

On the relationship between sloppiness and identifiability

Oana-Teodora Chis^{a,b}, Alejandro F. Villaverde^a, Julio R. Banga^a, Eva Balsa-Canto^{a,1}

^a*Bioprocess Engineering Group, Institute of Marine Research (IIM-CSIC), Vigo 36208, Spain*

^b*Technological Institute for Industrial Mathematics (ITMATI), Santiago de Compostela 15782, Spain*

Abstract

Dynamic models of biochemical networks are often formulated as sets of non-linear ordinary differential equations, whose states are the concentrations or abundances of the network components. They typically have a large number of kinetic parameters, which must be determined by calibrating the model with experimental data. In recent years it has been suggested that dynamic systems biology models are universally sloppy, meaning that the values of some parameters can be perturbed by several orders of magnitude without causing significant changes in the model output. This observation has prompted calls for focusing on model predictions rather than on parameters. In this work we examine the concept of sloppiness, investigating its links with the long-established notions of structural and practical identifiability. By analysing a set of case studies we show that sloppiness is not equivalent to lack of identifiability, and that sloppy models can be identifiable. Thus, using sloppiness to draw conclusions about the possibility of estimating parameter values can be misleading. Instead, structural and practical identifiability analyses are better tools for assessing the confidence in parameter estimates. Furthermore, we show that, when designing new experiments to decrease parametric uncertainty, designs that optimize practical identifiability criteria are more informative than those that minimize sloppiness.

Keywords: sloppiness, identifiability, parameter estimation, optimal experimental design

1. Introduction

Dynamic models of cellular processes describe the interactions among molecular entities – for example, proteins, transcripts or regulatory sites – that determine cellular behaviour. Such models typically consist of non-linear ordinary differential equations, whose state variables represent the concentrations of the network components and whose parameters characterize the reaction kinetics. Unfortunately, in most cases the parameter values are unknown, or only rough estimates are available. It is therefore necessary to calibrate the model using time-series experimental data [28]. The task of estimating the parameter values is an optimization problem, whose objective is to minimize a cost function that quantifies the differences between model predictions and experimental data. In dynamic models of biochemical systems this problem is typically characterized by limited observability, large number of parameters, and scarce, poor quality data [7]. As a consequence, its solution is in general challenging and computationally expensive, even

¹To whom correspondence should be addressed. Email: ebalsa@iim.csic.es

with efficient optimization methods. In addition, data limitations often lead to great uncertainty in the parameter estimates [47, 48].

During the last decade, several works [52, 22, 23, 44, 45, 46, 34] have introduced and elaborated the concept of *sloppiness*. The parameters of a dynamic model can be divided into stiff (those that can be determined with great certainty) and sloppy (those that can vary by orders of magnitude without influencing significantly the model output), although it is difficult to establish a clear cut-off between both categories. The sloppiness of a model is quantified from the distribution of the eigenvalues of its Fisher information matrix; a separation of more than 3 orders of magnitude in the eigenvalues qualifies a model as sloppy.

It has been claimed that dynamic systems biology models are *universally sloppy* [23], and therefore it is not possible to obtain accurate estimates of their parameters. This idea has been cited in many publications and, unfortunately, has sometimes led to misinterpretations. Since parameter estimation is often an arduous task in practice, it is tempting to use the notion of sloppiness to argue that it is not necessary nor possible to uniquely determine the parameter values, thus justifying that no further efforts are invested to it (see for example [39, 16, 31, 51, 19]). The suggestion that sloppiness is a universal – or, more precisely, ubiquitous – property of systems biology models has spurred a debate: should modellers desist from trying to estimate precise values for the parameters and, instead, focus on characterizing model *predictions*? Exploring this direction, Cedersund and coworkers [13, 11, 12] coined the term “core predictions” to denote specific model outcomes that can be uniquely determined, even if parameter values cannot. The parameter regions complying with core predictions can be found using optimization, at least for models of moderate size [11].

In [43] the origin of sloppiness was traced back to the structure of the sensitivity matrix, which contains the sensitivities of the model outputs with respect to the parameters. *Experimental design* was proposed as a way of reducing sloppiness, concluding that the intensity of the effect is highly dependent on the available data, thus challenging the universality of the property. The importance of the role played by experimental design in this task had been stressed e.g. in [2, 32]. Finally, it is worth mentioning that sloppiness has sometimes been considered as an indication of biological robustness [21, 30].

In this work we aim at clarifying the application and implications of sloppiness, to avoid certain misconceptions. We study the role played by model parameters using the well established framework of *identifiability*, which has a long history of application in dynamical systems [50], including biological models. Indeed, while the study of parameter identifiability has been present in the systems and control literature for decades, many methodological advances in the field have been motivated by biological applications, starting in the 1970s and 1980s [8, 37, 20, 27] and continuing until the present day [49]. Identifiability-based concepts are rigorously defined and well understood, and they can be analysed with a large number of techniques. Hence it is of interest to clarify the connection between sloppiness and (lack of) identifiability [41, 19, 38], which, despite recent developments, is still incompletely understood. Here we study the relationship between sloppiness and identifiability from both structural and practical points of view. Using a set of case studies, we inquire to which extent is sloppiness determined by the model structure, and how it is influenced by the quantity and quality of experimental data. Then we explore optimal experiment design alternatives that reduce sloppiness and improve identifiability, and clarify the connections between both concepts. We conclude that identifiability analysis can be more insightful than sloppiness for characterizing the mapping between parameters and outputs.

2. Methods

We consider general nonlinear models of the form:

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

where $\mathbf{x} = (x_1, \dots, x_{n_x}) \in \mathbf{R}^{n_x}$ is the state vector, $\mathbf{u} = (u_1, \dots, u_{n_u}) \in \mathbf{R}^{n_u}$ a n_u -dimensional input (control) vector, 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 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 $\mathbf{p} \in \mathbf{P}$.

Note that the model in Eqns. (1) is composed of two different elements: (i) a set of ordinary differential equations (ODEs), describing the system dynamics and (ii) the observation function, which relates states (typically concentrations or amounts) and measurements. In this work we consider that the mathematical structure of the system dynamics (\mathbf{f} , \mathbf{g}) can by no means be modified, whereas the mathematical structure of the observation function (\mathbf{h}) may eventually be modified by the experimental scheme (which, in fact, leads to a different model).

2.1. Parameter estimation

The above representation (1) is a sufficiently accurate mathematical description of the real system, i.e. the only uncertainty is represented by the vector of unknown parameters. This means that, in principle, there is a unique ‘‘true’’ value of the parameters, denoted by $\mathbf{p}^* = (p_1^*, \dots, p_{n_p}^*)$, which allows the model to reproduce a given data set and to predict the system behaviour. This vector \mathbf{p}^* is computed by means of data fitting, i.e. by solving an optimization problem devoted to minimizing the log-likelihood function, which for Gaussian experimental noise reads:

$$\chi^2(\mathbf{p}) = \sum_{e=1}^{n_e} \sum_{o=1}^{n_y} \sum_{s=1}^{n_s} \frac{[\mathbf{y}_{e,o,s}(\mathbf{p}, t_s) - \tilde{\mathbf{y}}_{e,o,s}]^2}{\sigma_{e,o,s}^2}, \quad (2)$$

where n_e is the number of experiments, n_y the number of observables for each experiment, and n_s the number of sampling times; $\mathbf{y}_{e,o,s}(\mathbf{p}, t_s)$ denotes the output of the model (1) for the sampling time t_s under the experimental conditions e ; $\tilde{\mathbf{y}}_{e,o,s}$ is the corresponding experimental data; and $\sigma_{e,o,s}^2$ is the variance of the measurement noise.

2.2. Structural identifiability

Structural identifiability analysis studies the possibility of finding a unique solution to the parameter estimation problem, assuming perfect experimental data (i.e. noise-free and continuous in time) [50]. A parameter p_i , $i = 1, \dots, n_p$ is *structurally globally (or uniquely) identifiable* if for almost any $\mathbf{p}^* \in \mathbf{P}$, $\Sigma(\mathbf{p}) = \Sigma(\mathbf{p}^*) \Rightarrow p_i = p_i^*$, whereas 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 $\mathbf{p} \in \mathbf{V}(\mathbf{p}^*)$ and $\Sigma(\mathbf{p}) = \Sigma(\mathbf{p}^*) \Rightarrow p_i = p_i^*$.

In some cases, an unidentifiable parameter may be made identifiable by including more measured outputs in the observation function, \mathbf{h} . This modification leads to a new model with a different structure. In other cases, however, the model may be structurally unidentifiable even if all states are accessible to the experimentation, i.e. $\mathbf{y} = \mathbf{x}$. In this case it will not be possible to avoid the lack of identifiability.

Recent reviews [35, 18] compare alternative methods to perform global structural identifiability analysis for nonlinear models; additionally, state of the art local methods are described in [15, 49]. In this work we adopt the MATLAB based GenSSI toolbox [17], which combines the generating series approach with identifiability *tableaus*. The underlying idea of the generating series approach is that the observables \mathbf{y} can be expanded in series with respect to time and inputs around a given time point (t_0), and that the uniqueness of the series coefficients guarantees the structural identifiability of the model. The series coefficients are computed by means of successive Lie derivative of \mathbf{h} along the vector fields \mathbf{f} and \mathbf{g} . The identifiability *tableaus* correspond to the Jacobian of the Lie derivatives with respect to the model parameters, and help to decide on global or local structural identifiability of the model [5].

2.3. Sloppiness

Parameter sloppiness can be quantified by means of the eigenvalues of the Hessian of the log-likelihood function (Eqn. 2) as evaluated in the optimal value of the parameters \mathbf{p}^* . The Fisher information matrix (\mathcal{F}) can be used as an approximation of the Hessian:

$$\mathcal{F} = E\left(\left[\frac{\partial \chi^2(\mathbf{p})}{\partial \mathbf{p}}\right]^T \left[\frac{\partial \chi^2(\mathbf{p})}{\partial \mathbf{p}}\right]\right), \quad (3)$$

In [23] Gutenkunst et al. suggest using the following normalized version of \mathcal{F} to assess sloppiness:

$$F_{i,j} = \frac{1}{n_d} \sum_{e=1}^{n_e} \sum_{o=1}^{n_y} \sum_{s=1}^{n_s} \frac{1}{\sigma_{e,o,s}^2} \frac{\partial y_{e,o,s}}{\partial \log(\mathbf{p}_i)} \frac{\partial y_{e,o,s}}{\partial \log(\mathbf{p}_j)}, \quad (4)$$

where n_d is the total amount of data, and logarithms of the parameters are used to avoid scaling issues due to different orders of magnitude of the parameters.

The largest eigenvalues correspond to “stiff” parameters and the smallest to “sloppy” parameters [10]. Models are said to be sloppy when the maximum eigenvalue (λ_{max}) is orders of magnitude larger than the minimum eigenvalue (λ_{min}). Even if there is no clear cut-off between sloppy and stiff parameters, from previous works [53, 23] it can be considered that a model is sloppy if $C_F = \frac{\lambda_{min}}{\lambda_{max}} \lesssim 10^{-3}$.

Sloppiness is a local property that depends on the parameter vector of reference, \mathbf{p}^* . A generalization called multiscale sloppiness, which accounts for non-infinitesimal deviations from \mathbf{p}^* , has been proposed recently [38].

2.4. Practical identifiability

Practical identifiability analysis assesses the quality of parameter estimates. Several methods are available for this purpose. Confidence intervals can be robustly calculated using Monte-Carlo based approaches such as bootstrap [29], which performs many calibrations after generating pseudo-measurements. Likewise, the profile likelihood method involves re-optimization with respect to all parameters for each value of a given parameter [40]. These methods provide robust estimations of the confidence intervals, although the associated computational cost complicates their application to large scale models. Alternatively, the confidence interval (ρ_i) of θ_i^* may be obtained through the covariance matrix. Assuming a normal distribution for the parameters, the confidence interval can be calculated as:

$$\rho_i = \pm t_{\alpha/2}^\gamma \sqrt{\mathbf{C}_{ii}} \quad (5)$$

where $t_{\alpha/2}^\gamma$ is given by Student's t-distribution, $\gamma = N_d - \eta$ degrees of freedom and $(1 - \alpha)100\%$ is the confidence interval selected, typically 95%.

For non-linear models there is no exact way of obtaining \mathbf{C} . The most widely used approximation is possibly the one based on the Crammèr-Rao inequality. It establishes that the inverse of the Fisher information matrix is a lower bound of the covariance matrix (under certain assumptions on the number of data and the non-linearity of the model): $\mathbf{C} \geq \mathcal{F}^{-1}$.

From the covariance matrix it is possible to compute the correlation matrix:

$$Cr_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \quad i = 1, \dots, n_p \quad j = 1, \dots, n_p \quad (6)$$

in such a way that two parameters (p_i, p_j) are completely uncorrelated if $Cr_{ij} = 0$ or completely correlated if $Cr_{ij} = \pm 1$.

If a model is structurally non identifiable, the Fisher information matrix, \mathcal{F} (Eqn. 3) and its modified version (Eqn. 4) should be theoretically singular for any parameter values. Note, however, that numerical inaccuracies when solving the model (Eq. 1) and computing the parametric sensitivities may lead to a non-singular \mathcal{F} , even for a structurally identifiable model. Note also that the fact that both matrices are singular does not imply that the model is structurally unidentifiable.

2.5. Role of the experimental error

In order to analyse the role of the amount of experimental noise, we generate a significant number (200) of realizations of noisy pseudo-experimental data. We do it by adding noise to the output of the models, considering a Gaussian distribution with known time-varying variance:

$$\tilde{\mathbf{y}}_{e,o,s} = \mathbf{y}_{e,o,s} + \epsilon_{e,o,s}, \quad \text{with} \quad \epsilon_{e,o,s} = \sigma \times \text{rand}C_{e,o,s} \times \max(\mathbf{y}_{e,o}) \quad (7)$$

where $\epsilon_{e,o,s}$ are normally distributed independent random variables with standard deviation $\sigma \times \max(\mathbf{y}_{e,o})$, and $\text{rand}C_{e,o,s}$ is a random number drawn from the standard normal distribution $\mathcal{N}(0, 1)$ defined for every experiment, observable and sampling time. We consider three scenarios: $\sigma = 0.05$, $\sigma = 0.1$, and $\sigma = 0.2$, and solve the subsequent parameter estimation problem for each of the 200 pseudo-data realizations. Then we compute the sloppiness (C_F) and mean value of the confidence intervals (ρ) for each of the optimal solutions obtained.

2.6. Optimal experimental design

The quality of parameter estimates may be improved if more data is available in the calibration process. Experiments that maximize the information content of the newly added data can be found via optimal experimental design. The idea is to formulate a dynamic optimization problem where the objective is to find those experimental conditions (number of experiments, external inputs, number and location of sampling times, initial conditions, experiment duration, and possibly observables for each experiment) which result in maximum information content, as measured e.g. by \mathcal{F} , subject to the system dynamics of Eqn. (1) and experimental constraints. The problem can be solved by a combination of the control vector parameterization (CVP) method and an optimizer that enables the simultaneous design of several dynamic experiments with optimal sampling times [4, 5].

Several \mathcal{F} -based criteria can be used for optimal experimental design. The most widely used are the D-optimum and E-optimum criteria, which correspond to the maximization of the determinant of \mathcal{F} and the maximization of its minimum eigenvalue, respectively. Here we will also explore the minimization of the sloppiness, i.e. the minimization of C_F . This type of design will be called S-optimum.

In this work we use the MATLAB toolbox AMIGO (Advanced Model Identification using Global optimization) [6] for the purposes of parameter estimation, sensitivity and practical identifiability analysis, and optimal experimental design.

2.7. Model reduction

Given a complex system, the model reduction problem consists in finding an approximation that generates a similar output with less computational cost, while maintaining the properties of the original system. Many model reduction techniques have been developed in control theory and related areas of engineering [1]. Most of them can be classified in two categories, Singular Value Decomposition (SVD) and moment matching. A popular example of the first type of methods is the balanced order reduction presented by Moore [36], which uses principal component analysis (PCA), a tool introduced in 1933 [26]. It is possible to apply SVD to nonlinear systems by extending Proper Orthogonal Decomposition methods (POD) [54].

The idea that a sloppy model is insensitive to parameter changes along sloppy directions, but highly sensitive along stiff combinations of parameters, has been exploited to propose reduced order modelling techniques [42, 3, 9]. The aim is to obtain a non-sloppy model that captures the relevant dynamics of the original model, but with a reduced number of states and parameters. Sloppy models in physics and biology show weak dependence of macroscopic observables on microscopic details, which allows effective descriptions with reduced dimensionality [33].

3. Models

To perform the different analysis proposed in this work we have selected a collection of examples representative of biological systems and processes. They are of different sizes and include several types of non-linear terms, such as general mass action, Michaelis-Menten or Hill kinetics. All the selected models are sloppy.

3.1. Example 1: Fed-batch reactor for ethanol production

This model [25] describes ethanol production using the anaerobic fermentation of glucose with the help of *Sacharomyces cerevisiae*. The unsteady state material balances are as follows:

$$\begin{aligned}\dot{x} &= x\mu - u\frac{x}{V}, \\ \dot{S} &= -v_0x\mu + u\frac{S_0 - S}{V}, \\ \dot{P} &= x\epsilon - u\frac{P}{V}, \\ \dot{V} &= u,\end{aligned}\tag{8}$$

where x, S, P are the cell mass, substrate, and ethanol concentrations (g/L), u corresponds to glucose concentration (g/L), V is the volume (L), $\mu = \frac{\mu_0}{1 + \frac{P}{K_P}} \frac{S}{K_S + S}$ is the specific growth rate, and $\epsilon = \frac{1}{1 + \frac{P}{K_P}} \frac{S}{K_S + S}$ the specific productivity. The parameter values are $\mu_0^* = 0.408$, $K_P^* = 16$, $K_S^* = 0.22$, $v_0^* = 10$, $K_P'^* = 71.5$, $K_S'^* = 0.44$, and the initial conditions are $[x = 1, S = 150, P = 0, V = 10]$; x, S and P are measured.

3.2. Example 2: A linear biochemical pathway with 14 steps

The system, that could describe a linear biochemical network, is represented by fourteen differential equations and sixteen parameters as follows:

$$\begin{aligned}
 \dot{x}_1 &= -\frac{v_m x_1}{(k_m + x_1)} + p_1 u, \\
 \dot{x}_2 &= p_1 x_1 - p_2 x_2, \\
 \dot{x}_3 &= p_2 x_2 - p_3 x_3, \\
 \dot{x}_4 &= p_3 x_3 - p_4 x_4, \\
 \dot{x}_5 &= p_4 x_4 - p_5 x_5, \\
 \dot{x}_6 &= p_5 x_5 - p_6 x_6, \\
 \dot{x}_7 &= p_6 x_6 - p_7 x_7, \\
 \dot{x}_8 &= p_7 x_7 - p_8 x_8, \\
 \dot{x}_9 &= p_8 x_8 - p_9 x_9, \\
 \dot{x}_{10} &= p_9 x_9 - p_{10} x_{10}, \\
 \dot{x}_{11} &= p_{10} x_{10} - p_{11} x_{11}, \\
 \dot{x}_{12} &= p_{11} x_{11} - p_{12} x_{12}, \\
 \dot{x}_{13} &= p_{12} x_{12} - p_{13} x_{13}, \\
 \dot{x}_{14} &= p_{13} x_{13} - p_{14} x_{14},
 \end{aligned} \tag{9}$$

where x_i stands for the components concentrations. The initial conditions are assumed to be zero for all components but $x_1(0) = 2$, and the nominal parameter values are: $v_m = 1.6$, $k_m = 2.8$, $p_1 = 0.5$, $p_2 = 1.0$, $p_3 = 1.6$, $p_4 = 1.2$, $p_5 = 1.2$, $p_6 = 0.3$, $p_7 = 1.3$, $p_8 = 0.8$, $p_9 = 1.4$, $p_{10} = 1.5$, $p_{11} = 1.0$, $p_{12} = 1.8$, $p_{13} = 1.2$, $p_{14} = 0.4$. All states can be measured.

3.3. Example 3: Six-gene regulatory network

This model, proposed in the DREAM6 parameter estimation challenge², assumes linear kinetics for mRNA degradation and protein synthesis (translation) and degradation. The mRNA synthesis follows a Hill-type kinetics, unless for some cases where there is no regulatory input to a gene. The mathematical model reads:

$$\begin{aligned}
 \dot{m}r_1 &= p s t_1 - m d r_1 m r_1, \\
 \dot{p}_1 &= r s t_1 m r_1 - p_1 p d r, \\
 \dot{m}r_2 &= p s t_2 a s_1 r s_1 - m d r_2 m r_2, \\
 \dot{p}_2 &= r s t_2 m r_2 - p_2 p d r, \\
 \dot{m}r_3 &= p s t_3 a s_3 r s_4 - m d r_3 m r_3, \\
 \dot{p}_3 &= r s t_3 m r_3 - p_3 p d r, \\
 \dot{m}r_4 &= p s t_4 a s_2 r s_2 - m d r_4 m r_4, \\
 \dot{p}_4 &= r s t_4 m r_4 - p_4 p d r, \\
 \dot{m}r_5 &= p s t_5 r s_3 - m d r_5 m r_5, \\
 \dot{p}_5 &= r s t_5 m r_5 - p_5 p d r, \\
 \dot{m}r_6 &= p s t_6 r s_5 - m d r_6 m r_6, \\
 \dot{p}_6 &= r s t_6 m r_6 - p_6 p d r, \\
 a s_1 &= \frac{(p_1/k_2)^{h_2}}{1+(p_1/k_2)^{h_2}}; a s_2 = \frac{(p_1/k_2)^{h_1}}{1+(p_1/k_1)^{h_1}}; a s_3 = \frac{(p_1/k_3)^{h_3}}{1+(p_1/k_3)^{h_3}}, \\
 r s_1 &= \frac{1}{1+(p_6/k_5)^{h_5}}; r s_2 = \frac{1}{1+(p_5/k_8)^{h_8}}; r s_3 = \frac{1}{1+(p_4/k_6)^{h_6}}, \\
 r s_4 &= \frac{1}{1+(p_2/k_4)^{h_4}}; r s_5 = \frac{1}{1+(p_4/k_7)^{h_7}}
 \end{aligned} \tag{10}$$

²<http://www.the-dream-project.org/challenges/dream6-estimation-model-parameters-challenge>

where the p_i are proteins and mr_i mRNA concentrations, pdr is a protein degradation constant, mdr_i corresponds to the mRNA degradation rate, the pst_i are the corresponding promoter strengths, rs_i the strengths of the ribosomal binding sites, k_i the binding affinities, and h_i the Hill coefficients. All mRNA degradation rate constants are assumed to be known and equal to 1. All other parameters are unknown. In this work we chose the following nominal values: $pdr = 0.8$, $pst_1 = pst_5 = pst_6 = 3$, $pst_2 = 4.7$, $pst_3 = 5$, $pst_4 = 20$, $rst_1 = 1$, $rst_2 = rst_3 = rst_6 = 4$, $rst_4 = 0.4$, $rst_5 = 6$, $k_1 = k_5 = 1$, $k_2 = 0.9$, $k_3 = k_6 = k_7 = 0.1$, $k_4 = 9.5$, $k_8 = 0.2$, $h_1 = h_4 = h_8 = 4$, $h_2 = h_3 = h_6 = h_7 = 2$, $h_5 = 1$. At initial time all protein concentrations are set to 1, and mRNAs are set to 0. It is assumed that all protein and mRNA concentrations can be measured.

4. Results

4.1. The geometry of sloppiness

Before proceeding to analyse the models presented in Section 3 we illustrate the geometric implications of the definition of sloppiness, and how it relates to identifiability, using simple examples.

We begin by remarking that *two equally sloppy experimental schemes may have very different identifiability properties*. The reason is that it is possible to modify the confidence volume of the confidence hyper-ellipsoid without modifying the ratio between the maximum and the minimum semi-axis, i.e. the sloppiness. This fact can be trivially illustrated using a two dimensional example, as in Fig. 1.

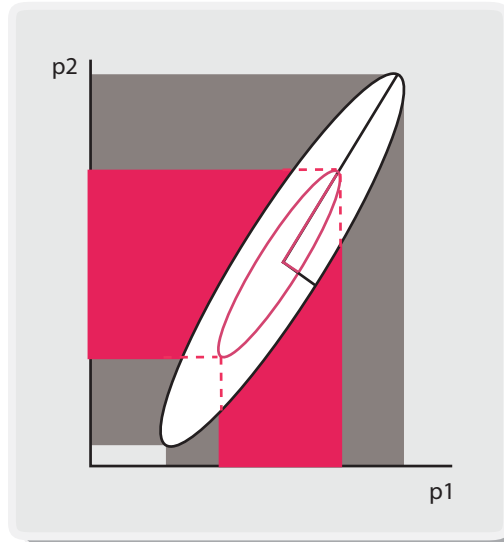


Figure 1: **Analysis of sloppiness from a geometric point of view.** The figure represents two confidence ellipses for an arbitrary example with two-parameters. The semi-axes of the larger ellipse are twice as long as the semi-axes of the smaller ellipse. The parameter values can be determined with higher confidence for the smaller ellipse, whereas sloppiness is the same in both cases.

Next we show that *the experimental setup influences sloppiness and identifiability in different ways*. To this end we use a simple dynamic model that can be analysed symbolically. Consider the following system, which could be a description of a linear biochemical network:

$$\begin{aligned}\dot{x}_1 &= -p_1x_1, \\ \dot{x}_2 &= p_1x_1 - p_2x_2,\end{aligned}\tag{11}$$

where x_1 and x_2 represent the concentrations of two reaction components, with initial conditions $x_1(0) = x_{10}$ and $x_2(0) = 0$. If we take a single measurement at time t of both states, the sloppiness can be computed symbolically and it can be shown (see Supplementary Information) that *sloppiness depends on the values of the parameters and on the sampling time*, but it is independent of the initial condition of the experiments. In contrast, *practical identifiability depends also on the initial condition*, because the eigenvalues of the Fisher information matrix – and hence the confidence on the parameters – are dependent on the system’s state at the beginning of the experiment. Figure 2 illustrates this fact, showing that, if it is possible to increase the initial amount of x_1 in the system, the identifiability of both parameters improves – without altering the sloppiness.

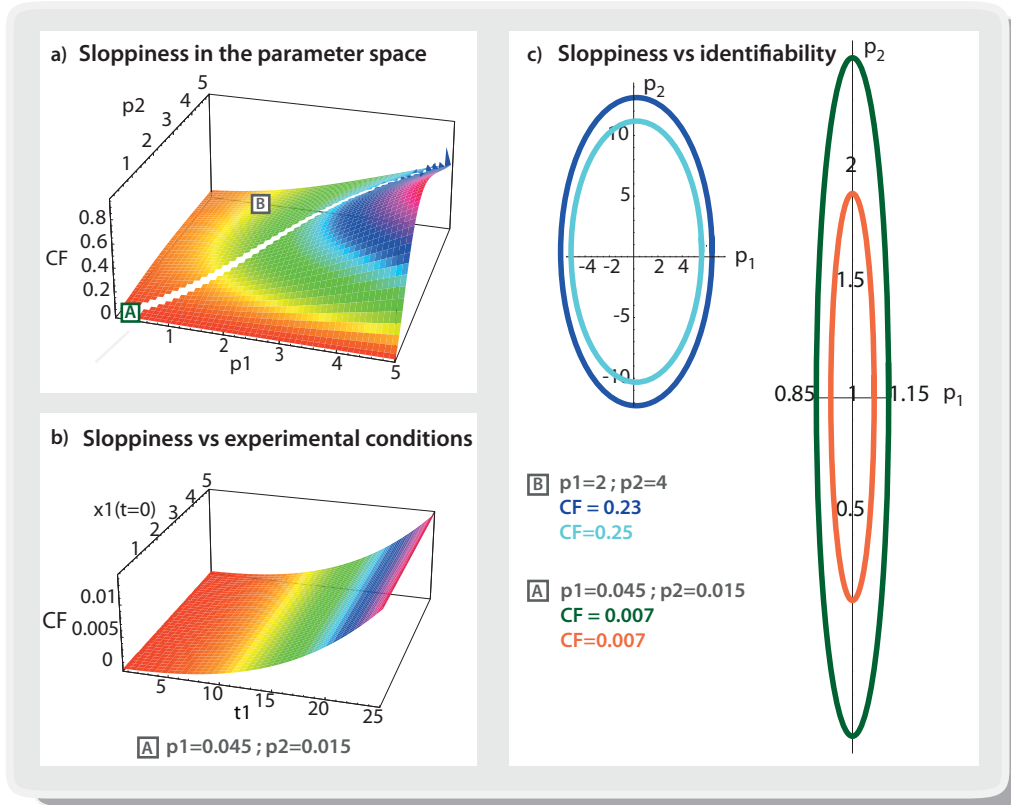


Figure 2: **Effect of parameter values and experimental setup on sloppiness and practical identifiability.** Figure (a) shows that sloppiness is highly dependent on parameter values for given experimental conditions (i.e. initial amount of x_1 and sampling time). Figure (b) shows the influence of the experimental setup: sloppiness varies significantly with the choice of the sampling time t_1 , but it is independent of the initial condition $x_1(t=0)$. As a consequence, Figure (c) illustrates the fact that it is possible to estimate parameters with reasonable accuracy despite being sloppy. The blue ellipses represent relative confidence intervals for initial condition $x_1(0) = 0.5$ and parameter values $p_1 = 2$ and $p_2 = 4$, coloured in dark blue (for a choice of sampling time $t_1 = 1$) and light blue (for $t_1 = 20$). Both cases are non-sloppy and non-identifiable. The green and red ellipses represent relative confidence intervals for parameter values $p_1 = 0.045$ and $p_2 = 0.015$, when 3 samples are measured at times $t = [15, 18, 22]$, and for different initial conditions: the green ellipsis corresponds to $x_1(0) = 3$, and the red one to $x_1(0) = 5$. While sloppiness is the same in both cases, practical identifiability is affected by the initial condition. For the case $x_1(0) = 5$, at least p_1 can be considered as identifiable, and the confidence level for p_2 is $\approx 60\%$, which can be reasonable in many cases.

4.2. Sloppiness versus structural identifiability

We analysed the structural identifiability of the three models presented in Section 3 using the GenSSI toolbox [17], as outlined in Section 2.2. The three models are structurally identifiable, because they have full rank identifiability *tableaus* (see Figure 3).

A *tableau* represents the non-zero elements of the Jacobian of the generating series coefficients with respect to the parameters. If the Jacobian is rank deficient, i.e. the *tableau* presents empty columns, the corresponding parameters may be non-identifiable. If the rank of the Jacobian equals the number of parameters, the model is at least locally identifiable. A unique non-zero element in a given row indicates a globally identifiable parameter. The *tableau* can

be “reduced” by removing such parameters, which may in turn lead to new unique non-zero elements, and so on.

Figure 3 shows that examples 2 (biochemical pathway) and 3 (six-gene regulatory network) are globally structurally identifiable, while in example 1 (fed-batch reactor) all parameters are globally structurally identifiable except K_P and K'_P , which are only locally structurally identifiable. As a consequence, it is in principle possible to identify all the parameters in these examples, provided that the appropriate datasets are used for parameter estimation.

To determine the quality that can be expected from the parameter estimates we use practical identifiability techniques, as explained next (Section 4.3).

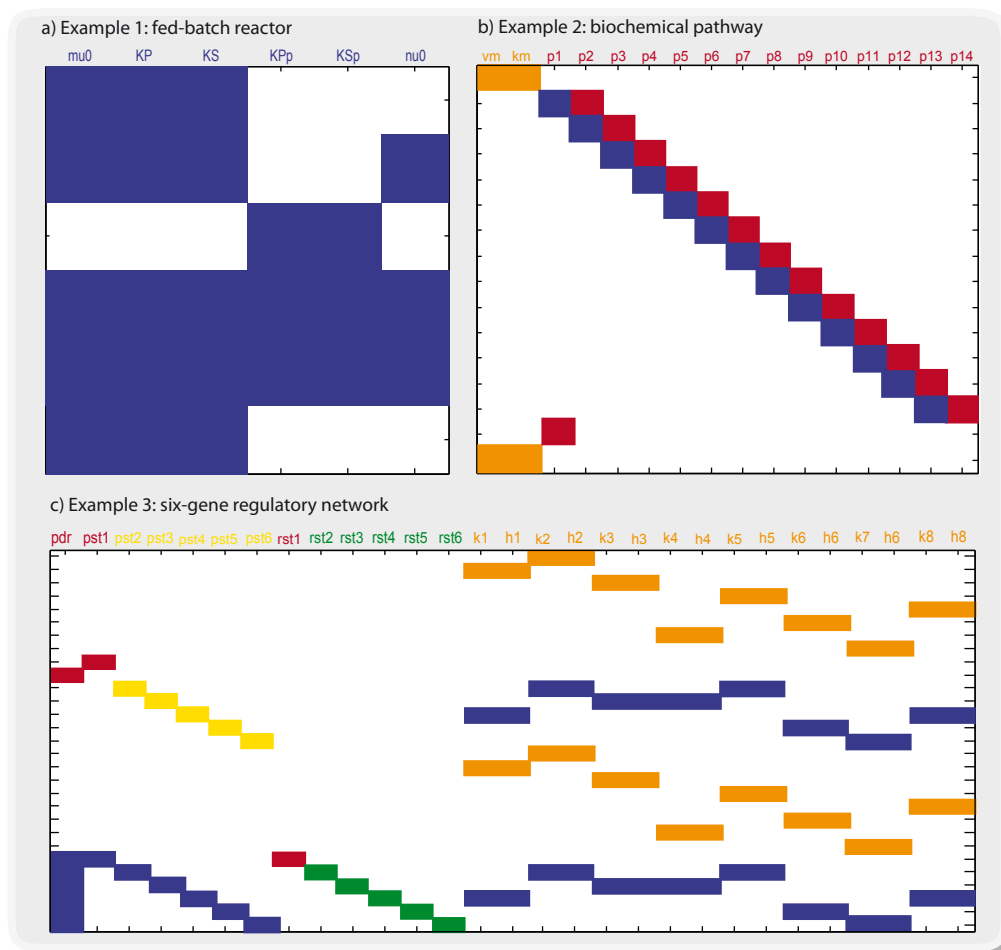


Figure 3: **Identifiability tableaus for the three examples.** a) Fed-batch reactor for ethanol production. b) Biochemical pathway: in red, direct globally identifiable parameters; in orange, 2nd level globally identifiable parameters. c) Six-gene regulatory network: in red, direct globally identifiable parameters; in orange, yellow and green, 2nd to 4th level globally identifiable parameters.

4.3. Sloppiness versus practical identifiability

4.3.1. Example 1: Fed-batch reactor for ethanol production

Here we analyse the practical identifiability of this model for an experiment in which the reactor is fed with a constant amount of glucose $u = 10$ for $t_f = 24h$. We assume that x , S , and P are measured and consider two possible scenarios: with 30 (Case 1a) or 60 (Case 1b) samples, which are equidistantly distributed throughout the experiment duration. The total number of data points is 90 in Case 1a and 180 in Case 1b.

Although the addition of sampling times reduces the sloppiness (from $C_{\mathcal{F}} = 1 \times 10^{-9}$ to $C_{\mathcal{F}} = 4.3 \times 10^{-7}$, as shown in Figure 4.a) and the confidence intervals (Figure 4.b), the model is both sloppy and practically unidentifiable in both scenarios, specially for K_s and K'_s . Several parameters (K_p , K_s , K'_s , and K'_p) are strongly correlated, as shown in Figure 4.c. Increasing the number of data points decreases correlation, although the effect is small; particularly, K'_s and K_s are highly correlated in both cases. Figure 4.d plots model predictions obtained by changing the value of K'_s while keeping all other parameters fixed to the optimum. Cell mass and substrate concentration are insensitive to changes in K'_s , whereas ethanol production is clearly affected at the end of the experiment. This result shows that *the fact that a parameter is sloppy does not necessarily mean that model predictions will remain the same regardless of the parameter value*. Note that, since K'_s and K_s are highly correlated, the effect of changes in K'_s can be compensated by changing K_s . Hence, if both parameters are simultaneously manipulated the predictions may remain unchanged. This case study provides a clear example of the differences between sloppiness and (practical) identifiability.

It should be also noted that only the last samples of ethanol provide information for estimating K_s and K'_s . Therefore, it is expected that the accumulation of sampling times at the end of the experiment will improve identifiability. To assess this intuition we consider a third scenario (Case 1c) in which the data points are not equidistant. We obtain that, with 30 non equidistant sampling times, it is possible to reduce both sloppiness ($C_{\mathcal{F}} = 1.3 \times 10^{-6}$) and confidence intervals for K_s and K'_s , with respect to the use of 60 equidistant sampling times (Case 1b). This result stresses that the *location* of sampling times is sometimes more determining than their *number*, for the purpose of reducing sloppiness and improving identifiability. Thus, while the quantity of data is important, it may not be indispensable to have massive amounts of high-quality measurements, as has been suggested sometimes [14].

Finally, we analyse the effect of experimental noise. We proceed as described in section 2.5 for the case with constant glucose feeding and 30 non equidistant sampling times (Case 1c). Figure 5.a shows the distribution of $C_{\mathcal{F}}$ values, while Figure 5.b presents the distribution of the mean of the confidence intervals. Both sloppiness and confidence intervals depend on the amount of experimental noise; in general, the larger the variance, the greater the sloppiness and the confidence intervals.

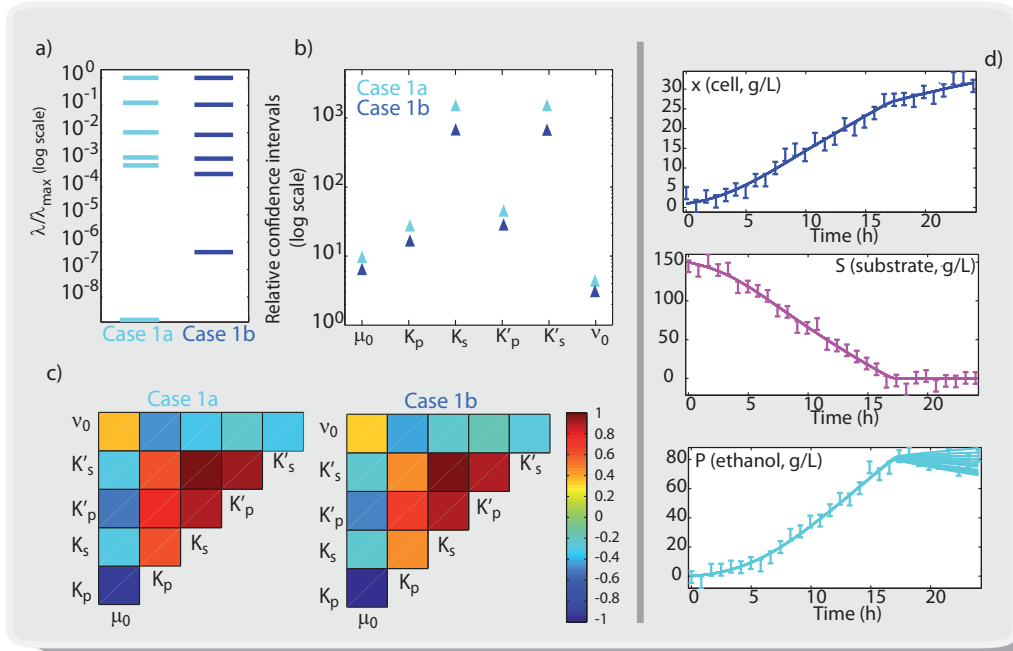


Figure 4: **Sloppiness vs practical identifiability for the example 1: ethanol production.** a) Normalized eigenvalue spectra provide a measure of sloppiness. In the two scenarios the eigenvalues span between 7 and 9 orders of magnitude, indicating sloppiness in both cases. b) Normalized confidence intervals for the model parameters for both scenarios. c) Correlation matrices. d) Model predictions when modifying the value of K'_s between 0.5 and 2 times its optimal value, with all other parameters fixed to the optimum. Plots show that x and P are not sensitive to K'_s , and therefore those measurements are not providing information to estimate its value, at least for the selected constant feeding profile.

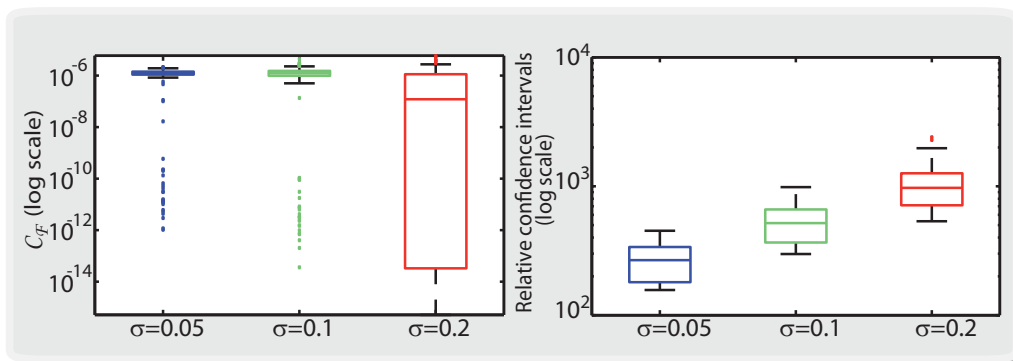


Figure 5: **Role of the experimental noise in the sloppiness and practical identifiability for the example 1: ethanol production.** a) Sloppiness: distribution of the values of C_F (log scale) out of 200 runs for each of the three different experimental noise amounts, for 30 non equidistant sampling times (Case 1c). b) Practical identifiability: distribution of the mean value of the confidence intervals ρ (log scale).

4.3.2. Example 2: biochemical pathway

For this example we set the stimulus to $u = 1$ and the experiment duration to $t_f = 20$. Since there are 16 unknown parameters, a reasonable number of data was selected to be 56. We consider three different ways of placing the total of 56 measurements: 2a), 2b), and 2c). In case 2a) we only measure the concentrations of the first and last components, 28 samples each. In case 2b) we measure half of the components, 8 samples each; from the more than 3000 different combinations we choose to measure $(x_1, x_2, x_5, x_7, x_{10}, x_{13}$ and $x_{14})$. In case 2c) we measure all the components in the network, 4 samples each. In all cases, sampling times are selected to be equidistantly distributed.

Figure 7 presents an overview of the results. The system is sloppy in all scenarios (Figure 7.a), with $C_{\mathcal{F}} < 1 \times 10^{-5}$. However, the distribution of eigenvalues and the values of $C_{\mathcal{F}}$ change with the number of observables. The possibility of measuring all the components in the pathway significantly reduces the sloppiness of the system and improves the confidence on the parameter values (Figure 7.b), even if the number of measurements per observable decreases (the total amount of data is the same in all cases). Note that this is a characteristic of linear pathways. Figure 7.c shows the sensitivity of the observables with respect to the different parameters, showing that the sensitivity matrix is almost triangular, as expected (since the concentration of a component is not affected by the downstream subnetwork). Figure 7.c also reveals (see yellow arrows) that it would have been necessary to measure x_3, x_4 or x_6, x_9 and x_{12} to improve the identifiability of p_3, p_4, p_9 and p_{12} . The addition of observables improves confidence (Figure 7.b) and tends to decorrelate parameters (Figures 7.d).

Figure 6 shows the effect of the amount of experimental noise for the scenario 2c). As in the previous example, the value of C_F is influenced by the amount of experimental noise, and all cases are regarded as sloppy. The mean value of the confidence intervals also increases with the experimental error, as expected.

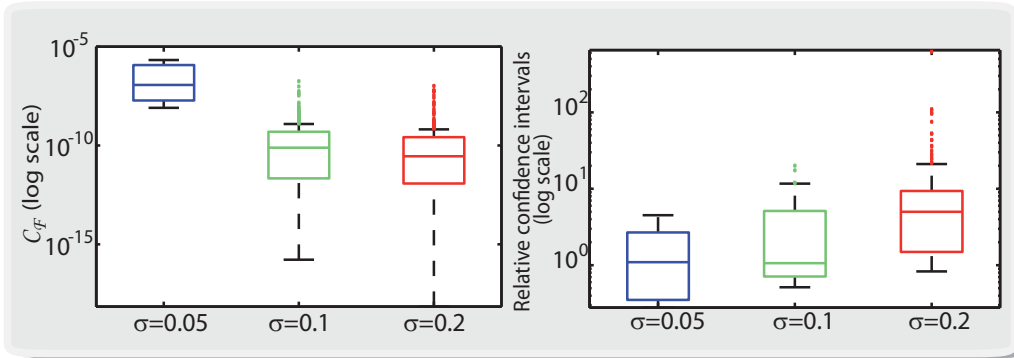


Figure 6: **Role of the experimental noise in the sloppiness and practical identifiability for the example 2: biochemical pathway.** a) Sloppiness: distribution of the values of $\log(C_F)$ out of 200 runs for each of the three different experimental noise scenarios. Values expand several orders of magnitude, and the model is sloppy in all cases. b) Practical identifiability: distribution of the mean value of the confidence intervals ρ . Confidence intervals are larger for the cases with larger variance in the experimental error.

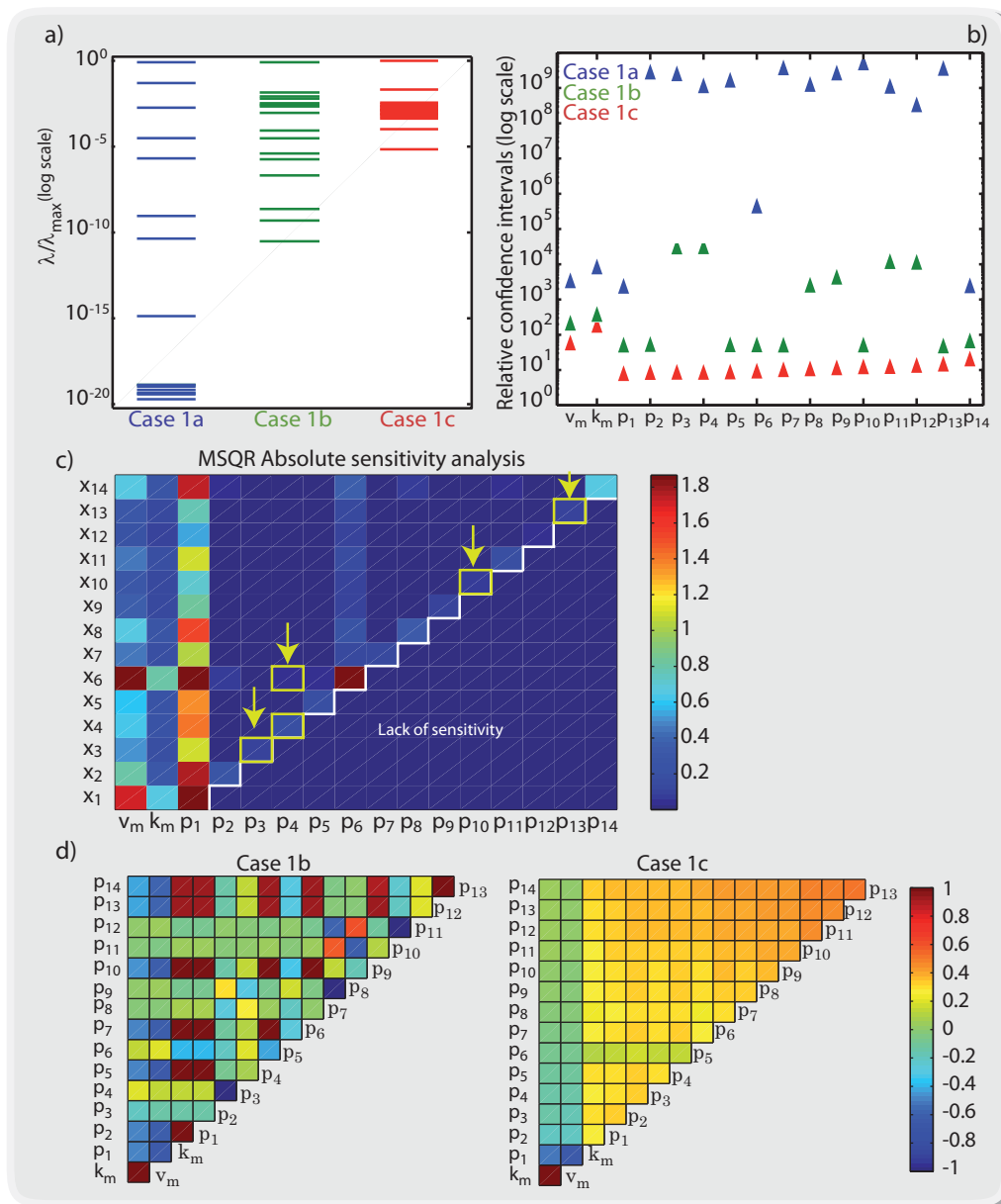


Figure 7: **Sloppiness vs practical identifiability for the example 2: biochemical pathway.** a) Sloppiness: normalized eigenvalue spectra for the different scenarios. The range of the eigenvalues spans between 20 orders of magnitude – for the case where only the first and the last component of the network are measured – to around 6 for the case in which all components are measured, indicating sloppiness in all scenarios. b) Normalized confidence intervals of the parameters in different scenarios. Results show that confidence intervals improve with the addition of observables. c) Sensitivity matrix illustrating how the observables depend on the parameters. For example, all observables are sensitive to p_1 , thus its value is computed with the highest confidence. The most informative observables for the purpose of parameter estimation are x_6 , x_1 , x_{14} and x_2 . d) Correlation matrix for cases 2b and 2c, showing that decorrelation increases with the addition of observables.

4.3.3. Example 3: six-gene regulatory network

In this case study, time courses of all mRNA concentrations may be obtained by microarray experiments, and protein concentrations can be measured by means of fluorescence microscope experiments. We consider an experiment in which all mRNA and protein concentrations are measured during 20 time units, with 21 equidistant sampling times (252 total data points), and experimental errors with $\sigma = 0.20$. Under these conditions, a sensitivity analysis reveals that it will be problematic to estimate several parameters, specially binding affinities and Hill coefficients (Figure 8.a). Indeed, solving the parameter estimation problem for 200 different realizations of the experimental data yields large distributions of parameter values (Figure 8.b), which reveals poor practical identifiability. In addition, C_F values are between 1×10^{-20} and 1×10^{-12} , indicating that the model is sloppy independently of the realization of the data.

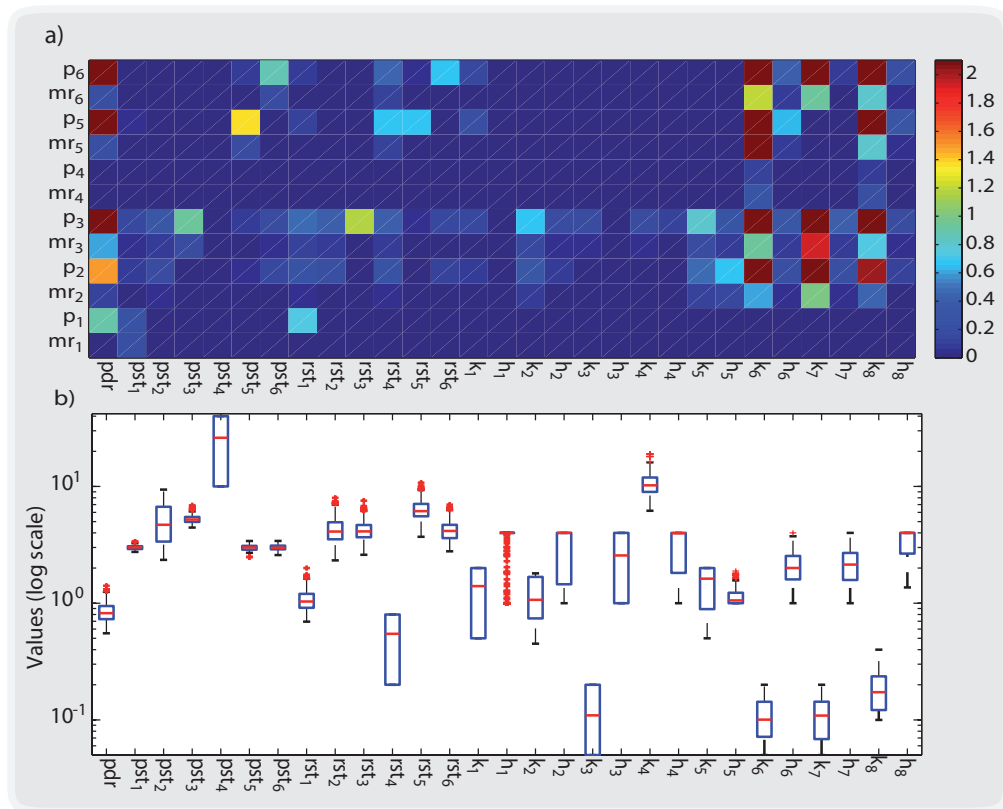


Figure 8: **Role of the experimental noise in the sloppiness and practical identifiability for the 6-gene regulatory network** a) Sensitivity matrix showing how some observables are almost not sensitive to parameter changes under wild type conditions (mr_1, mr_4 or p_4) and how some parameters do almost not influence model outputs ($pst_4, h_1, k_3, h_3, k_4$ or h_4). b) Distribution of optimal parameter values achieved out of 200 realizations of experimental data. Most of the parameters cannot be estimated with precision.

4.4. The role of optimal experimental design and sloppiness

The results presented in Section 4.3 have shown how the experimental setup influences sloppiness and practical identifiability. By appropriately choosing the number and location of the sampling times, the stimulation conditions, and (when possible) the measured outputs, it is possible to improve identifiability and reduce sloppiness. Note that, strictly speaking, the latter case (changing the observation function) involves redefining the model. Choosing the right experimental conditions is not easy, and may be counter intuitive. It is thus useful to apply optimal experimental design (OED) techniques, which automatically find the best experiment for a specific objective. The goal can be to improve identifiability, i.e. by minimizing the confidence intervals or the correlation among parameters, or, alternatively, to minimize sloppiness. Here we analyse the impact of experiment design, using Examples 2 and 3 from Section 3 as case studies. To this end we solve the associated D-optimum, E-optimum and S-optimum problems (see Section 2.6) under different experimental constraints and compare the results.

4.4.1. Example 2: biochemical pathway

For this example we assume that we have already performed the experiment labeled “Case 2b” in Section 4.3.2, that is, we have 8 measurements per each of the 7 observables (a total of 56 data points), obtained under sustained stimulation $u = 1$. We want to design a second experiment that complements the information provided by the first one. In this optimal experimental design task we consider the following degrees of freedom: the observed quantities with a maximum of 7, the stimulation conditions ($0 \leq u(t) \leq 2$), and the experiment durations ($15 \leq t_f \leq 25$). Sampling times are assumed to be equidistant, and the total amount of data is again $n^e \times 56$. Table 1 summarizes the results.

n^e / n_d	Criterion	C_F	$\rho_i(\%)$
2/112	D-opt	1.3×10^{-5}	15.9, <u>23.2</u> , 3.0, 5.0, 6.1, 6.1, 5.4, 6.2, 6.1, 6.9, 7.2, 6.9, 7.7, 8.0, 8.4, 12.1
	E-opt	4.8×10^{-5}	11.0, 12.2, <u>6.8</u> , 8.3, 9.9, 10.2, 9.3, <u>26.4</u> , 7.0, 8.6, 8.8, 7.9, 9.4, 9.8, 9.5, 13.6
	S-opt	6.6×10^{-5}	12.6, 14.2, <u>3.2</u> , 4.7, 6.1, 6.7, 6.5, 23.8, 8.6, <u>52.0</u> , 5.7, 7.4, 5.9, 6.1, 8.7, 12.1
3/168	D-opt	1.7×10^{-4}	5.1, 5.0, <u>2.2</u> , 4.0, 3.2, 3.1, 4.4, 3.1, 5.2, 3.5, 3.7, 6.1, 4.0, 4.1, 7.5, <u>10.6</u>
	E-opt	2.0×10^{-4}	4.8, 4.8, <u>2.1</u> , 4.0, 3.3, 3.2, 4.4, 3.3, 5.2, 3.7, 3.8, 6.1, 4.1, 4.3, 7.4, <u>10.5</u>
	S-opt	2.2×10^{-4}	5.2, 5.1, <u>2.1</u> , 4.0, 3.5, 3.4, 4.4, 3.5, 5.2, 3.9, 4.1, 6.0, 4.4, 4.6, 7.4, <u>10.4</u>

Table 1. Optimal design of experiments for the biochemical pathway example. First column: total number of experiments used (n^e , i.e. Case 2b plus the number of optimally designed experiments, and total amount of data (n_d). Second column: criterion used for optimal experimental design. Third column: sloppiness. Fourth column: relative confidence intervals for all parameters. Underlined values correspond to the minimum and maximum relative confidence interval in each scenario.

Recall that for experiment Case 2b the sloppiness C_F was in the order of 10^{-10} . The addition of a single experiment reduces the sloppiness by 5 orders of magnitude, in all experimental designs. Further reductions are achieved by adding a second experiment. However, the addition of further experiments does not lead to a non-sloppy scenario. Although the model is sloppy for all experimental designs, the confidence intervals are substantially reduced for all the parameters (see Figure 9), to the point that all of them can be considered practically identifiable. For example, with the *E-opt* experimental design the mean confidence interval is $\bar{\rho} = 4.7\%$ and the maximum is $\max(\rho_i) = 10.5\%$.

The final experimental designs with two designed experiments are shown in Figure 10. D-opt and E-opt criteria result in qualitatively similar experiments. In both cases the optimum choice is to measure $x_3, x_4, x_6, x_8, x_9, x_{11}, x_{12}$ (as already expected from the sensitivity analysis, Figure 7.c). The resulting feeding profiles are time-varying and complement the information obtained from experiment Case 2b, where there was a constant input $u = 1$; the new feeds are either clearly above or below 1. The S-opt design results in the measurement of $x_1, x_2, x_3, x_4, x_9, x_{11}, x_{12}$ in the first experiment and $x_3, x_4, x_6, x_8, x_9, x_{11}, x_{12}$ in the second. The feed is kept at low levels in both cases, even though the profiles lead to different dynamic behaviours of the system.

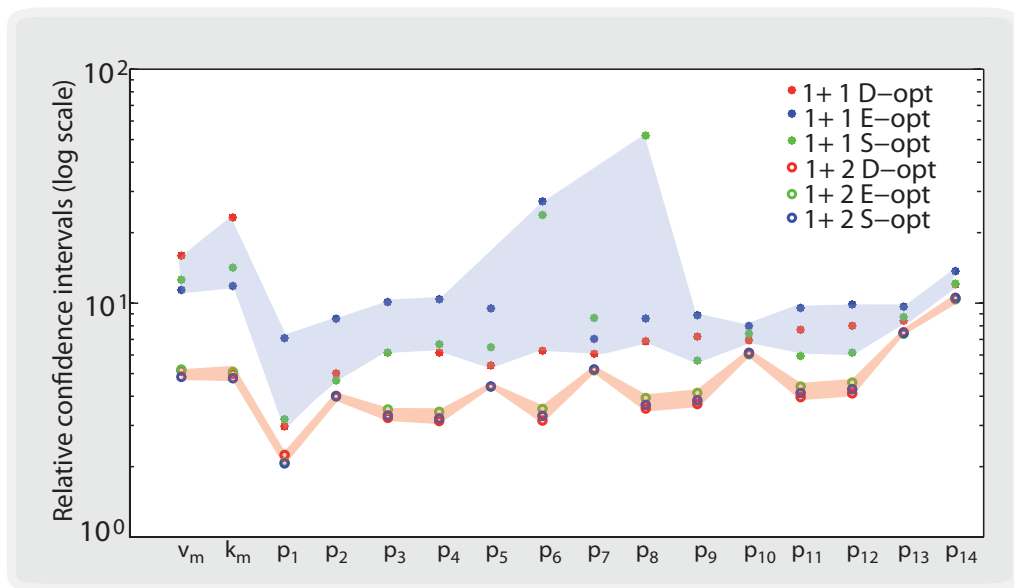


Figure 9: **Evolution of the relative confidence intervals with optimal experimental design for the biochemical pathway.** Stars: results obtained with the experiment “Case 2b” plus one optimally designed experiment. The use of different criteria for OED results in substantial differences in the confidence intervals (wide area shaded in blue). Circles: results of “Case 2b” plus two optimally designed experiments. All criteria converge to practically the same result (narrow area shaded in orange), given the experimental constraints.

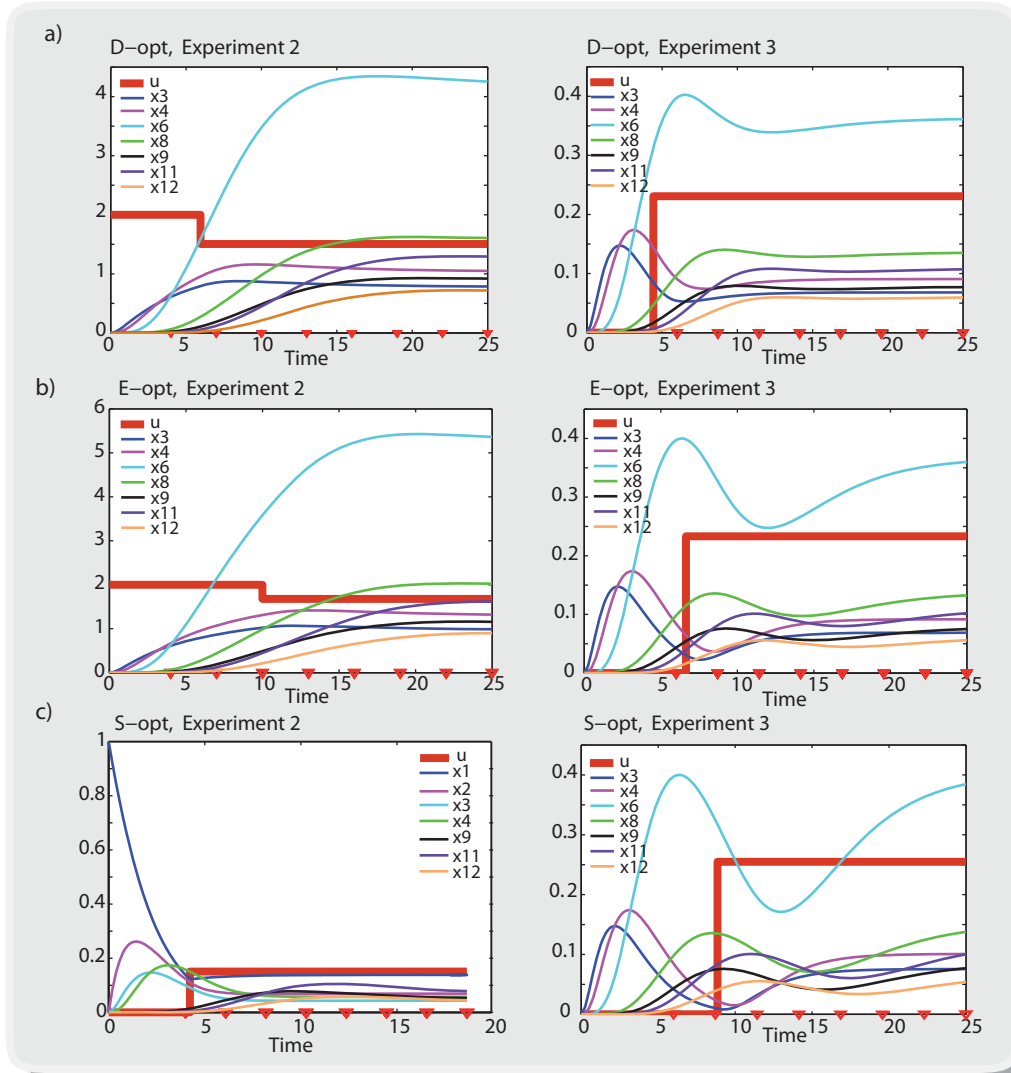


Figure 10: **Optimally designed experiments for the biochemical pathway example.** Experiments obtained with different criteria: a) D-opt, b) E-opt, c) S-opt. Lines represent stimulation conditions and measured states, triangles in the x-axis represent sampling times.

4.4.2. Example 3: six-gene regulatory network

Following the instructions of the DREAM6 parameter estimation challenge, we assume that the following perturbations can be implemented:

- GD: Gene deletion to eliminate both mRNA and protein production for a specified gene.
- KD: mRNA knockdown using siRNA to achieve a 5-fold increase in mRNA degradation.
- RBA: Increase of RBS activity by 100% (i.e. double) to increase translation rate.

For every perturbation we can measure either all mRNA time courses, by means of microarray measurements (M), or time courses of protein abundance for a maximum of 2 proteins per experiment, by means of fluorescent protein fusion (P). Measuring mRNA provides 21×6 data, whereas measuring protein yields only 41×2 data per experiment. In addition, a gel-shift assay (GS) can be chosen to determine the binding affinity (Kd) and Hill coefficient (h) of any transcription factor. A maximum budget of $20 \times C1$, where $C1$ is a unit cost, is allowed for buying experiments. The cost of each type of experiment is as follows:

- GD: $2 \times C1$, KD: $1 \times C1$ and RBA: $1 \times C1$
- M: $2 \times C1$, P: $1 \times C1$
- GS: $3 \times C1$

The sequences of experiments obtained with the D-optimum, E-optimum and S-optimum design criteria are shown in Table 2. The selected experiments vary significantly depending on the criteria. D-optimum design avoids gel shift assays, since those experiments reduce the dimension and thus the determinant of \mathcal{F} . The tendency is to use all other types of experiments and to measure all states, when enough budget is available. E-optimum design exploits a combination of gel shift assays with mRNA knockdown and gene deletion. S-optimum design combines gene deletion and gel shift assays.

Criterion	Sequence of experiments	$C_{\mathcal{F}}$	$\bar{\rho}$	$\max(\rho_i)$
D-opt	mRNA 5 knockdown, all mRNA measured	2.1×10^{-11}	2451%	$4.726 \times 10^4\%$
	Gene 1 deletion, all mRNAs measured	1.2×10^{-7}	50.5%	125%
	mRNA 5 knockdown, all proteins measured	7.8×10^{-8}	37.7%	122%
	Gene 1 deletion, all proteins measured	8.6×10^{-8}	32.5%	104%
	mRNA 1 knockdown, all mRNA measured	2.7×10^{-7}	26.6%	80.9%
	Increase of RBS 1 activity, all mRNAs measured	3.1×10^{-7}	24.3%	77.6%
	Increase of RBS 1 activity, p2 & p3 measured	3.2×10^{-7}	23.3%	76.1%
E-opt	mRNA 5 knockdown, all proteins measured	4.7×10^{-10}	481%	6987%
	Gel shift (h_1, k_1)	1.3×10^{-9}	418%	6897%
	mRNA 5 knockdown, all mRNAs measured	1.8×10^{-8}	111%	1721%
	Gene 1 deletion, all mRNAs measured	2.9×10^{-7}	29.6%	108%
	Gel shift (h_3, k_3)	3.7×10^{-7}	23.7%	64.9%
	Gel shift (h_4, k_4)	5.6×10^{-7}	21.3%	64.9%
Gene 1 deletion, p3 & p5 measured	8.3×10^{-7}	16.9%	57.8%	
S-opt	Gene 5 deletion, all proteins measured	7.3×10^{-9}	735%	$1.0 \times 10^4\%$
	Gene 1 deletion, all mRNAs measured	5.6×10^{-8}	67.1%	252%
	Gene 5 deletion, all mRNAs measured	6.1×10^{-7}	50.2%	215%
	Gel shift (h_1, k_1)	6.1×10^{-6}	36.3%	122%
	Gel shift (h_6, k_6)	1.3×10^{-5}	32.3%	122%
	Gene 1 deletion, p1, p2, p3 & p5 measured	2.0×10^{-5}	22.8%	80.5%

Table 2. Optimal design of experiments for the six-gene regulatory network. First column: criterion used for optimal experimental design. Second column: experiments. Third column: sloppiness. Fourth and fifth columns: mean and max relative confidence intervals, respectively.

Under the imposed constraints it was not possible to obtain a non-sloppy scenario. The best result in terms of identifiability was achieved with the E-optimum criterion; note that for this design the sloppiness is two orders of magnitude worse than for the S-optimum design.

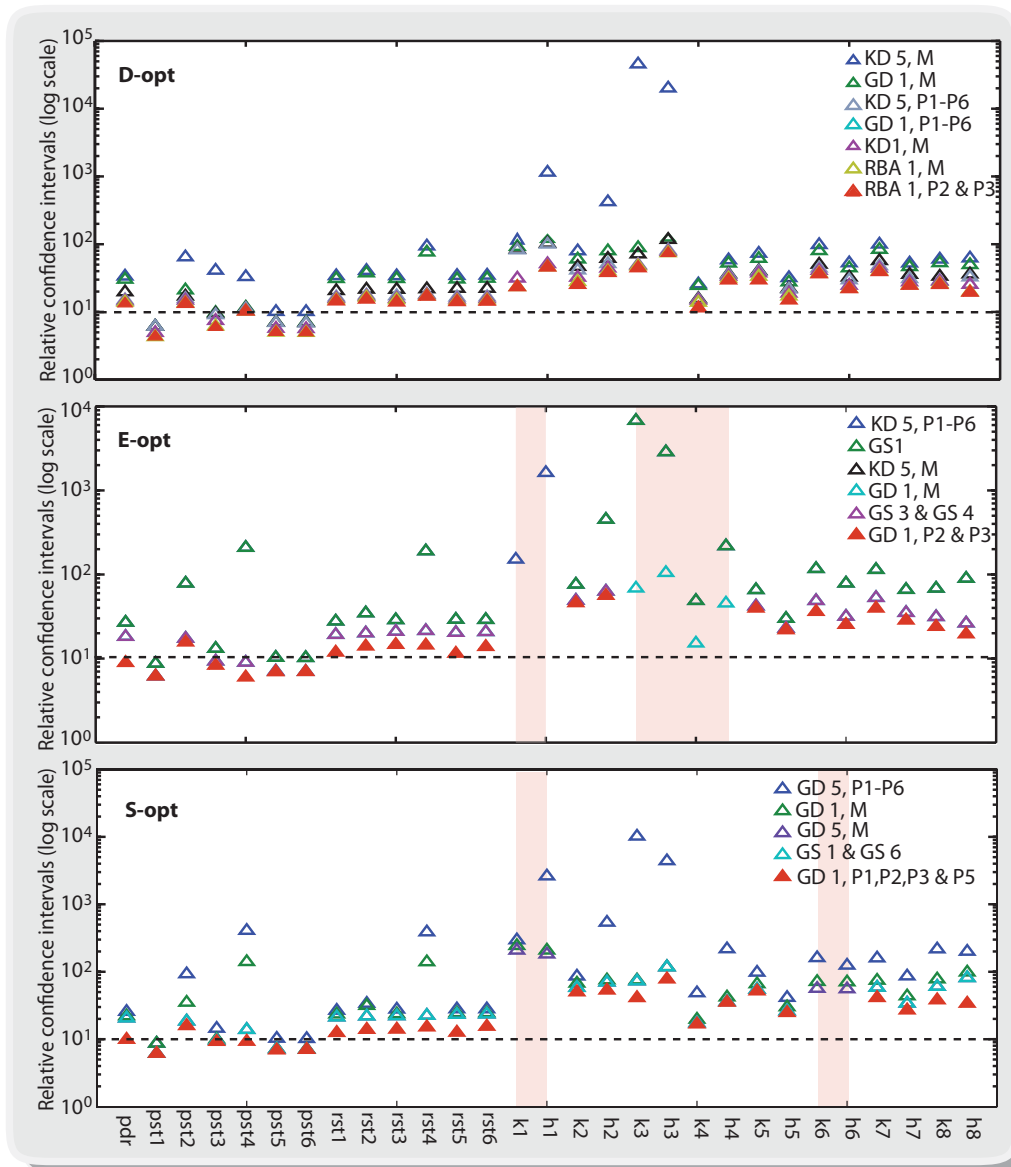


Figure 11: **Relative confidence intervals in OED for the six-gene regulatory network example.** The figures show the evolution of relative confidence intervals as new experiments are added using D-opt, E-opt and S-opt experimental designs. Shaded areas represent parameters measured by means of gel shift experiments, whose confidence intervals are 0%. Red triangles represent the best value achieved. Best overall result corresponds to the E-opt design, for which the confidence on the protein degradation rate is below 10%, for all promoter strengths and ribosomal binding sites strengths the relative confidence is around 10%, and for binding affinities and Hill coefficients it is below 60%.

5. Discussion and conclusions

The present work was motivated by the claim that sloppiness is an inherent property of systems biology models [22, 23]. Intrigued by this statement, we started by assessing to what extent “inherent” might be equivalent to “structural”. To this end we analysed the structural identifiability of several sloppy models with different characteristics of size and type of non-linearity. Results revealed that sloppiness is not equivalent to structural non-identifiability, since the models turned out to be structurally identifiable. This means that it is possible – at least in principle – to find unique values of sloppy parameters.

Since the origin of sloppiness is not – or at least not exclusively – related to the model structure (in fact, the dependence of sloppiness on data was already known [23]), it should be rooted elsewhere. Therefore our next question was: given a sloppy but structurally identifiable model, is it possible to determine the values of its parameters with reasonable confidence, for a realistic amount of noisy experimental data? In other words: can sloppy models be practically identifiable in reality? Practical identifiability depends on the experimental conditions (such as number of experiments, number of observables, number and location of sampling times, initial state of the system, and stimulation conditions) and on the type and amount of measurement noise. It is not uncommon to find in the recent literature the assumption that sloppiness is equivalent to practical non-identifiability, which sometimes leads to the conclusion that, since systems biology models are universally sloppy, they are also practically unidentifiable and, therefore, parameter estimation is doomed to fail, see e.g. [19]. However, we have found that sloppiness is not equivalent to lack of practical identifiability, although they are related. We have provided examples of sloppy models for which the confidence intervals of the parameters are reasonable, and thus the model may be considered practically identifiable – using feasible, realistic experimental setups.

As we have seen, sloppiness is neither equivalent to lack of structural nor practical identifiability. Thus the fact that a model is sloppy does not necessarily mean that its parameters cannot be estimated: sloppiness analysis cannot replace structural or practical identifiability analysis. Hence it is always advisable to analyse practical identifiability, for example by means of a non-parametric bootstrap approach, before making a decision on whether the parameters of a model can or cannot be estimated.

It is known that the practical identifiability of a model can be improved by means of an adequate experimental design [2, 4, 41, 24], and a similar procedure can be used to reduce sloppiness. Apgar and coworkers [2] analysed the evolution of sloppiness with the number of designed experiments, suggesting that sloppiness can decrease significantly following a careful experimental design. Later, in a reply to [2], Chandra and coworkers [14] objected that (i) even after experimental design the model remains sloppy, and (ii) the amount of data required is too high. We compared experiments designed with the aim of minimizing sloppiness with conventional D-optimum and E-optimum designs, whose goal is to improve confidence intervals. We found that the distribution of relative confidence intervals depends on the optimization criterion, as expected, and so does sloppiness. We observed that designs which minimize sloppiness do not end up in the best compromise between mean and maximum expected confidence on the parameters. The additive nature of the Fisher information matrix implies that the addition of informative data will increase its determinant and thus reduce the mean confidence intervals. However, this reduction does not necessarily modify the ratio between the minimum and the maximum eigenvalue (C_F), which is what defines sloppiness. This can be easily seen from a geometric perspective, as shown in Section 4.1. Taking this into account, we suggest using E-optimum or D-optimum designs to improve parameter estimates, instead of minimizing sloppiness. Regarding experimental

design, we also found that while the addition of new observables is very effective for reducing sloppiness, the amount of data is less critical than an adequate selection of sampling times.

The idea that a good fit to data is a guarantee for the validity of a model and, thus, for the validity of the corresponding predictions, although tempting, requires further analysis. Similarly, the idea that sloppy parameters can take almost any value without modifying model predictions also requires further analysis. We have seen here that this is not always the case, as shown e.g. in Figure 4. Hence parameter estimation remains a fundamental step in the model building process, and we argue that, despite the associated difficulties, a properly performed model calibration is an effort worth making.

Sloppiness is not a consequence of biological robustness, but a mathematical property of a given model and the data used for its calibration. While sloppiness has been linked to robustness [21], we argue that both concepts are only partially related. Robustness in biological behaviours with respect to different sources of noise can be associated with a certain insensitivity of models predictions to parameter values. However, whether this insensitivity is rooted in a lack of structural or practical identifiability makes a difference. We can distinguish three situations. The first one is lack of structural identifiability, in which case the model can be robust against parameter value modifications, but this has little or nothing to do with the biological robustness. The second situation is if a model is both structurally and practically identifiable, and sloppy. In this case the model may be robust to perturbations around the parameter values, and robustness can be assessed by the uncertainty associated with model predictions. Finally, if the model is sloppy but practically non-identifiable (or poorly identifiable), then the model (or, more precisely, some states in the model) may not be robust against parameter modifications. This situation would be, in principle, reverted if more informative experimental data becomes available.

The fact that sloppiness is not equivalent to structural nor practical identifiability (or to any other properties) indicates that it is not a redundant concept. This realization poses a new question: where exactly does the value of sloppiness reside? This point has so far been controversial. On the one hand, a number of researchers have expressed critical or sceptical views about the utility of the concept [2, 24, 43, 48]. On the other hand, it has been argued that sloppiness can not only explain why tools such as principal component analysis (PCA) and the Levenberg-Marquardt algorithm are effective, but also that it is the phenomenon that enables biological evolution; and, ultimately, that sloppiness is the reason why science is possible and the universe is comprehensible [46]. While these are certainly impressive claims, in the present work we have investigated a more specific topic, namely the identification of biological models. It is in this context that we inquire: what does sloppiness exactly contribute in order to be considered a pertinent addition to the set of conceptual tools employed by systems biology modellers? From our analyses, we conclude that what sloppiness brings to the table is of little use. Identifiability is a more informative concept, both in terms of the inverse and forward problems. Therefore, identifiability-based approaches are a better option for assessing the quality of parameter estimates and the predictive capabilities of calibrated models.

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 686282 (“CANPATHPRO”) and from the Spanish government (MINECO) and the European Regional Development Fund (ERDF) through the projects “SYNBIOFACTORY” (grant number DPI2014-55276-C5-2-R), and “IMPROWINE” (grant number AGL2015-67504-C3-2-R).

Supplementary material

Symbolic analysis of a simple linear model.

References

- [1] Antoulas, A. C., Sorensen, D. C., Gugercin, S., 2001. A survey of model reduction methods for large-scale systems. *Contemporary mathematics* 280, 193–220.
- [2] Apgar, J., Witmer, D., White, F., Tidor, B., 2010. Sloppy models, parameter uncertainty, and the role of experimental design. *Mol. BioSyst.* 6, 1890–1900.
- [3] Apri, M., de Gee, M., Molenaar, J., 2012. Complexity reduction preserving dynamical behavior of biochemical networks. *J. Theor. Biol.* 304(7), 16–26.
- [4] Balsa-Canto, E., Alonso, A., Banga, J., 2008. Computational procedures for optimal experimental design in biological systems. *IET Syst. Biol.* 2(4), 163–172.
- [5] Balsa-Canto, E., Alonso, A., Banga, J., 2010. An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Syst. Biol.* 4:11.
- [6] Balsa-Canto, E., Banga, J., 2011. AMIGO, a toolbox for advanced model identification in systems biology using global optimization. *Bioinformatics* 27(16), 2311–2313.
- [7] Banga, J. R., Balsa-Canto, E., 2008. Parameter estimation and optimal experimental design. *Essays Biochem.* 45, 195–210.
- [8] Bellman, R., Astrom, K. J., 1970. On structural Identifiability. *Math. Biosci.* 7, 329–339.
- [9] Berthomieux, S., Brilli, M., Kahn, D., de Jong, H., Cinquemani, E., 2012. On the identifiability of metabolic network models. *J. Math. Biol.*, DOI: 10.1007/s00258–012–0614–x.
- [10] Brown, K., Sethna, J.P., 2003. Statistical mechanical approaches to models with many poorly known parameters. *Phys Rev E Stat Nonlin Soft Matter Phys.* 68.
- [11] Cedersund, G., 2012. Conclusions via unique predictions obtained despite unidentifiability- new definitions and a general method. *FEBS J.* 279, 3513–3527.
- [12] Cedersund, G., 2016. Prediction uncertainty estimation despite unidentifiability: an overview of recent developments. In: *Uncertainty in Biology*. Springer, pp. 449–466.
- [13] Cedersund, G., Roll, J., 2009. Model based evaluation and comparison of potential explanations for given biological data. *FEBS J.* 276(4), 903–922.
- [14] Chandra, R., Transtrum, M., Sethna, J., 2011. Comment on Sloppy models, parameter uncertainty, and the role of experimental design. *Mol. Biosyst.* DOI: 10.1039/c1mb05046j.
- [15] Chatzis, M. N., Chatzi, E. N., Smyth, A. W., 2015. On the observability and identifiability of nonlinear structural and mechanical systems. *Struct. Control Health Monit.* 22 (3), 574–593.
- [16] Chen, W., Schoeberl, B., Jasper, P., Niepel, M., Nielsen, U., Lauffenburger, D., Sorger, P., 2009. Input output behavior of erbb signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Syst. Biol.* 5(239).
- [17] 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.
- [18] Chis, O., Banga, J., Balsa-Canto, E., 2011. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE* 6(11).
- [19] Cirit, M., Haugh, J. M., 2012. Data-driven modelling of receptor tyrosine kinase signalling networks quantifies receptor-specific potencies of pi3k-and ras-dependent erk activation. *Biochem. J.* 441 (1), 77–85.
- [20] Cobelli, C., DiStefano, J., 1980. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 239 (1), R7–R24.
- [21] Daniels, B. C., Chen, Y.-J., Sethna, J. P., Gutenkunst, R. N., Myers, C. R., 2008. Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* 19 (4), 389–395.
- [22] Gutenkunst, R., Casey, F., Waterfall, J., Myers, C., Sethna, J., 2007. Extracting falsifiable predictions from sloppy models. *Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference* *Annals of the New York Academy of Sciences* 1115 (203–211), 1115–203.
- [23] Gutenkunst, R., Waterfall, J., Casey, F., Brown, K., Myers, C., Sethna, J., 2007. Universally sloppy parameter sensitivities in systems biology models. *Plos Comput. Biol.* 3 (10), 1871–1878.
- [24] Hagen, D. R., White, J. K., Tidor, B., 2013. Convergence in parameters and predictions using computational experimental design. *Interface Focus* 3 (4), 20130008.
- [25] Hong, J., 1986. Optimal substrate feeding policy for a fed batch fermentation with substrate and product inhibition kinetics. *Biotechnol. and Bioeng.* 28, 1421–1431.

- [26] Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 24 (6), 417.
- [27] Jacquéz, J. A., Greif, P., 1985. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences* 77 (1), 201–227.
- [28] Jaqaman, K., Danuser, G., 2006. Linking data to models: data regression. *Nat. Rev. Mol. Cell Bio.* 7 (11), 813–819.
- [29] Joshi, M., Seidel-Morgenstern, A., Kremling, A., 2006. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab. Eng.* 8, 447–455.
- [30] Kirouac, D. C., Madlambayan, G. J., Yu, M., Sykes, E. A., Ito, C., Zandstra, P. W., 2009. Cell–cell interaction networks regulate blood stem and progenitor cell fate. *Molecular systems biology* 5 (1), 293.
- [31] Kotte, O., Heinemann, M., 2009. A divide-and-conquer approach to analyze underdetermined biochemical models. *Bioinformatics* 25(4), 519–525.
- [32] Liepe, J., Filippi, S., Komorowski, M., Stumpf, M., 2013. Maximizing the information content of experiments in systems biology. *PLoS Computational Biology* 9.
- [33] Machta, B., Chachra, B., Transtrum, M., Sethna, J., 2013. Parameter space compression underlies emergent theories and predictive models. *Science* 342, 604–607.
- [34] Mannakee, B. K., Ragsdale, A. P., Transtrum, M. K., Gutenkunst, R. N., 2016. Sloppiness and the geometry of parameter space. In: *Uncertainty in Biology*. Springer, pp. 271–299.
- [35] 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.
- [36] Moore, B., 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE transactions on automatic control* 26 (1), 17–32.
- [37] Pohjanpalo, H., 1978. System identifiability based on power-series expansion of solution. *Math. Biosci.* 41 (1-2), 21–33.
- [38] Raman, D. V., Anderson, J., Papachristodoulou, A., 2016. Delineating parameter unidentifiabilities in complex models. *arXiv preprint arXiv:1607.07705*.
- [39] Rand, D., 2008. Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation law. *J. R. Soc. Interface* 5, 59–69.
- [40] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., aug 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25 (15), 1923–9.
- [41] Raue, A., Kreutz, C., Maiwald, T., Klingmüller, U., Timmer, J., 2011. Addressing parameter identifiability by model-based experimentation. *IET Systems Biology* 5 (2), 120–130.
- [42] Sunnaker, M., Cedersund, G., Jirstrand, M., 2011. A method for zooming of nonlinear models of biochemical systems. *BMC Syst. Biol.* 5, 140.
- [43] Tönsing, C., Timmer, J., Kreutz, C., 2014. Cause and cure of sloppiness in ordinary differential equation models. *Physical Review E* 90 (2), 023303.
- [44] Transtrum, M., Machta, B. B., Sethna, J., 2010. Why are nonlinear fits to data so challenging? *Phys. rev. Lett.* 104, 060201.
- [45] Transtrum, M., Machta, B. B., Sethna, J., 2011. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E.* 83, 036701.
- [46] Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., Sethna, J. P., 2015. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of chemical physics* 143 (1), 010901.
- [47] Vanlier, J., Tiemann, C., Hilbers, P., van Riel, N., 2013. Parameter uncertainty in biochemical models described by ordinary differential equations. *Mathematical Biosciences* 246, 305–314.
- [48] Villaverde, A. F., Banga, J. R., 2014. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of The Royal Society Interface* 11 (91), 20130505.
- [49] Villaverde, A. F., Barreiro, A., 2016. Identifiability of large nonlinear biochemical networks. *MATCH Commun. Math. Comput. Chem.* 76 (2), 259–296.
- [50] Walter, E., Pronzato, L., 1997. *Identification of Parametric Models from Experimental Data*. Springer, Masson.
- [51] Wang, C., Cirit, M., Haugh, J., 2009. Pi3k-dependent cross-talk interactions converge with ras as quantifiable inputs integrated by erk. *Molecular Systems Biology* 5:246.
- [52] Waterfall, J., 2006. Universality in multiparameter fitting: sloppy models. Ph.D. thesis, Cornell University.
- [53] Waterfall, J., Casey, F., Gutenkunst, R., Brown, K., Myers, C., Brouwer, P., Elser, V., Sethna, J., 2006. Sloppy-model universality class and the vandermonde matrix. *Physical Review Letters* 97(15).
- [54] Willcox, K., Peraire, J., 2002. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal* 40 (11), 2323–2330.