relevant developments in the area, and
3. providing critical comparisons of the strengths and limitations of the existing approaches.

The focus will be on *large* models, since they provide the appropriate level of description for many applications of interest, and they pose the kind of real-world challenges that do not arise in small or “toy” models. Roughly speaking, a dynamic model of a biochemical network—for example, a metabolic, signaling, or gene regulatory network—can be considered “large” by current standards if it has > 10 states and > 100 parameters. A set of examples of such large-scale dynamic models can be found for example in the BioPreDyn-bench collection [118].

The properties we will study can be classified as structural or non-structural, among other possible distinctions. The former can be derived from the model structure; two examples are (structural) identifiability and distinguishability [125]. Given a model structure with unknown parameters, structural identifiability analysis studies whether it is possible to uniquely determine the parameter values by inspection of the model equations [14]. Given several model structures, distinguishability analysis studies whether it is possible to differentiate between their outputs [110]. Non-structural properties, on the other hand, depend on other factors such as data quality and availability. Parameter sloppiness and the so-called practical or numerical identifiability analysis fall in this category [13, 19].

It should be kept in mind that identifiability analysis is embedded in the more general problem of dynamic model building, which consists of a number of tasks with tight links among them, as shown schematically in Figure 3.

This paper is organized as follows: we begin by introducing in section 2 the concept of sensitivity, which plays an important role in many of the techniques reviewed here. Quantifying the model sensitivity to parameter variations provides a way of ranking the model parameters according to their relative importance. However, they do not paint the whole picture, and a parameter can be difficult or impossible to identify even if the model has a high sensitivity to it, for example if it is correlated with another parameter. Thus in section 3 we study the (structural) identifiability problem, which consists of determining whether the model parameters can be identified in principle. However structural identifiability is not the full picture either, because it does not consider limitations caused by data availability or numerical issues. Therefore we introduce practical identifiability tools in section 4, which can be used for quantifying the uncertainty in the parameter estimates. After that, in section 5 we connect the theory and methodologies of identifiability analysis with other related tasks in the modelling cycle, such as parameter estimation, experiment design, and model discrimination and distinguishability, finishing by discussing whether, despite the ubiquitous uncertainty in model parameters, we can still have confidence in the model predictions.

Notation

In the remaining of the text we will use the following notation: let M be a model structure with a real-valued vector of parameters $\mathbf{p} \in \mathfrak{R}^p$, of inputs $\mathbf{u} \in \mathfrak{R}^r$, of states $\mathbf{x} \in \mathfrak{R}^n$, and of

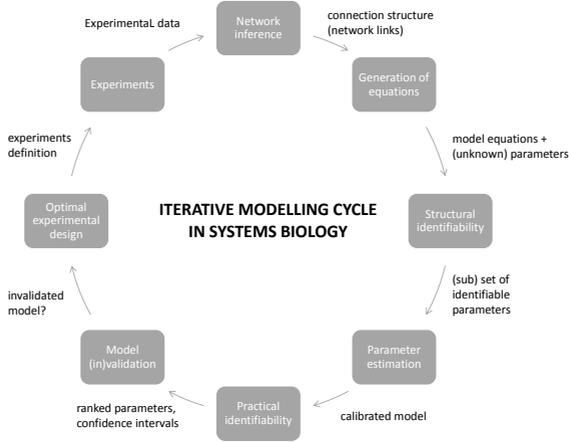


Figure 3: Overview of the main tasks in the modelling cycle. Building a model of a complex biochemical network is usually a time-consuming process, with uncertainties of different types arising at every step. Ideally, it should be carried out in an iterative way, obtaining successive refinements until a the resulting model is deemed satisfactory.

measured outputs $\mathbf{y} \in \mathfrak{R}^m$. The model dynamics are defined by the following equations:

$$\dot{\mathbf{x}}(t) = f[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), t, \mathbf{p}] \quad (2)$$

$$\mathbf{y}(t) = g[\mathbf{x}(t, \mathbf{p}), \mathbf{p}] \quad (3)$$

$$\mathbf{x}_0 = \mathbf{x}(t_0, \mathbf{p}) \quad (4)$$

$$h_{eq}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] = 0 \quad (5)$$

$$h_{in}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] > 0 \quad (6)$$

where f , g , h_{eq} , and h_{in} are nonlinear vector functions, with h_{eq} and h_{in} representing algebraic equality and inequality constraints respectively. In general, simulating M entails determining the values of $\mathbf{x}(t)$ for all t , which in turn requires knowledge of \mathbf{p} .

Abbreviations

DA: Differential Algebra, DSM: Discrete-time Sensitivity Matrix, EE: Elementary Effect, FIM: Fisher Information Matrix, FOLS: First-Order Local Sensitivity, LTI: Linear Time-Invariant, OAT: One parameter At a Time, ODE: Ordinary Differential Equation, PCA:

Principal Component Analysis, PL: Profile Likelihood, SA: Sensitivity Analysis, SVD: Singular Value Decomposition, TS: Taylor Series.

2 Model sensitivity to parameter variations

Given a mathematical model with a possibly large number of parameters, it is natural to ask which parameters are more important, that is, which have a greater effect in the model output. To answer this question we begin by defining the concept of sensitivity. The state variable sensitivities are the partial derivatives of the measured state variables with respect to the parameter values. Mathematically they are defined as follows. Let us consider first, for simplicity, a model with only one state x , which we further assume to be measured, and one parameter p . The (time-varying) *relative sensitivity* of the state variable x to the parameter p is:

$$s_p^x(t) = \frac{\partial x(t, p)/x(t, p)}{\partial p/p} = \frac{\partial x(t, p)}{\partial p} \frac{p}{x(t, p)} \quad (7)$$

where p is the reference or typical value of the parameter and $x(t, p)$ the corresponding state. If the sensitivity is zero (or close to zero, for practical purposes), changes in the parameter p will have no (or almost no) effect on x . It will not be possible to estimate the value of p from measurements of x , and, as will be explained in more detail in the following sections, parameter p will be classified as not identifiable.

Let us now generalize this simple notion to a model with m outputs, n states and p parameters, like the one defined by equations (2)–(6). To this end we perform a Taylor series (TS) expansion of the state variables. We write the TS expansion of state variable x_i around the parameter vector \mathbf{p} as:

$$x_i(t, \mathbf{p} + \Delta\mathbf{p}) = x_i(t, \mathbf{p}) + \frac{\partial x_i}{\partial \mathbf{p}} \Delta\mathbf{p} + \frac{1}{2} \Delta\mathbf{p}^T \frac{\partial^2 x_i}{\partial \mathbf{p}^2} \Delta\mathbf{p} + \dots \quad (8)$$

where the terms of order higher than two are truncated. The total variation $\Delta x_i(t, \mathbf{p})$ is:

$$\Delta x_i(t, \mathbf{p}) = \frac{\partial x_i}{\partial \mathbf{p}} \Delta\mathbf{p} + \frac{1}{2} \Delta\mathbf{p}^T \frac{\partial^2 x_i}{\partial \mathbf{p}^2} \Delta\mathbf{p} + \dots \quad (9)$$

where the first order partial derivative $\frac{\partial x_i}{\partial \mathbf{p}}$ is a p -dimensional vector, and the second order partial derivative, $\frac{\partial^2 x_i}{\partial \mathbf{p}^2}$, is the $p \times p$ Hessian matrix whose elements are $\frac{\partial^2 x_i}{\partial p_j \partial p_k}$.

To obtain the *total* variation in the state x_i due to perturbations Δp_j in the parameters around their nominal value, we must consider *all* the terms present in equation (9).

However, in the following subsection we will neglect the second-order and higher order terms, and we will restrict ourselves to the analysis of *first-order sensitivities*, those that include only the first partial derivative. These sensitivities will not only be first-order but also *local*, since they depend on a particular value of \mathbf{p} , and therefore valid only in its neighbourhood. Note that in doing so we are considering a simplified case. We will extend these notions to the more general *global sensitivity analysis* in section 2.2.

2.1 First order local sensitivities

The *first-order local sensitivity* (FOLS) of state variable x_i to parameter p_j is defined as:

$$\overline{v_{ij}}(t, \mathbf{p}) \equiv v_{ij} = \frac{\partial x_i(t, \mathbf{p})}{\partial p_j} \equiv \frac{\partial x_i}{\partial p_j} \quad (10)$$

The FOLS can be included in a $n \times p$ matrix of first-order local sensitivity functions, $V(t, \mathbf{p})$, whose elements are the v_{ij} . It is important to note that, in general, sensitivities depend not only on the *parameter values*, but also on *time*. Since FOLS are a function of time, their evolution can be described by differential equations. Indeed, since the FOLS also depend on the system states, the ODEs of the model and the ODEs of the FOLS are coupled, and they must be solved simultaneously in order to calculate the time course of the sensitivities. The ODEs of the FOLS can be written as:

$$\begin{aligned} \frac{dv_{ij}}{dt} &= \frac{d}{dt} \left(\frac{\partial x_i}{\partial p_j} \right) = \frac{\partial \dot{x}_i}{\partial p_j} = \frac{\partial}{\partial p_j} f_i[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), t, \mathbf{p}] \\ &= \left(\frac{\partial f_i}{\partial \mathbf{x}} \right) \left(\frac{\partial \mathbf{x}}{\partial p_j} \right) + \left(\frac{\partial f_i}{\partial \mathbf{p}} \right) \left(\frac{\partial \mathbf{p}}{\partial p_j} \right) \end{aligned} \quad (11)$$

$$= \left(\frac{\partial f_i}{\partial \mathbf{x}} \right) \left(\frac{\partial \mathbf{x}}{\partial p_j} \right) + \frac{\partial f_i}{\partial p_j} \quad (12)$$

$$= \sum_{k=1}^n \frac{\partial f_i}{\partial x_k} v_{kj} + \frac{\partial f_i}{\partial p_j} \quad (13)$$

The set of equations can be written in matrix form as a $n \times p$ ODE:

$$\dot{V} = \frac{\partial f}{\partial \mathbf{x}} V + \frac{\partial f}{\partial \mathbf{p}} \quad (14)$$

We have seen how parameter variations affect the state variables; let us now focus on how they affect the model outputs. The $m \times p$ matrix of *output sensitivities* is

$$\frac{\partial \mathbf{y}}{\partial \mathbf{p}} = \left[\frac{\partial y_i}{\partial p_j} \right] = \begin{bmatrix} \frac{\partial y_1}{\partial p_1} & \dots & \frac{\partial y_1}{\partial p_p} \\ \frac{\partial y_2}{\partial p_1} & \dots & \frac{\partial y_2}{\partial p_p} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial p_1} & \dots & \frac{\partial y_m}{\partial p_p} \end{bmatrix} \quad (15)$$

where each element can be calculated, similarly to equations (11)–(13), as

$$\frac{\partial y_i}{\partial p_j} = \frac{\partial g_i[\mathbf{x}(t), t, \mathbf{p}]}{\partial p_j} = \left(\frac{\partial g_i}{\partial \mathbf{x}} \right) \left(\frac{\partial \mathbf{x}}{\partial p_j} \right) + \left(\frac{\partial g_i}{\partial \mathbf{p}} \right) \left(\frac{\partial \mathbf{p}}{\partial p_j} \right) = \sum_{k=1}^n \frac{\partial g_i}{\partial x_k} v_{kj} + \frac{\partial g_i}{\partial p_j} \quad (16)$$

Since the output sensitivities are computed from the state sensitivities, the same remarks that were made for state sensitivities apply here.

If we have a set of d measurements, we can build a *discrete-time sensitivity matrix* (DSM) with $(m \times d)$ rows and p columns as:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{p}} = \left[\frac{\partial y_i}{\partial p_j} \right] = \begin{bmatrix} \frac{\partial y_1(t_1)}{\partial p_1} & \frac{\partial y_1(t_1)}{\partial p_2} & \dots & \frac{\partial y_1(t_1)}{\partial p_p} \\ \frac{\partial y_1(t_2)}{\partial p_1} & \frac{\partial y_1(t_2)}{\partial p_2} & \dots & \frac{\partial y_1(t_2)}{\partial p_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_1(t_d)}{\partial p_1} & \frac{\partial y_1(t_d)}{\partial p_2} & \dots & \frac{\partial y_1(t_d)}{\partial p_p} \\ \frac{\partial y_2(t_1)}{\partial p_1} & \frac{\partial y_2(t_1)}{\partial p_2} & \dots & \frac{\partial y_2(t_1)}{\partial p_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m(t_d)}{\partial p_1} & \frac{\partial y_m(t_d)}{\partial p_2} & \dots & \frac{\partial y_m(t_d)}{\partial p_p} \end{bmatrix} \quad (17)$$

Remark on calculating sensitivities. Note that in equation (14) one can obtain $\frac{\partial f}{\partial \mathbf{x}}$ and $\frac{\partial f}{\partial \mathbf{p}}$ analytically by symbolic differentiation of f . However, to obtain V it is necessary to solve the ODEs of the original dynamical system (2)–(3) together with the sensitivity equations (14), and for this purpose numerical integration is needed. Solvers available for this task include ODESSA [55–57] and CVODES [92], which can be called e.g. from the AMIGO toolbox [12].

2.2 Global sensitivities

Calculation of local sensitivities, taking into account the variation in one parameter at a time, is both the simplest and most common way of performing sensitivity analysis. However, this approach is limited, since it does not consider interaction effects, and not entirely appropriate for assessing the relative importance of several parameters in the output [89]. If possible, one should resort instead to the more general *global sensitivity analysis*, which

explores the whole parameter space and varies all parameters simultaneously—note that in local sensitivity analysis the parameters are changed one at a time (or “OAT”) [38].

Methods for calculating global sensitivities can be classified as variance-based or derivative-based. Variance-based methods aim at estimating the total sensitivity index S_{Ti} , which is the sum of all effects involving a parameter p_i . The total sensitivity index can be thought of as the expected fraction of variance that would remain after all the parameter values, except p_i , have been determined. The Fourier Amplitude Sensitivity Test (FAST) method [35] assigns an integer frequency to each unknown parameter, and calculates the sensitivity indexes by scanning the parameter space. When a particular index S_{Ti} is computed, a high scanning frequency is selected for factor i , while the remaining factors are assigned low frequencies. A Fourier analysis allows to recover the sensitivity indexes. An enhanced version of this method, extended FAST, was presented in [90]. The method of Sobol’ [95] follows a similar approach, although it is less computationally efficient than the extended FAST. A review of these methods can be found in [89]; for an extensive treatment of the subject the reader is referred to the book [88].

A different approach is the one taken by so-called derivative-based methods, which estimate sensitivities by calculating derivatives, as shown in the preceding subsection. While those estimations are local (see e.g. the FOLS), since they are calculated around a particular point, they may be extended to a range of values by performing multiple calculations around different values. However, the remaining parameters are kept constant in this analysis, which is an important difference with respect to variance-based methods.

While variance-based methods can provide more information than derivative-based ones, they are also more computationally expensive, which limits their applicability to large-scale models. As an implementation example, the AMIGO toolbox [12] provides derivative-based procedures for estimating both local and global parameter sensitivities. However, it should be noted that while the “global” sensitivities cover a user-defined range of values in the parameter space, they are calculated OAT, and therefore they would not fit completely in the definitions of “global” sensitivities used e.g. in [38, 89].

2.3 Parameter ranking

Sensitivity functions allow ranking the parameters according to their influence in the model outputs. The parameter ranking may not be unique, as several criteria can be used

for measuring influence, specially in systems with multiple outputs. For example, the following five sensitivity-based indicators have been suggested for assessing the identifiability of parameter sets [20]

$$\begin{aligned} \delta_j^{\text{msqr}} &= \sqrt{\frac{1}{n} \sum_{i=1}^p v_{ij}^2}; & \delta_j^{\text{mabs}} &= \frac{1}{n} \sum_{i=1}^p |v_{ij}|; & \delta_j^{\text{mean}} &= \frac{1}{n} \sum_{i=1}^p v_{ij}; \\ \delta_j^{\text{max}} &= \max_i v_{ij}; & \delta_j^{\text{min}} &= \min_i v_{ij} \end{aligned} \quad (18)$$

These quantities—or others that may be used—are calculated from local sensitivities, that is, they depend on a specific choice of parameter values. Hence the corresponding ranking is local. To escape from the local nature of such an approach, the calculations may be repeated several times by randomly sampling the parameter space. Analysis of the distribution of the resulting sensitivities is then used to assess the relative importance of the inputs globally. When sampling the parameter space computational efficiency is a key factor, since an exhaustive evaluation of different combinations of parameter values would be generally prohibitive. Hence a fundamental challenge is to design a parameter screening method that allows to detect as many significant changes as possible with a reduced number of runs. While a simple Monte Carlo random sampling would be the most obvious choice, other approaches such as the popular Latin Hypercube Sampling (LHS) [66] or the Morris method [75] have been shown to be statistically more efficient. Although the method of Morris can be used to find the subset of most important parameters, it provides only qualitative sensitivity measures. Since it is computationally cheaper than quantitative approaches such as FAST and Sobol, it may be preferred for analysis of large models. Implementations of some of these techniques are available; for example, the AMIGO Matlab toolbox computes parameter rankings from local sensitivity analysis, and also allows using LHS to sample the parameter space efficiently [10, 12].

3 Structural identifiability

The concept of structural identifiability was introduced by Bellman and Åström [14] with the aim of addressing the following question: is it possible to know, by examining the model equations, whether the model parameters can be uniquely determined from measurements of the model outputs? Other authors have referred to this concept with the equivalent terms *a priori* identifiability [7], theoretical identifiability [79], or qualitative identifiability [111], among other names. This fact, together with the introduction of many *ad hoc* definitions, quickly led to a number of ambiguities and misunderstandings,

as already acknowledged by DiStefano and Cobelli in 1980 [30]. In the following we will avoid the specialized jargon as much as possible.

Structural identifiability is a model property depending on the system dynamics, observable functions, external stimuli, and initial conditions [125]. It does not depend on the amount or quality of the available data. Assuming that the model structure M is correct, that the data is noise-free, and that the inputs to the system can be chosen freely, it is always possible to choose an estimated parameter vector $\hat{\mathbf{p}}$ such that the model output $M(\hat{\mathbf{p}})$ equals the one obtained with the true parameter vector, $M(\mathbf{p}^*)$. If $\hat{\mathbf{p}} = \mathbf{p}^*$ this is obviously the case. Parameter p_i is structurally globally (or uniquely) identifiable (s.g.i.) if, for almost any \mathbf{p}^* in P ,

$$M(\hat{\mathbf{p}}) = M(\mathbf{p}^*) \Rightarrow \hat{p}_i = p_i^* \tag{19}$$

A model M is s.g.i. if all its parameters are s.g.i.

A parameter p_i is structurally locally identifiable (s.l.i.) if for almost any \mathbf{p}^* in P there is a neighbourhood $V(\mathbf{p}^*)$ such that

$$\hat{\mathbf{p}} \in V(\mathbf{p}^*) \text{ and } M(\hat{\mathbf{p}}) = M(\mathbf{p}^*) \Rightarrow \hat{p}_i = p_i^* \tag{20}$$

A model M is s.l.i. if all its parameters are s.l.i.

If equation (20) does not hold in any neighborhood of \mathbf{p}^* , parameter p_i is structurally unidentifiable (s.u.i.). A model M is s.u.i. if at least one of its parameters is s.u.i.

3.1 Global structural identifiability

As has been already mentioned, the term structural identifiability was coined by Bellman and Åström. Since they worked in systems and control theory, it was natural for them to study it with the tools commonly used in that field. Thus in their original paper [14] they employed the transfer function, that is, the input-output map of the system, which can be obtained for *linear time-invariant (LTI) systems* with a Laplace transform. They noted that, since the input-output relation of a system is given by its impulse response, if the system is identifiable it can be identified from impulse response measurements. Bellman and Åström also reported results for systems in diagonal and companion form, as well as for some simple compartmental structures, all of which were shown to be structurally identifiable under certain conditions. Other methods proposed for analyzing

structural identifiability of LTI systems include power series expansion [79] and similarity transformation [121], although extensions of these methods for nonlinear systems have been developed too.

For *nonlinear* dynamic models, assessing the global structural identifiability is much more complicated. Analytical approaches generate large symbolic expressions whose treatment quickly becomes impractical as the problem dimension increases, even for models of moderate size. Hence this is currently an open research problem, for which many different approaches have been suggested. In two reviews published in 2011 Chiş *et al* [26] and Miao *et al* [73] compared approaches based on Taylor series [79], generating series [122], similarity transformation [109], differential algebra [61], direct test [37,120], implicit function theorem [127], and test for reaction networks [32,36]. An overview of the main methods is given below; we refer the reader interested in the full details to the original publications or to the review articles [26,73].

The Direct Test approach is arguably the simplest, and hence also very limited in its application. Its basic idea is to compare the right hand side, f , of equation (2). If $f(\mathbf{p}) = f(\mathbf{p}^*) \Rightarrow \mathbf{p} = \mathbf{p}^*$, the two systems (which must be uncontrolled and autonomous) are unidentifiable. Solving this analytically is difficult in practice for all but the simplest models.

The Taylor series approach (TS, also called power series) expands the model output $\mathbf{y}(t, \mathbf{p})$ and its derivatives in a Taylor series around $t = 0$:

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{y}(0, \mathbf{p}) + \frac{d\mathbf{y}(0, \mathbf{p})}{dt}t + \frac{d^2\mathbf{y}(0, \mathbf{p})}{dt^2} \frac{t^2}{2!} + \dots + \frac{d^i\mathbf{y}(0, \mathbf{p})}{dt^i} \frac{t^i}{i!} + \dots \quad (21)$$

The coefficients of the TS are unique. Equating the derivatives $\frac{d^i\mathbf{y}(0, \mathbf{p})}{dt^i}$ with their analytical expressions yields a set of algebraic equations that relate the unknown parameters with the output derivatives evaluated at time $t = 0$. If the parameters can be determined from these equations, they are structurally identifiable.

The Generating Series, or Volterra series coefficient approach (GS), has conceptual similarities with the TS approach. It expands the observables in series with respect to time *and inputs* with the use of Lie derivatives. The Lie derivative of a function g along another function f is defined as

$$L_f g(x) = \frac{\partial g(x)}{\partial x} f(x, u) \quad (22)$$

The GS approach requires that the model is linear in the inputs, that is,

$$\dot{\mathbf{x}}(t) = f_0(\mathbf{x}, t, \mathbf{p}) + \sum_{i=1}^{n_u} u_i(t) \cdot f_i(\mathbf{x}, t, \mathbf{p}) \quad (23)$$

The coefficients of the series are then the output functions $g(\mathbf{x}, t = 0, \mathbf{p})$ and their Lie derivatives along the functions f_i , that is, $L_{f_{j_0}} \dots L_{f_{j_k}} g(\mathbf{x}, t = 0, \mathbf{p})$. If the coefficients are unique, then the system is structurally globally identifiable. A disadvantage shared by the TS and GS approaches is that the required number of derivatives is unknown. On the other hand, the expressions obtained with GS are usually simpler than with TS. This is important since the nonlinear algebraic relations in the parameters obtained by the GS may be too challenging to solve even with state of the art symbolic software [26].

The Differential Algebra approach (DA) consists in rewriting the model in an equivalent form, which leaves out the unobservable state variables. This is done using Ritt's pseudo-division algorithm [15]. The newly obtained model is expressed in terms of the input and output variables and the parameters, called the input-output map or characteristic set. The characteristic set gives a Gröbner basis for the model equations and its derivatives. The result is a set of nonlinear algebraic equations among the parameters. By solving them, the structurally identifiable parameters and combinations of parameters can be retrieved.

The Implicit Function Theorem proposed by Xia and Moog [127] also seeks to eliminate the unobserved states from the system equations, by obtaining high order time derivatives of the observables. The procedure yields a matrix consisting of the partial derivatives of f with respect to the parameters, which must be non singular for the system to be identifiable.

The Similarity Transformation approach, initially proposed for linear systems by Walter and Lecourtier [121], is based on finding a transformation of the state variables that does not modify the input-output mapping. For the linear case, given a system $\dot{x} = A \cdot x + B \cdot u$ it seeks a similar matrix $S = P^{-1} \cdot A \cdot P$ such that $\dot{x} = P^{-1} \cdot A \cdot P \cdot x + B \cdot u$. If it is possible to find a matrix $P \neq I$ that satisfies this condition, the system is not identifiable. This idea was later extended to the nonlinear case by Vajda, Godfrey, and Rabitz [109], and compared to the TS approach in [23]. Two main drawbacks hamper the application of this method to nonlinear systems in practice: the complexity arising from the need of solving a set of partial differential equations, and the requisite that the system under analysis must be observable and controllable [73]. Observability and controllability

are defined in section 3.2.

The main conclusions of the analyses and comparisons performed in [26] is that no single method is amenable to all types of problems. Furthermore, when benchmarked on a set of small and medium-sized nonlinear models, the results varied widely from one method to another. Typically, some methods were either not applicable for a particular problem, or—if applicable—they were not capable of reaching a result due to computational errors, or—even if a result was obtained—the result was sometimes not conclusive. This variation shows that great care should be taken when selecting a particular method. This applies not only to the theoretical approach, but also to the implementation details, which can also have a large influence on the results. Regarding implementations, there are currently at least two publicly available toolboxes for computing global structural identifiability of nonlinear models: DAISY [15], which implements the differential algebra approach, and GenSSI [25], which combines the generating series approach with identifiability *tableaus*.

As can be noted from the descriptions above, the common idea of these methods is to obtain analytic relations between the model outputs and the parameters (usually by differentiation of some kind) and then establish whether these relations are unique. In general, identifiability is very difficult to assess for models with a low ratio between number of observables and number of parameters. If this is the case, one can only expect to obtain (in the best case) partial results about subsets of parameters. Such limitation appears, for example, in a medium-sized model of the circadian clock of *Arabidopsis Thaliana*, which has 7 differential equations, 29 parameters and two measured outputs [62]. This shows that much more progress is still needed in order to assess the structural identifiability of medium- to large-scale nonlinear models.

3.2 Local structural identifiability

The *sensitivity* of the model outputs—or more generally, of the model states—with respect to changes in the parameter values, which was described in section 2, provides an indication of the parameter identifiability. Clearly, if the sensitivity function of a parameter p_k is zero or close to zero, p_k is not identifiable. Furthermore, if two sensitivity functions are linearly dependent, the corresponding parameters cannot be identified either (even if the model is highly sensitive to both of them). If the sensitivity functions are *almost* linearly dependent, their parameters are highly correlated and they are very difficult to identify

in practice [13].

These conditions can be tested by means of the DSM of equation (17). The j^{th} column in the DSM contains all the output sensitivities to parameter p_j . If the DSM has linearly dependent columns, its rank will be less than the number of its columns, p . The rank of the DSM can be calculated by singular value decomposition (SVD). If all its singular values are different from zero, the DSM has full column rank and all the parameters in the model are practically identifiable. On the other hand, one or more singular values equal to zero indicate that there are unidentifiable parameters.

A method for assessing local structural identifiability that exploits this idea was recently presented by Stigter and Molenaar [97]. It combines numerical and symbolic approaches with the aim of exploiting the strengths of both, that is, the relatively fast computations allowed by the former and the rigorous results provided by the latter. Specifically, the numerical part consists of performing a SVD of the output sensitivity matrix for a set of randomly chosen parameter vectors. If a near-zero singular value is found, the parameters in the corresponding singular vector are correlated, and thus taken as candidates for constituting a non empty null space of the model. To confirm this guess, symbolic calculations in the form of Lie derivatives are then performed on the reduced model. By first using SVD to guide the search, the dimension of the problem to be analyzed symbolically is greatly reduced, thus decreasing its computational complexity. This method has been tested on models with up to ≈ 50 parameters.

A simplified version of the DSM in equation (17) was used by Li and Vu [58]. Instead of calculating each element as in equation (16), that is, $\frac{\partial y_i}{\partial p_j} = \sum_{k=1}^n \frac{\partial g_i}{\partial x_k} v_{kj} + \frac{\partial g_i}{\partial p_j}$, the elements are simply $\frac{\partial g_i}{\partial p_j}$. This amounts to discarding the part corresponding with $\sum_{k=1}^n \frac{\partial g_i}{\partial x_k} v_{kj}$; in this way the calculation of v_{kj} is avoided, which in turns eliminates the need for numerical integration of the model, since $\frac{\partial g_i}{\partial p_j}$ can be derived analytically. However, by discarding part of the mathematical expressions of the sensitivities, only a *sufficient* condition for unidentifiability (not *necessary*) can be obtained. Thus, even if the condition does not hold, the model may still be unidentifiable (conversely, the method can provide only necessary conditions for identifiability, not sufficient).

Note that calculating the rank of the DSM in equation (17) numerically entails choosing specific values for the parameter vector. Additionally, to build the DSM it is necessary to choose the set of time points in which the partial derivatives will be evaluated. Since

these points correspond to “artificial measurements”, there is no need to have experimental data nor to assume any measurement errors. Noiseless data, simulated with a particular parameter vector, can be used for this purpose. However, the requirement of specifying the number and location of measurement time points is in principle contradictory with the notion of structural identifiability, which should be analyzed only from the model structure, without depending on the experimental setup. Hence, some authors have argued [73] that this type of analysis is somehow intermediate between structural and practical identifiability, and avoid using the term “structural local identifiability” for this reason. Instead, they refer to this approach simply as “sensitivity-based identifiability analysis”. According to this criterion, the method presented by Stigter and Molenaar [97] and the Profile Likelihood approach (PL) [83] reviewed afterwards in this section would not be considered as structural identifiability methods, despite being presented as such by their authors. Acknowledging that the difference between structural and practical identifiability is not as obvious as it could seem at first sight, we choose to include the aforementioned methods in this section and encourage the reader to extract her/his own conclusions.

Instead of using the sensitivity matrix, an alternative approach is to consider structural local identifiability as a particular case of observability, where parameters are seen as state variables which satisfy $\dot{p} = 0$. Observability and controllability are dual concepts introduced by R.E. Kalman in 1960 [49]. Initially developed for linear systems, extensions for nonlinear systems were already proposed in the 1970s [31,46,47,100]. Broadly speaking, controllability can be defined as the ability of a system to be driven from an initial state to a target state in finite time and with finite inputs. Observability, on the other hand, refers to the ability to determine the states of a system from knowledge of its inputs and outputs. More formally, a system is observable at time t_1 if it is possible to determine its state $\mathbf{x}(t_1)$ from future measurements, that is, from a set of $\mathbf{y}(t)$ such that $t_1 < t < t_2$, where t_2 is a finite time. A system is called completely observable if it is observable for any state and any time.

Before turning to nonlinear systems, we provide some classical results available for linear time-invariant (LTI) systems. Consider a LTI system defined by the equations:

$$\begin{aligned} \dot{x} &= A \cdot x + B \cdot u \\ y &= C \cdot x \end{aligned} \tag{24}$$

where $u \in \mathfrak{R}^r$, $x \in \mathfrak{R}^n$, and $y \in \mathfrak{R}^m$. The linear observability matrix is defined as:

$$\mathcal{O} = \begin{pmatrix} C \\ C \cdot A \\ C \cdot A^2 \\ \vdots \\ C \cdot A^{n-1} \end{pmatrix} \quad (25)$$

The linear observability condition states that the system in equation (24) is observable if $\text{rank}(\mathcal{O}) = n$.

The linear controllability matrix is defined as:

$$C = (B|A \cdot B|A^2 \cdot B|\dots|A^{n-1} \cdot B) \quad (26)$$

The linear controllability condition states that the system is controllable if $\text{rank}(C) = n$.

The notions of observability and controllability can be generalized to nonlinear systems of the form (2)–(5) with the use of Lie algebra. The nonlinear observability matrix is

$$\mathcal{O}_{NL} = \begin{pmatrix} \frac{\partial}{\partial x} g(x) \\ \frac{\partial}{\partial x} (L_f g(x)) \\ \frac{\partial}{\partial x} (L_f^2 g(x)) \\ \vdots \\ \frac{\partial}{\partial x} (L_f^{n-1} g(x)) \end{pmatrix} \quad (27)$$

where $L_f g(x)$ is the Lie derivative of $g(x)$ along f , which was defined in equation (22) and is reproduced below for convenience:

$$L_f g(x) = \frac{\partial g(x)}{\partial x} f(x, u) \quad (28)$$

And the i^{th} Lie derivatives are defined recursively as follows:

$$\begin{aligned} L_f^2 g(x) &= \frac{\partial L_f g(x)}{\partial x} f(x, u) \\ \dots & \\ L_f^i g(x) &= \frac{\partial L_f^{i-1} g(x)}{\partial x} f(x, u) \end{aligned} \quad (29)$$

Like in the linear case, the condition for nonlinear observability is that $\text{rank}(\mathcal{O}_{NL}) = n$.

Local structural identifiability can be seen as a particular case of observability, if model parameters p are considered as state variables which satisfy $\dot{p} = 0$ [108]. Identifiability

is then assessed by calculating the rank of a generalized observability matrix, in which the state vector is augmented to include also the unknown parameters. August and Pappachristodoulou [8] used semidefinite programming to assess whether this matrix had full rank or not, by means of a sum of squares decomposition. A problem of such an approach is that the computational complexity associated with this symbolic rank calculation increases very fast with system size, making it infeasible in practice for large models. As an alternative, Sedoglavic [91] proposed a numerical algorithm which efficiently calculates this rank in the case of rational systems. Sedoglavic's algorithm was implemented in Mathematica and extended in [50] to cover more generally parametrized initial conditions. In [82] Raue *et al* compared this approach, which they referred to as the Exact Arithmetic Rank (EAR) method, with the already mentioned Differential Algebra Identifiability of Systems (DAISY) approach of Saccomani *et al* [87] and the Profile Likelihood approach (PL) [83]. PL is an *a posteriori* (i.e. data-based) approach. It evaluates identifiability by performing repeated parameter estimation runs as follows. For each parameter p_i , a set of possible values are chosen. Then p_i is fixed to the first in the set, and all the remaining parameters in the model are estimated with an optimization algorithm. Repeating this procedure for all the possible values results in a set of values of the likelihood function (i.e. the objective to be minimized in the optimization procedure). If these values are equal a flat profile is obtained, which means the parameter is not identifiable. The PL approach can be used to assess structural or practical identifiability, depending on whether the data used for calibration is simulated or experimental, respectively. It should be noted that the EAR and PL approaches assess *local* identifiability, while DAISY's results are *global*. Hence the computational cost of DAISY is larger than that of EAR and PL, and the scope of models that it can analyse is more reduced.

Finally, a related challenge is to determine the outputs that must be measured in order to guarantee local structural identifiability [3].

3.3 Finding structurally identifiable subsets of parameters and reparameterizations

The result of an identifiability analysis is often that a model is unidentifiable. Note, however, that for a model to be considered as unidentifiable it is sufficient that only one of its parameters is unidentifiable. Obviously, much more information is provided by determining *which* parameters are identifiable and which are not. Most of the aforementioned

methods for structural identifiability are capable of reporting (albeit with certain limitations) which parameters of an unidentifiable model are identifiable. Once the subset of identifiable parameters has been determined, a new question naturally arises: given the remaining set of unidentifiable parameters, is it possible to combine them in order to find one or several identifiable reparameterizations? This task is more difficult than “simply” determining the identifiability of a set of parameters, and only a few of the aforementioned methods are also capable of finding identifiable combinations of parameters. The similarity transformation approach inspired a procedure for finding *locally* identifiable reparameterizations [24] which was extended using Taylor series [41]. This approach was also adopted more recently by Stigter and Molenaar [97]. In a series of publications [70, 71], DiStefano and coworkers extended the differential algebra approach [87] to find parameter combinations (‘combos’) which are globally identifiable, using Gröbner bases to determine the simplest set of identifiable parameter combinations. Notably, the method is capable in principle of determining whether a parameter is *globally* or *locally* identifiable, and how many local solutions exist. A web-based application that implements this approach [72] can be accessed at <http://biocyb1.cs.ucla.edu/combos/>. However, limitations in model size are very strict for this task, which can typically be achieved for models with only a few (ten or less) parameters.

4 Practical identifiability

Assessing the structural identifiability of a model is only part of the general identifiability problem. Even in the most favourable case (that a model is structurally identifiable), it may not be possible to determine its parameter values in practice. The reason is that structural identifiability analyses assume that unlimited, noiseless data are available for parameter estimation; however, in practice data are always noisy and limited.

Practical identifiability is about quantifying the uncertainty in the estimated parameter values and calculating their confidence intervals, taking into account not only the model structure but also the information contained in the available data. To avoid confusion, it should be noticed that practical identifiability has also been called numerical identifiability [38], quantitative identifiability [111], or *a posteriori* identifiability [85]. Other authors have called it estimability [48, 67], reserving the term “identifiability” for structural identifiability.

For an estimated parameter vector $\hat{\mathbf{p}}$ the covariance matrix provides information about the variability, both for individual parameters and for pairs. It is defined as:

$$\text{COV} = E \left[(\hat{\mathbf{p}} - \bar{\mathbf{p}}) (\hat{\mathbf{p}} - \bar{\mathbf{p}})^T \right] \begin{bmatrix} \sigma^2(\hat{p}_1) & \cdots & \text{cov}(\hat{p}_1\hat{p}_p) \\ \vdots & \ddots & \vdots \\ \text{cov}(\hat{p}_p\hat{p}_1) & \cdots & \sigma^2(\hat{p}_p) \end{bmatrix} \quad (30)$$

where: $\bar{\mathbf{p}}$ is the mean value; $E(\cdot)$ is the expected value, that is, $E(\hat{\mathbf{p}}) = \bar{\mathbf{p}}$; $\sigma^2(\hat{p}_i)$ is the estimated variance; and $\text{cov}(\hat{p}_i\hat{p}_p)$ is the estimated covariance. The correlation coefficients between pairs of parameters can be calculated as:

$$\text{corr}(\hat{p}_i\hat{p}_j) = \frac{\text{cov}(\hat{p}_i\hat{p}_j)}{\sqrt{\sigma^2(\hat{p}_i) \cdot \sigma^2(\hat{p}_j)}} \quad (31)$$

The Hessian matrix, $\hat{H}(\hat{\mathbf{p}}) = \left(\frac{\partial^2 \mathbf{y}}{\partial \mathbf{p}^2} \right)_{\mathbf{p}=\hat{\mathbf{p}}}$, is asymptotically equal to the covariance matrix,

$$\text{COV}(\hat{\mathbf{p}}) \approx \hat{H}(\hat{\mathbf{p}}) \quad (32)$$

Hence the covariance matrix can be approximated from estimations of the Hessian.

The covariance matrix (COV) and the Fisher information matrix (FIM) are closely related. The FIM measures the amount of information contained in the experimental data. It can be written as a function of the sensitivity functions as

$$\text{FIM}(\mathbf{p}) = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{p}} \right)^T W \left(\frac{\partial \mathbf{y}}{\partial \mathbf{p}} \right) \quad (33)$$

where W is a data weighting matrix. Typically, weights should be chosen as the inverse of the error variance for each experimental data point.

For a set of d measurements, the FIM can be calculated as

$$\text{FIM}(\mathbf{p}) = \sum_{k=1}^d \left(\frac{\partial \mathbf{y}(t_k, \mathbf{p})}{\partial p_i} \right) W \left(\frac{\partial \mathbf{y}(t_k, \mathbf{p})}{\partial p_j} \right)^T \quad (34)$$

The Cramér-Rao theorem [33] states that, if $\hat{\mathbf{p}}$ is an unbiased estimate of \mathbf{p} (i.e. $E(\hat{\mathbf{p}}) = \bar{\mathbf{p}}$), the inverse of the FIM provides a lower bound estimate for the COV,

$$\text{COV}(\hat{\mathbf{p}}) \geq \text{FIM}^{-1}(\hat{\mathbf{p}}) \quad (35)$$

Thus, we see how the correlations between parameters can be derived from sensitivity-based criteria like the Fisher Information Matrix (FIM). Like the DSM, the FIM provides information about parameter identifiability. Since the FIM includes information about

experimental measures (variances), it can be used for *practical* identifiability analysis. If the FIM is singular, there are unidentifiable parameters; if it is near-singular, there are highly correlated parameters which will be very difficult or impossible to identify in practice.

Several scalar measures have been proposed to summarize the information contained in the FIM, in order to be used as criteria in optimal experimental design or subset selection [113]. Two commonly used ones are the D and E criteria, which are defined as follows:

$$\text{D-criterion} = \max(\det(\text{FIM})) \tag{36}$$

$$\text{E-criterion} = \max(\lambda_{\min}(\text{FIM}^{-1})) \tag{37}$$

where λ_{\min} is the minimum eigenvalue of the FIM. The D-criterion minimizes the volume of the confidence ellipsoids, and therefore the geometric mean of the errors in the parameters. The E-criterion minimizes the largest error.

The FIM is easy to calculate. However, since the sensitivity analysis based on FIM relies on a linearization of the model, it may be misleading if strong nonlinearities are present. As a consequence, confidence intervals estimated from the FIM can be overly optimistic. An alternative is to use other computational approaches such as jackknife or bootstrap [13]. These are much more expensive computationally speaking: the bootstrap approach involves solving the parameter estimation problem a large number of times (usually in the hundreds or thousands), starting from different initial solutions. It should be noted that even a single parameter estimation can require very high computation times in realistic systems biology applications (typically at least several hours for dynamic nonlinear models of medium or large-scale size [118]).

A concept related to practical identifiability is parameter sloppiness, which was introduced by Sethna and coworkers [19, 44, 65]. The basic idea is that the output of so-called sloppy models is mostly determined by a reduced subset of parameters, or of combinations of parameters, while the remaining parameter directions have little or no influence—and therefore cannot be estimated from data. In [44] a total of 17 systems biology models were analysed, concluding that sloppiness is a universal property. However, as other authors have shown [5], sloppiness can be avoided (at least in some cases) by obtaining new

data from rationally designed experiments. Thus, sloppiness does not have to be taken for granted in dynamic biological models. On the theoretical side, it has also been argued [115] that the concept of sloppiness is strongly related to identifiability (despite the fact that this relation was not acknowledged nor investigated in [44]). Indeed, it seems to refer, somehow ambiguously, to both structural and practical identifiability.

4.1 Finding practically identifiable subsets of parameters

Even when a model is structurally identifiable, it may be practically unidentifiable: the parameter estimation problem may be ill-conditioned, meaning that its optimal solution will be highly sensitive to variations in the data. This may be due to an excess of parameters or lack of sufficiently informative experimental data. Such an overparameterized model can lead to overfitting when trying to estimate the parameter values [45]. The ill-conditioned nature of the PE can be remedied with the use of regularization techniques [43], modifying the objective function of the associated optimization problem in order to make its minimum more robust. Subset selection can be seen as a way of regularizing the problem adopting a model reduction perspective [53], and it is particularly interesting because it maintains model interpretability. It consists of finding the subset(s) of parameters which are identifiable in practice, that is, taking into account the limitations caused by the quantity and quality of available data. From a practical identifiability point of view, there will often be several possible subsets of identifiable parameters.

A number of approaches for subset selection have been presented. They usually involve calculations of the model's output sensitivities to parameter values [53, 84]. Two aspects related to parameter sensitivities must be taken into account: their magnitude (i.e. a parameter cannot be identified if the model output is hardly sensitive to it) and correlation (i.e. two or more parameters cannot be estimated if their effects can be mutually compensated). Thus, a general sequential procedure for subset selection can be as follows:

1. Rank the parameters according to their sensitivities, e.g. using one of the measures in eq. (18).
2. Select the parameter with the highest sensitivity.
3. From the remaining parameters, choose the one that:

- (a) is least correlated with the parameter(s) already included in the set
 - (b) has the largest influence on the model output
4. Repeat step 3 until a stopping criterion (typically based on a threshold on the dependence of parameters) is met.

Several variations of this procedure have been proposed, differing on the specific choices made in the implementation of steps 3 and 4. For example, orthogonalization approaches do it by projecting the sensitivity vectors of the unselected parameters to an orthogonal space of the already selected parameters [63]. Alternatively, it has been proposed to add parameters based on the collinearity index, a measure of “compensability” introduced in [20], which is defined as the minimum achievable norm of a linear combination of the sensitivity functions with normalized coefficients,

$$\gamma(\mathbf{p}) = \frac{1}{\min_{\|\beta\|=1} \|\tilde{s}_1\beta_1 + \dots + \tilde{s}_m\beta_m\|} \quad (38)$$

where \tilde{s}_i is the normalized sensitivity of parameter i .

Other possibilities for choosing the new parameters in the subset include adding them based on the D or E criteria of equations (36), (37) [28, 64], using principal component analysis [59], combining subset selection with the profile likelihood approach [39], or using hierarchical parameter clustering [29], possibly combined with mutual information [76].

It is important to realize that sequential procedures do not guarantee finding the “best” subset, that is, the one containing the most sensitive and least correlated parameters. An alternative is to use optimization methods [28], which entails a higher computational cost.

5 Related Challenges in Modelling

Identifiability analysis is not an isolated task, but part of the broader process of model building. Model building can be seen as a cycle, as shown in Figure 3, consisting of several procedures that should ideally be performed iteratively until obtaining a satisfactory model [10]. Of course, it will not always be necessary to perform all the steps shown in Figure 3; for example, often the structure of the system (i.e. the set of chemical reactions, including stoichiometry and type of kinetics) will be assumed known. In such scenario we can directly formulate the differential equations that define the model dynamics, and pro-

ceed to analyse its structural identifiability with the techniques reviewed in the preceding sections.

After identifiability analysis has determined which parameters in a model are identifiable, the next step is naturally identifying them. *Parameter Estimation* (PE), also called model calibration, is the task of determining the values of unknown parameters from experimental data [6, 125]. This is carried out by finding the parameter values that minimize a measure of the difference between the model output and the measurements. Typical measures are the (weighted) least squares (LS), maximum likelihood (ML), and Bayesian estimators [60], in increasing order of the amount of information required to calculate them. The LS estimator consists simply of a (possibly weighted) sum of squared differences between model outputs and experimental measures. The ML estimator uses the probability distribution of the parameters, and Bayesian measures need additionally the conditional probability distribution of the measurements. Minimization of the error is achieved by an optimization procedure, where one of these measures is selected as the objective function that must be minimized. For the large, nonlinear models on which this paper focuses, the search space of the objective function will usually be also large, non-convex, and will have several minima, specially if noisy data are used for calibration (which is in practice unavoidable) [115]. Hence the methods that will be more likely to provide good solutions in reasonable computation times will generally be non-deterministic global optimization algorithms [40, 74], such as metaheuristics, which may be combined with local deterministic methods if it is suspected that the search is in the vicinity of a minimum [98].

Often the available data is not sufficient to calibrate the model properly. If this is the case, new data should be collected if possible, by designing and performing new experiments. *Optimal Experiment Design* (OED) deals with defining new experiments which will produce data with the maximum possible information content [13, 99]. It is tightly related with practical identifiability analysis, since OED can be used to reduce the uncertainty in the estimated parameter values, obtaining narrower confidence intervals for them. Optimization techniques can be used to design experimental setups by minimizing a objective function that represents some measure of the uncertainty in the parameters [11]. The formulation of this objective will depend on the purpose of the experiment. OED can be used not only with the aim of improving parameter estimates, but also

to characterize new or specific parts of a model, or to discriminate between competing modelling structures [69, 93, 112].

Until now we have always considered the problem of identifying a given model, but if instead of a single model structure there are a number of candidate models that compete for describing a system, a question naturally arises: can we select the correct (or best) model, and if so, how? *Model selection*, or discrimination, is a non-trivial matter, since the selected model should not necessarily be the one that achieves the best fit to the data. Other criteria, such as biochemical interpretation and degree of complexity (measured e.g. by the number of parameters) should be taken into account; to this end, statistical measures such as the Akaike or Bayesian information criteria can be used to quantify the degree of complexity of the model [17, 119]. The model choice should reflect a balance between these factors [80].

A prerequisite to choose among different models is *distinguishability*, a concept that addresses this question: can two different models produce the same output for any allowed input? Although they are sometimes mistaken, distinguishability and identifiability are different concepts [123, 124]: identifiability is a property of a single model, while distinguishability refers to several models. As an example, assume that a given model $M_1(\mathbf{x}, \mathbf{p}, \mathbf{u}, t)$ (where we have made explicit the dependence on the states, parameters, inputs, and time) is identifiable. As we have seen, this means that, assuming that its structure is known, we can uniquely determine the values of its parameters \mathbf{p} by observing their inputs \mathbf{u} and outputs \mathbf{y} . Now imagine that we have a second model, $M_2(\mathbf{x}, \mathbf{p}, \mathbf{u}, t)$, which has a different internal structure than M_1 but the same input-output behaviour. In this case, although M_1 is identifiable, it is not distinguishable from M_2 . In general, two models may be identifiable but their outputs can be indistinguishable (and clearly the opposite also holds). The identifiability of two structures is neither necessary nor sufficient for their distinguishability. That said, although both concepts are different, they are clearly related, and similar approaches as those used for determining structural identifiability can be applied to distinguishability [42, 81, 110]. For the specific case of chemical reaction networks, several methods for finding alternative structures have been presented in recent years by Szederkényi and coworkers [86, 101–106].

Finally, before concluding this paper we cannot avoid a pending question: if, after all, we cannot escape from having an unidentifiable model, can we still use it to make

predictions? At first sight, it may appear that we cannot trust predictions made by unidentifiable models. Indeed, when calibrating a model with identifiability issues it is easy to incur in overfitting. In that case, although the model manages to fit the available data very well, it is being trained to fit the noise present in the data instead of the true underlying dynamics. This is due to the model being more complex than it is necessary to describe the system under study [45]. An undesired consequence is that the model predictions in different conditions may be very different from the actual system outputs. Thus, we may ask: to which extent does the uncertainty in parameter values translate into *uncertainty in the model predictions*? Is it possible to find bounds for this uncertainty and, even better, reduce it?

A possible approach to take prediction uncertainty into account is to use an ensemble of plausible models instead of a single one. In this way, instead of predicting a single point, or a single time-series, it is possible to generate a cloud of such predictions. Ensemble modelling has been applied to genetic [51], cell signalling [54], and metabolic networks [107]. Bever [16] showed that the consensus between models in the ensemble may be used to find high-confidence predictions, and recently such consensus has been shown to be correlated with prediction quality in biochemical models [114, 117]. A step further is represented by so-called core predictions, a concept introduced by Cedersund [21] to denote well-determined predictions, which can be obtained even from unidentifiable models. The idea can be seen as a further refinement of the cloud of predictions made by ensemble models, by finding the boundaries (extreme values) of the point-cloud generated with all the possibly acceptable parameters. Different variants of this approach, including connections with the profile likelihood methodology, were discussed in [22].

In conclusion, in this paper we hope to have conveyed the idea that there is a wealth of mathematical and computational techniques that can provide invaluable help in the task of modelling complex biochemical systems. When used appropriately, they can aid in identifying possible sources of uncertainty and remedying them, eventually leading to accurate and useful representations of the system under study.

Acknowledgment: This work was funded by the Galician government (Xunta de Galicia) through the I2C postdoctoral program (fellowship ED481B2014/133-0), and by the Spanish Ministry of Economy and Competitiveness (grant DPI2013-47100-C2-2-P).

References

- [1] J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, M. Jirstrand, Kinetic models in industrial biotechnology—improving cell factory performance, *Metab. Eng.* **24** (2014) 38–60.
- [2] U. Alon, Biological networks: the tinkerer as an engineer, *Science* **301** (2003) 1866–1867.
- [3] M. Anguelova, J. Karlsson, M. Jirstrand, Minimal output sets for identifiability, *Math. Biosci.* **239** (2012) 139–153.
- [4] P. M. Antony, R. Balling, N. Vlassis, From systems biology to systems biomedicine, *Curr. Opin. Biotechnol.* **23** (2012) 604–608.
- [5] J. F. Apgar, D. K. Witmer, F. M. White, B. Tidor, Sloppy models, parameter uncertainty, and the role of experimental design, *Mol. Biosyst.* **6** (2010) 1890–1900.
- [6] M. Ashyraliyev, Y. Fomekong-Nanfack, J. Kaandorp, J. Blom, Systems biology: parameter estimation for biochemical models, *FEBS J.* **276** (2008) 886–902.
- [7] S. Audoly, G. Bellu, L. D’Angiό, M. P. Saccomani, C. Cobelli, Global identifiability of nonlinear models of biological systems, *IEEE. Trans. Biomed. Eng.* **48** (2001) 55–65.
- [8] E. August, A. Papachristodoulou, A new computational tool for establishing model parameter identifiability, *J. Comput. Biol.* **16** (2009) 875–885.
- [9] E. Balsa-Canto, Advanced model identification using global optimization, at http://gingproc.iim.csic.es/~amigo/ICSB2010_IntroTalk_AMIGO_toolbox.pdf.
- [10] E. Balsa-Canto, A. Alonso, J. R. Banga, An iterative identification procedure for dynamic modeling of biochemical networks, *BMC Syst. Biol.* **4** (2010) 1–18.
- [11] E. Balsa-Canto, A. A. Alonso, J. R. Banga, Computational procedures for optimal experimental design in biological systems, *IET Syst. Biol.* **2** (2008) 163–172.
- [12] E. Balsa-Canto, J. R. Banga, AMIGO, a toolbox for advanced model identification in systems biology using global optimization, *Bioinformatics* **27** (2011) 2311–2313.
- [13] J. Banga, E. Balsa-Canto, Parameter estimation and optimal experimental design, *Essays Biochem.* **45** (2008) 195–209.
- [14] R. Bellman, K. J. Åström, On structural identifiability, *Math. Biosci.* **7** (1970) 329–339.

- [15] G. Bellu, M. P. Saccomani, S. Audoly, L. D'Angio, DAISY: a new software tool to test global identifiability of biological and physiological systems, *Comput. Methods Programs Biomed.* **88** (2007) 52–61.
- [16] C. Bever, *Selecting high-confidence predictions from ordinary differential equation models of biological networks*, PhD thesis, MIT, (2008).
- [17] H. Bozdogan, Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika* **52** (1987) 345–370.
- [18] I. M. Brockman, K. L. Prather, Dynamic metabolic engineering: New strategies for developing responsive cell factories, *Biotechnol. J.* **10** (2015) 1360–1369.
- [19] K. S. Brown, J. P. Sethna, Statistical mechanical approaches to models with many poorly known parameters, *Phys. Rev. E* **68** (2003) 021904.
- [20] R. Brun, P. Reichert, H. R. Künsch, Practical identifiability analysis of large environmental simulation models, *Water Resour. Res.* **37** (2000) 1015–1030.
- [21] G. Cedersund, Conclusions via unique predictions obtained despite unidentifiability – new definitions and a general method, *FEBS J.* **279** (2012) 3513–3527.
- [22] G. Cedersund, Prediction uncertainty estimation despite unidentifiability: an overview of recent developments, in: L. Geris, D. Gomez-Cabrero (Eds.), *Uncertainty in Biology*, Springer, 2016, pp. 449–466.
- [23] M. J. Chappell, K. R. Godfrey, S. Vajda, Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods, *Math. Biosci.* **102** (1990) 41–73.
- [24] M. J. Chappell, R. N. Gunn, A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach, *Math. Biosci.* **148** (1998) 21–41.
- [25] O. Chiş, J. R. Banga, E. Balsa-Canto, GenSSI: a software toolbox for structural identifiability analysis of biological models, *Bioinformatics* **27** (2011) 2610–2611.
- [26] O. T. Chiş, J. R. Banga, E. Balsa-Canto, Structural identifiability of systems biology models: a critical comparison of methods, *PLoS One* **6** (2011) e27755.
- [27] A. Chowdhury, A. Khodayari, C. D. Maranas, Improving prediction fidelity of cellular metabolism with kinetic descriptions, *Curr. Opin. Biotechnol.* **36** (2016) 57–64.
- [28] Y. Chu, J. Hahn, Parameter set selection for estimation of nonlinear dynamic systems, *AIChE J.* **53** (2007) 2858–2870.

- [29] Y. Chu, J. Hahn, Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters, *Ind. Eng. Chem. Res.* **48** (2008) 6000–6009.
- [30] C. Cobelli, J. DiStefano, Parameter and structural identifiability concepts and ambiguities: a critical review and analysis, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **239** (1980) R7–R24.
- [31] C. Cobelli, G. Romanin-Jacur, Controllability, observability and structural identifiability of multi input and multi output biological compartmental systems, *IEEE Trans. Biomed. Eng.* **2** (1976) 93–100.
- [32] G. Craciun, C. Pantea, Identifiability of chemical reaction networks, *J. Math. Chem.* **44** (2008) 244–259.
- [33] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, 1999.
- [34] M. E. Csete, J. C. Doyle, Reverse engineering of biological complexity, *Science* **295** (2002) 1664–1669.
- [35] R. Cukier, C. Fortuin, K. E. Shuler, A. Petschek, J. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory, *J. Chem. Phys.* **59** (1973) 3873–3878.
- [36] F. P. Davidescu, S. B. Jørgensen, Structural parameter identifiability analysis for dynamic reaction networks, *Chem. Eng. Sci.* **63** (2008) 4754–4762.
- [37] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, Some effective approaches to check the identifiability of uncontrolled nonlinear systems, *Math. Comput. Simul.* **57** (2000) 35–44.
- [38] J. DiStefano, *Dynamic Systems Biology Modeling and Simulation*, Academic Press, Waltham, 2016.
- [39] M. C. Eisenberg, M. A. Hayashi, Determining identifiable parameter combinations using subset profiling, *Math. Biosci.* **256** (2014) 116–126.
- [40] W. R. Esposito, C. A. Floudas, Global optimization for the parameter estimation of differential-algebraic systems, *Ind. Eng. Chem. Res.* **39** (2000) 1291–1310.
- [41] N. D. Evans, M. J. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, *Math. Biosci.* **168** (2000) 137–159.

- [42] N. D. Evans, M. J. Chappell, M. J. Chapman, K. Godfrey, Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems, *Automatica* **40** (2004) 1947–1953.
- [43] A. Gábor, J. R. Banga, Robust and efficient parameter estimation in dynamic models of biological systems, *BMC Syst. Biol.* **9** (2015) #74.
- [44] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models, *PLoS Comput. Biol.* **3** (2007) e189.
- [45] D. M. Hawkins, The problem of overfitting, *J. Chem. Inf. Comput. Sci.* **44** (2004) 1–12.
- [46] G. Haynes, H. Hermes, Nonlinear controllability via Lie theory, *SIAM J. Control* **8** (1970) 450–460.
- [47] R. Hermann, A. J. Krener, Nonlinear controllability and observability, *IEEE Trans. Autom. Control* **22** (1977) 728–740.
- [48] J. A. Jacquez, P. Greif, Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design, *Math. Biosci.* **77** (1985) 201–227.
- [49] R. E. Kalman, Contributions to the theory of optimal control, *Bol. Soc. Mat. Mexicana* **5** (1960) 102–119.
- [50] J. Karlsson, M. Anguelova, M. Jirstrand, An efficient method for structural identifiability analysis of large dynamic systems, in: M. Kinnaert (Ed.), *16th Symposium on System Identification*, IFAC, 2012, pp. 941–946.
- [51] S. Kauffman, A proposal for using the ensemble approach to understand genetic regulatory networks, *J. Theor. Biol.* **230** (2004) 581–590.
- [52] P. Kirk, D. Silk, M. P. Stumpf, Reverse engineering under uncertainty, in: L. Geris, D. Gomez-Cabrero (Eds.), *Uncertainty in Biology*, Springer, 2016, pp. 15–32.
- [53] C. Kravaris, J. Hahn, Y. Chu, Advances and selected recent developments in state and parameter estimation, *Comput. Chem. Eng.* **51** (2013) 111–123.
- [54] L. Kuepfer, M. Peter, U. Sauer, J. Stelling, Ensemble modeling for analysis of cell signaling dynamics, *Nat. Biotechnol.* **25** (2007) 1001–1006.
- [55] J. R. Leis, M. A. Kramer, Sensitivity analysis of systems of differential and algebraic equations, *Comput. Chem. Eng.* **9** (1985) 93–96.

- [56] J. R. Leis, M. A. Kramer, Algorithm 658: ODESSA – an ordinary differential equation solver with explicit simultaneous sensitivity analysis, *ACM TOMS* **14** (1988) 61–67.
- [57] J. R. Leis, M. A. Kramer, The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations, *ACM TOMS* **14** (1988) 45–60.
- [58] P. Li, Q. D. Vu, Identification of parameter correlations for parameter estimation in dynamic biological models, *BMC Syst. Biol.* **7** (2013) #91.
- [59] R. Li, M. Henson, M. J. Kurtz, Selection of model parameters for off-line parameter estimation, *IEEE Trans. Contr. Syst. Technol.* **12** (2004) 402–412.
- [60] L. Ljung, *System Identification – Theory for the User*, Prentice Hall, Upper Saddle River, 1999.
- [61] L. Ljung, T. Glad, On global identifiability for arbitrary model parameterizations, *Automatica* **30** (1994) 265–276.
- [62] J. Locke, A. Millar, M. Turner, Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*, *J. Theor. Biol.* **234** (2005) 383–393.
- [63] B. F. Lund, B. A. Foss, Parameter ranking by orthogonalization–applied to nonlinear mechanistic models, *Automatica* **44** (2008) 278–281.
- [64] V. C. Machado, G. Tapia, D. Gabriel, J. Lafuente, J. A. Baeza, Systematic identifiability study based on the fisher information matrix for reducing the number of parameters calibration of an activated sludge model, *Environ. Modell. Softw.* **24** (2009) 1274–1284.
- [65] B. B. Machta, R. Chachra, M. K. Transtrum, J. P. Sethna, Parameter space compression underlies emergent theories and predictive models, *Science* **342** (2013) 604–607.
- [66] M. D. McKay, R. J. Beckman, W. J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* **21** (1979) 239–245.
- [67] K. A. McLean, K. B. McAuley, Mathematical modelling of chemical processes – obtaining the best model predictions and parameter estimates using identifiability and estimability procedures, *Can. J. Chem. Eng.* **90** (2012) 351–366.

- [68] M. Á. Medina, Systems biology for molecular life sciences and its impact in biomedicine, *Cell. Mol. Life Sci.* **70** (2013) 1035–1053.
- [69] B. Mélykúti, E. August, A. Papachristodoulou, H. El-Samad, Discriminating between rival biochemical network models: three approaches to optimal experiment design, *BMC Syst. Biol.* **4** (2010) #38.
- [70] N. Meshkat, C. Anderson, J. J. DiStefano, Finding identifiable parameter combinations in nonlinear ode models and the rational reparameterization of their input–output equations, *Math. Biosci.* **233** (2011) 19–31.
- [71] N. Meshkat, M. Eisenberg, J. J. DiStefano, An algorithm for finding globally identifiable parameter combinations of nonlinear ode models using Gröbner bases, *Math. Biosci.* **222** (2009) 61–72.
- [72] N. Meshkat, C. E. Z. Kuo, J. DiStefano, On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: A novel web implementation, *PLoS One* **9** (2014) #e110261.
- [73] H. Miao, X. Xia, A. S. Perelson, H. Wu, On identifiability of nonlinear ode models and applications in viral dynamics, *SIAM Rev.* **53** (2011) 3–39.
- [74] C. Moles, P. Mendes, J. Banga, Parameter estimation in biochemical pathways: a comparison of global optimization methods, *Genome Res.* **13** (2003) #2467.
- [75] M. D. Morris, Factorial sampling plans for preliminary computational experiments, *Technometrics* **33** (1991) 161–174.
- [76] K. Nienaltowski, M. Włodarczyk, T. Lipniacki, M. Komorowski, Clustering reveals limits of parameter identifiability in multi–parameter models of biochemical dynamics, *BMC Syst. Biol.* **9** (2016) #65.
- [77] R. P. Nolan, K. Lee, Dynamic model of CHO cell metabolism, *Metab. Eng.* **13** (2011) 108–124.
- [78] R. P. Nolan, K. Lee, Dynamic model for CHO cell engineering, *J. Biotechnol.* **158** (2012) 24–33.
- [79] H. Pohjanpalo, System identifiability based on the power series expansion of the solution, *Math. Biosci.* **41** (1978) 21–33.
- [80] D. Posada, T. R. Buckley, Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests, *Syst. Biol.* **53** (2004) 793–808.

- [81] A. Raksanyi, Y. Lecourtier, E. Walter, A. Venot, Identifiability and distinguishability testing via computer algebra, *Math. Biosci.* **77** (1985) 245–266.
- [82] A. Raue, J. Karlsson, M. P. Saccomani, M. Jirstrand, J. Timmer, Comparison of approaches for parameter identifiability analysis of biological systems, *Bioinformatics* **30** (2014) 1440–1448.
- [83] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, J. Timmer, Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, *Bioinformatics* **25** (2009) 1923–1929.
- [84] P. Reichert, P. Vanrolleghem, Identifiability and uncertainty analysis of the river water quality model no. 1 (RWQM1), *Water Sci. Technol.* **43** (2000) 329–338.
- [85] M. Rodríguez-Fernández, J. A. Egea, J. R. Banga, Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems, *BMC Bioinf.* **7** (2006) #483.
- [86] J. Rudan, G. Szederkényi, K. Hangos, Efficient computation of alternative structures for large kinetic systems using linear programming, *MATCH Commun. Math. Comput. Chem.* **71** (2014) 71–92.
- [87] M. P. Saccomani, S. Audoly, L. D’Angiò, Parameter identifiability of nonlinear systems: the role of initial conditions, *Automatica* **39** (2003) 619–632.
- [88] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global Sensitivity Analysis – The Primer*, Wiley, Oxford, 2008.
- [89] A. Saltelli, S. Tarantola, F. Campolongo, Sensitivity analysis as an ingredient of modeling, *Statist. Sci.* **15** (2000) 377–395.
- [90] A. Saltelli, S. Tarantola, K. S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics* **41** (1999) 39–56.
- [91] A. Sedoglavic, A probabilistic algorithm to test local algebraic observability in polynomial time, *J. Symb. Comput.* **33** (2002) 735–755.
- [92] R. Serban, A. C. Hindmarsh, CVODES: the sensitivity-enabled ODE solver in SUNDIALS, in: *Proc. of IDETC/CIE 2005*, ASME, 2005, pp. 257–269.
- [93] D. Skanda, D. Lebiedz, An optimal experimental design approach to model discrimination in dynamic biochemical systems, *Bioinformatics* **26** (2010) 939–945.
- [94] K. Smallbone, P. Mendes, Large-scale metabolic models: From reconstruction to differential equations, *Ind. Biotech.* **9** (2013) 179–184.

- [95] I. M. Sobol, On sensitivity estimation for nonlinear mathematical models, *Mat. Model.* **2** (1990) 112–118.
- [96] H. S. Song, F. DeVilbiss, D. Ramkrishna, Modeling metabolic systems: the need for dynamics, *Curr. Opin. Chem. Eng.* **2** (2013) 373–382.
- [97] J. D. Stigter, J. Molenaar, A fast algorithm to assess local structural identifiability, *Automatica* **58** (2016) 118–124.
- [98] J. Sun, J. M. Garibaldi, C. Hodgman, Parameter estimation using metaheuristics in systems biology: a comprehensive review, *IEEE/ACM Trans. Comput. Biol. Bioinform.* **9** (2012) 185–202.
- [99] N. Z. Sun, A. Sun, Optimal experimental design, in: N. Z. Sun, A. Sun (Eds.), *Model Calibration and Parameter Estimation*, Springer, 2016, pp. 459–507.
- [100] H. J. Sussmann, V. Djurdjevic, Controllability of nonlinear systems, *J. Differ. Eqs.* **12** (1972) 95–116.
- [101] G. Szederkényi, Computing sparse and dense realizations of reaction kinetic systems, *J. Math. Chem.* **47** (2010) 551–568.
- [102] G. Szederkényi, J. Banga, A. Alonso, Inference of complex biological networks: distinguishability issues and optimization-based solutions, *BMC Syst. Biol.* **5** (2011) #177.
- [103] G. Szederkényi, J. R. Banga, A. A. Alonso, Crnreals: a toolbox for distinguishability and identifiability analysis of biochemical reaction networks, *Bioinformatics* **28** (2012) 1549–1550.
- [104] G. Szederkényi, K. M. Hangos, Finding complex balanced and detailed balanced realizations of chemical reaction networks, *J. Math. Chem.* **49** (2011) 1163–1179.
- [105] G. Szederkényi, K. M. Hangos, T. Péni, Maximal and minimal realizations of reaction kinetic systems: computation and properties, *MATCH Commun. Math. Comput. Chem.* **65** (2011) 309–332.
- [106] G. Szederkényi, K. M. Hangos, Z. Tuza, Finding weakly reversible realizations of chemical reaction networks using optimization, *MATCH Commun. Math. Comput. Chem.* **67** (2012) 193–212.
- [107] L. Tran, M. Rizk, J. Liao, Ensemble modeling of metabolic networks, *Biophys. J.* **95** (2008) 5606–5617.

- [108] E. T. Tunali, T.-J. Tarn, New results for identifiability of nonlinear systems, *IEEE Trans. Autom. Control* **32** (1987) 146–154.
- [109] S. Vajda, K. R. Godfrey, H. Rabitz, Similarity transformation approach to identifiability analysis of nonlinear compartmental models, *Math. Biosci.* **93** (1989) 217–248.
- [110] S. Vajda, H. Rabitz, Identifiability and distinguishability of general reaction systems, *J. Phys. Chem.* **98** (1994) 5265–5271.
- [111] S. Vajda, H. Rabitz, E. Walter, Y. Lecourtier, Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models, *Chem. Eng. Commun.* **83** (1989) 191–219.
- [112] J. Vanlier, C. A. Tiemann, P. A. Hilbers, N. A. van Riel, Optimal experiment design for model selection in biochemical networks, *BMC Syst. Biol.* **8** (2014) #20.
- [113] P. Vanrolleghem, D. Dochain, Bioprocess model identification, in: J. F. van Impe, P. A. Vanrolleghem, D. M. Iserentant (Eds.), *Advanced Instrumentation, Data Interpretation, and Control of Biotechnological Processes*, Springer, 1998, pp. 251–318.
- [114] A. Villaverde, S. Bongard, J. Schmid, D. Müller, K. Mauch, E. Balsa-Canto, J. Banga, High-confidence predictions in systems biology dynamic models, *Adv. Intell. Soft-Computing* **294** (2014) 161–171.
- [115] A. F. Villaverde, J. R. Banga, Reverse engineering and identification in systems biology: strategies, perspectives and challenges, *J. R. Soc. Interface* **11** (2014) #20130505.
- [116] A. F. Villaverde, S. Bongard, K. Mauch, E. Balsa-Canto, J. R. Banga, Metabolic engineering with multi-objective optimization of kinetic models, *J. Biotechnol.*, in press.
- [117] A. F. Villaverde, S. Bongard, K. Mauch, D. Müller, E. Balsa-Canto, J. Schmid, J. R. Banga, A consensus approach for estimating the predictive accuracy of dynamic models in biology, *Comput. Methods Programs Biomed.* **119** (2015) 17–28.
- [118] A. F. Villaverde, D. Henriques, K. Smallbone, S. Bongard, J. Schmid, D. Cicin-Sain, A. Crombach, J. Saez-Rodriguez, K. Mauch, E. Balsa-Canto, P. Mendes, J. Jaeger, J. R. Banga, Biopredyn-bench: a suite of benchmark problems for dynamic modelling in systems biology, *BMC Syst. Biol.* **9** (2015) #8.
- [119] V. Vyshemirsky, M. A. Girolami, Bayesian ranking of biochemical system models, *Bioinformatics* **24** (2008) 833–839.

- [120] E. Walter, I. Braems, L. Jaulin, M. Kieffer, Guaranteed numerical computation as an alternative to computer algebra for testing models for identifiability, in: R. Alt, A. Frommer, R. B. Kearfott, W. Luther (Eds.), *Numerical Software with Result Verification*, Springer, 2004, pp. 124–131.
- [121] E. Walter, Y. Lecourtier, Unidentifiable compartmental models: what to do? *Math. Biosci.* **56** (1981) 1–25.
- [122] E. Walter, Y. Lecourtier, Global approaches to identifiability testing for linear and nonlinear state space models, *Math. Comput. Simul.* **24** (1982) 472–482.
- [123] E. Walter, Y. Lecourtier, J. Happel, On the structural output distinguishability of parametric models, and its relations with structural identifiability, *IEEE. Trans. Autom. Contr.* **29** (1984) 56–57.
- [124] E. Walter, L. Pronzato, On the identifiability and distinguishability of nonlinear parametric models, *Math. Comput. Simul.* **42** (1996) 125–134.
- [125] E. Walter, L. Pronzato, *Identification of Parametric Models from Experimental Data*, Springer, London, 1997.
- [126] O. Wolkenhauer, M. Mesarović, Feedback dynamics and cell function: why systems biology is called systems biology, *Mol. Biosyst.* **1** (2005) 14–16.
- [127] X. Xia, C. H. Moog, Identifiability of nonlinear systems with application to HIV/AIDS models, *IEEE Trans. Autom. Control* **48** (2003) 330–336.