

# Comparison of approaches for parameter identifiability analysis of biological systems

Andreas Raue<sup>1,2,\*</sup>, Johan Karlsson<sup>3</sup>, Maria Pia Saccomani<sup>4</sup>, Mats Jirstrand<sup>3</sup> and Jens Timmer<sup>1,5,6</sup>

<sup>1</sup>University of Freiburg, Institute for Physics, 79104 Freiburg, Germany, <sup>2</sup>Merrimack Pharmaceuticals Inc., 02139 Cambridge, MA, USA, <sup>3</sup>Fraunhofer-Chalmers Research Centre for Industrial Mathematics, Chalmers Science Park, SE-412 88 Göteborg, Sweden, <sup>4</sup>Department of Information Engineering, University of Padova, 35131 Padova, Italy, <sup>5</sup>BIOSS Centre for Biological Signalling Studies and <sup>6</sup>Zentrum für Biosystemanalyse (ZBSA), University of Freiburg, 79104 Freiburg, Germany

Associate Editor: Alfonso Valencia

## ABSTRACT

**Motivation:** Modeling of dynamical systems using ordinary differential equations is a popular approach in the field of Systems Biology. The amount of experimental data that are used to build and calibrate these models is often limited. In this setting, the model parameters may not be uniquely determinable. Structural or a priori identifiability is a property of the system equations that indicates whether, in principle, the unknown model parameters can be determined from the available data.

**Results:** We performed a case study using three current approaches for structural identifiability analysis for an application from cell biology. The approaches are conceptually different and are developed independently. The results of the three approaches are in agreement. We discuss strength and weaknesses of each of them and illustrate how they can be applied to real world problems.

**Availability and implementation:** For application of the approaches to further applications, code representations (DAISY, Mathematica and MATLAB) for benchmark model and data are provided on the authors webpage.

**Contact:** andreas.raue@fdm.uni-freiburg.de

Received on October 9, 2013; revised on December 3, 2013; accepted on December 31, 2013

## 1 INTRODUCTION

The dynamics of cellular processes such as signal transduction pathways can be described by models consisting of ordinary differential equations (ODEs). We used a model of IL13-induced JAK/STAT signaling (Raia *et al.*, 2011) to present a comprehensive comparison of three approaches for structural identifiability analysis. Mathematically, such dynamical models can be characterized by the ODE

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \quad \text{with} \quad \mathbf{x}(0) = \mathbf{x}_0 = \mathbf{g}(\boldsymbol{\theta}) \quad (1)$$

In this case, the vector of state variables  $\mathbf{x}(t)$  describes the dynamics of 15 molecular components involved in the JAK/STAT signaling pathway (for details see equation (3)). The function  $\mathbf{u}(t)$  represents possibly time-dependent

experimental treatments; in this case it represents a constant treatment with the cytokine IL13. The vector  $\boldsymbol{\theta}$  contains all parameters of the dynamical model, such as the reaction rate constants. For a specific cell type or biological context, the parameters  $\boldsymbol{\theta}$  are often not available from literature and have to be estimated from experimental data. The initial concentrations  $\mathbf{x}_0$  can be known or unknown; in the latter case they have to be estimated from experimental data as well. Each possible measurement is mathematically represented by a functional mapping

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \quad (2)$$

that might include additional parameters and thus increase the dimension of  $\boldsymbol{\theta}$ , such as scaling parameter for relative data obtained by immunoblotting in our case.

The inference problem of concern is the determination of the model parameters  $\boldsymbol{\theta}$  from the measurements  $\mathbf{y}$  and inputs  $\mathbf{u}$  described by Equation (2). In general, this may not be possible. *Structural or a priori identifiability* (Bellman and Åström, 1970) is a property of the systems (1) and (2) that guarantees that the unknown model parameters can, in principle, be determined from generic input and output functions of the model, provided they satisfy certain minimal conditions called ‘persistence of excitation’ [for a broad introduction and classical approaches Walter (1987)]. Structural identifiability is a prerequisite and necessary condition for any estimation procedure to render the recovery of  $\boldsymbol{\theta}$  from input–output measurements as a well-posed problem and to return meaningful results about  $\boldsymbol{\theta}$ . Well-known approaches to structural identifiability analysis often have problems with the realistically sized models (Karlsson *et al.*, 2012; Sedoglavic, 2002), but there are approaches that can handle them.

Let us consider a simple model for the illustration of structural identifiability. The model describes a transition between two components

$$\begin{cases} \dot{x}_1(t) &= -\theta_1 x_1 \\ \dot{x}_2(t) &= +\theta_1 x_1 \end{cases}$$

We assume that  $x_1(0) = \theta_2$  and  $x_2(0) = 0$ , and only the measurement  $y_1(t) = \theta_3 x_2(t)$ , are available. In this case, we can solve for

\*To whom correspondence should be addressed.

**Table 1.** Classification of the investigated approaches

Approach	Identifiability	Scope	Type	Runtime	Main reference
DAISY	Structural	Global	A priori	~120 min	Saccomani <i>et al.</i> (2003)
EAR	Structural	Local (around a generic point)	A priori	~30 s	Karlsson <i>et al.</i> (2012)
PL	Structural and practical	Local (region covered by profile)	Data-based	~30 min	Raue <i>et al.</i> (2009)

Note: 'Runtime' refers to one complete analysis for the case of Raia *et al.* (2011) using a standard desktop computer.

$x_2(t)$  analytically and substitute into the measurement equation. The result

$$y_1(t) = \theta_3 \theta_2 \exp(-\theta_1 t) \cdot (\exp(\theta_1 t) - 1)$$

shows that  $\theta_3$  and  $\theta_2$  are *structurally non-identifiable*.  $\theta_1$  and only the product of  $\theta_3$  and  $\theta_2$  are structurally identifiable. Such redundant parameterization can be detected directly for very simple examples. For analysis of realistic models, one has to resort to more sophisticated approaches.

We compared three current conceptually different approaches for identifiability analysis using the model and data of Raia *et al.* (2011) as case study. In particular, we compared the Differential Algebra Identifiability of Systems (DAISY) approach proposed by Saccomani *et al.* (2003), the Exact Arithmetic Rank (EAR) approach implemented by Karlsson *et al.* (2012) and the Profile Likelihood (PL) approach proposed by Raue *et al.* (2009). The results of all three approaches are in good agreement; however, each approach has specific strength and weaknesses that will be discussed.

## 2 METHODS

Approaches for identifiability analysis can be classified according to several criteria. The main difference is between a priori versus data-based type approaches. A priori approaches can be applied irrespective of which input functions are used and before the availability of experimental data. Data-based type approaches can be applied if actual experimental data are available or can be simulated under reasonable assumptions. Some a priori approaches allow to test *global* identifiability, a property holding for all possible parameter values, i.e. independently of the actual parameter value. Other approaches allow to test *local* identifiability, holding around a point in the parameter space. Some data-based approaches also allow for conclusions about *practical* non-identifiability (Raue *et al.*, 2009) that is caused by limited quality of experimental data. The classification of the three approaches investigated is shown in Table 1.

### 2.1 DAISY approach

This approach implements a differential algebra algorithm to perform a global parameter identifiability analysis for dynamic models described by polynomial or rational equations (Bellu *et al.*, 2007). The basic idea is that of manipulating algebraic differential equations as polynomials depending also on derivatives of the variable. Ritt's algorithm permits to eliminate the non-observed state variables  $x$  from the system of equations and to find the *input-output relation* of the system: a set of polynomial differential equations involving only the variables  $u$  and  $y$ , thus describing all input-output pairs satisfying (1) and (2). The input-output relation is linearly parameterized by certain algebraic functions of the unknown parameters called the *exhaustive summary*, which can be

easily extracted. These functions lead to a system of algebraic non-linear equations in the unknown  $\theta$ . By applying a computer algebra algorithm, i.e. the Buchberger algorithm, it is possible to check whether there is one or multiple solutions and hence distinguish between global, or local identifiability or non-identifiability of the original dynamic system. An additional advantage of using this computer algebra tool is that it does not require expertise on mathematical modeling by the experimenter.

### 2.2 EAR approach

This approach is based on applying the inverse function theorem to the system of algebraic equations relating higher order derivatives of the output  $y$  with respect to time at the initial time with the initial state and parameters (Pohjanpalo (1978). Using a differential algebra approach, an upper bound of the order of differentiation can be given, resulting in a non-linear algebraic system of equations in the parameters. The rank of the Jacobian matrix for this system of equations gives information about its solvability, and in case of a rank-deficient matrix, a more detailed analysis of the Jacobian provides information about which parameters are involved in relations rendering the system non-identifiable. The EAR method provides means to efficiently compute the generic rank of the Jacobian matrix to return a conclusive result if the system is structurally identifiable. It considers local identifiability but around a generic point, i.e. the computations are carried out for a random specialization of the unknown parameters, and initial state to integer values and a random input in terms of a truncated integer coefficient power series. Local structural identifiability is an *almost everywhere* property by definition, i.e. it holds everywhere apart from possibly on a set of measure zero. The EAR approach is based on a method for local algebraic observability (Sedoglavic, 2002). The EAR analysis is implemented as a fully automatic Mathematica function. The user simply inputs the equations and gets the answer. Based on the EAR identifiability analysis, it is also possible to find minimal sets of outputs giving identifiability (Anguelova *et al.*, 2012).

### 2.3 PL approach

This approach checks for non-identifiability by posing a parameter estimation problem using real or simulated data. The central idea is that non-identifiability manifests as a flat manifold in the parameter space of the estimation problem, e.g. the likelihood function. A profile can be calculated for each parameter  $\theta_i$  individually by repeated optimization of all parameters  $\{\theta_j | j \neq i\}$  for a series of fixed values of the parameter  $\theta_i$ . A flat profile indicates a structurally non-identifiable parameter. For detecting structural non-identifiability, simulated data are sufficient. In case real experimental data are available, practical non-identifiability can also be detected and confidence intervals for the parameter estimates can be calculated. The traces in parameter space that correspond to the profiles can be used to analyze the reason for non-identifiability and point to missing experimental information and interrelated parameters. It was demonstrated in a study by Raue *et al.* (2010) that the approach facilitates an iterative experimental design strategy and Kreutz *et al.* (2012)

extended the approach to detect non-observability of the dynamics directly. The approach was applied to biological data in Bachmann *et al.* (2011) and Becker *et al.* (2010).

## 2.4 Equations of benchmark model

In the following, we describe the equations of the IL13-Induced JAK/STAT signaling model (Raia *et al.*, 2011) that correspond to Equations (1) and (2). These equations determine the structural identifiability of the model parameters. The ODE system determining the time evolution of the state variables is given by

$$\begin{cases} \dot{x}_1(t) = -\theta_1 c_1 u_1 x_1 - \theta_5 x_1 + \theta_6 x_2 \\ \dot{x}_2(t) = +\theta_5 x_1 - \theta_6 x_2 \\ \dot{x}_3(t) = +\theta_1 c_1 u_1 x_1 - \theta_2 x_3 x_7 \\ \dot{x}_4(t) = +\theta_2 x_3 x_7 - \theta_3 x_4 \\ \dot{x}_5(t) = +\theta_3 x_4 - \theta_4 x_5 \\ \dot{x}_6(t) = -\theta_7 x_3 x_6 / (1 + \theta_{13} x_{14}) - \\ \quad -\theta_7 x_4 x_6 / (1 + \theta_{13} x_{14}) + c_2 \theta_8 x_7 \\ \dot{x}_7(t) = +\theta_7 x_3 x_6 / (1 + \theta_{13} x_{14}) + \\ \quad +\theta_7 x_4 x_6 / (1 + \theta_{13} x_{14}) - c_2 \theta_8 x_7 \\ \dot{x}_8(t) = -\theta_9 x_8 x_7 + c_2 \theta_{10} x_9 \\ \dot{x}_9(t) = +\theta_9 x_8 x_7 - c_2 \theta_{10} x_9 \\ \dot{x}_{10}(t) = +\theta_{11} x_9 \\ \dot{x}_{11}(t) = -\theta_{12} c_1 u_1 x_{11} \\ \dot{x}_{12}(t) = +\theta_{12} c_1 u_1 x_{11} \\ \dot{x}_{13}(t) = +\theta_{14} x_{10} / (\theta_{15} + x_{10}) - \theta_{16} x_{13} \\ \dot{x}_{14}(t) = +\theta_{17} x_9 \end{cases} \quad (3)$$

where  $\mathbf{x}$  is the 14D state vector, the dot denotes the time derivative,  $u_1$  is an input functions and  $c_{1,2}$  are constants. The initial conditions of the state variables are

$$\mathbf{x}(0) = [1.3, \theta_{23}, 0, 0, 0, 2.8, 0, 165, 0, 0, 0.34, 0, 0, 0] \quad (4)$$

the values are given in units of molecules per cell ( $\times 1000$ ). The set of measurement equations is defined by the following equation

$$\begin{cases} y_1(t) = x_1 + x_3 + x_4 \\ y_2(t) = \theta_{18}(x_3 + x_4 + x_5 + x_{12}) \\ y_3(t) = \theta_{19}(x_4 + x_5) \\ y_4(t) = \theta_{20} x_7 \\ y_5(t) = \theta_{21} x_{10} \\ y_6(t) = \theta_{22} x_{14} \\ y_7(t) = x_{13} \\ y_8(t) = x_9. \end{cases} \quad (5)$$

The components  $\mathbf{x}$ ,  $\mathbf{u}$  and  $\mathbf{c}$  of Equation (3) and  $\mathbf{y}$  of Equation (5) are described in Table 2. The model Equations (3–5) contain 23 unknown parameters  $\theta$  that are described in Table 3. They need to be determined from the available measurements given in Equation (5).

## 3 RESULTS

We present results on the identifiability for the benchmark model (Raia *et al.*, 2011) using the three approaches described in the Section 2: the Differential Algebra Identifiability of Systems (DAISY) approach proposed by Bellu *et al.* (2007); EAR approach implemented by Karlsson *et al.* (2012) and the PL approach proposed by Raue *et al.* (2009). For the last approach, we also use the original data of Raia *et al.* (2011). The application of the three approaches for identifiability analysis will be described in detail. In summary, all three approaches consistently classify five parameters as structurally non-identifiable:  $\theta_{11}$ ,  $\theta_{15}$ ,  $\theta_{17}$ ,  $\theta_{21}$ ,  $\theta_{22}$ .

## 3.1 DAISY approach

The differential-algebra-based approach provides a direct check of global identifiability of the above model, showing the non-identifiability of some model parameters. The approach suggests a reduction of the model to minimal form so that fundamental system theoretic properties, such as accessibility, hold and the model is more suitable for further mathematical investigations. Thus, before checking model identifiability from the designed experiment, it is convenient to check its minimality. In this case, just by visual inspection, it is easy to see that some of the 14 model equations defining the model are redundant.

In particular, the equations for  $x_6$ ,  $x_8$ ,  $x_{11}$  are the same as those for  $x_7$ ,  $x_9$ ,  $x_{12}$  with the opposite sign. For example, from the eighth and ninth equations, one obtains  $\dot{x}_9 = -\dot{x}_8$ . By integrating and using the known initial conditions, one arrives at  $x_9 = -x_8 + 165$ , thus eliminating one differential equation. Same procedure is followed for  $x_7$ ,  $x_{12}$ . Finally,  $x_{14}$  can be expressed as  $x_{14} = x_{10}\theta_{17}/\theta_{11}$ , and the last equation is also redundant. The model can then be rewritten in a simplified form involving only 10 state variables. This is done not only for the sake of mathematical simplification, but also to satisfy some important structural property, such as *minimality* and *accessibility* (Saccomani *et al.*, 2003). This is important, because the lack of minimality of the model may lead to spurious non-identifiability results for some parameters that may not occur in a minimal model. Also, if the model has two or more differential equations dependent on each other, as in this case, the model would not be *accessible*, making more difficult the identifiability check of the model from the given initial conditions.

One thus arrives at the following simplified 10D model:

$$\begin{cases} \dot{x}_1(t) = -\theta_1 x_1 c_1 u_1 - \theta_5 x_1 + \theta_6 x_2 \\ \dot{x}_2(t) = \theta_5 x_1 - \theta_6 x_2 \\ \dot{x}_3(t) = \theta_1 c_1 u_1 x_1 - \theta_2 x_3 (-x_6 + 2.8) \\ \dot{x}_4(t) = \theta_2 x_3 (-x_6 + 2.8) - \theta_3 x_4 \\ \dot{x}_5(t) = \theta_3 x_4 - \theta_4 x_5 \\ \dot{x}_6(t) = -\theta_7 x_3 x_6 / (1 + \theta_{13} x_{13}) - \\ \quad -\theta_7 x_4 x_6 / (1 + \theta_{13} x_{13}) + \theta_8 (-x_6 + 2.8) c_2 \\ \dot{x}_8(t) = -\theta_9 x_8 (-x_6 + 2.8) + \theta_{10} (-x_8 + 165) c_2 \\ \dot{x}_{10}(t) = \theta_{11} (-x_8 + 165) \\ \dot{x}_{11}(t) = -\theta_{12} c_1 u_1 x_{11} \\ \dot{x}_{13}(t) = x_{10} \theta_{14} / (\theta_{15} + x_{10}) - \theta_{16} x_{13} \end{cases} \quad (6)$$

with the corresponding initial conditions given in Equation (4).

To check global identifiability of this simplified 10D model with DAISY, the user has to write in the input file the ordered list of the output and state variables, the list of the unknown parameters, the model equations and the known initial conditions. Later in the text, the results will be illustrated. For explanations of the technical terms, one may consult Bellu *et al.* (2007).

DAISY automatically ranks the input, output, state variables and their derivatives, starts the pseudodivision algorithm, i.e. the Ritt algorithm, and calculates the *characteristic set* of the model. This is a minimal set of differential polynomials, which provides an equivalent description of the model. The subset made of the first eight (i.e. the number of model outputs) differential polynomials does not depend on the state variable  $\mathbf{x}$  and provides the so-called *input-output relation* of the model. In particular, these involve higher derivatives of the input and output signals.

**Table 2.** List of mathematical variables, their biological short names and their meaning

$x_1$	Rec	Receptor for IL-13
$x_2$	Rec_i	Internalized receptor for IL-13
$x_3$	IL13_Rec	IL-13 receptor complex
$x_4$	p_IL13_Rec	Phosphorylated IL-13 receptor complex
$x_5$	p_IL13_Rec_i	Internalized phosphorylated IL-13 receptor complex
$x_6$	JAK2	Janus kinase 2
$x_7$	pJAK2	Phosphorylated Janus kinase 2
$x_8$	STAT5	Signal Transducer and Activate of Transcription 5
$x_9$	pSTAT5	phosphorylated Signal Transducer and Activate of Transcription 5
$x_{10}$	SOCS3Mrna	Suppressor of cytokine signaling 3 mRNA
$x_{11}$	DecoyR	Decoy receptor for IL-13
$x_{12}$	IL13_DecoyR	IL-13 decoy receptor complex
$x_{13}$	SOCS3	Suppressor of cytokine signaling 3 protein
$x_{14}$	CD274Mrna	Cluster of Differentiation 274 mRNA
$c_1$		Conversion factor from ng/ml to molecules per cell ( $\times 1000$ ) with value 2.265
$c_2$	SHP1	Constant intracellular concentration of protein tyrosine phosphatase SHP-1 with value 91 molecules per cell ( $\times 1000$ )
$u_1$	IL13	Constant extracellular concentration of Interleukine-13 0, 4, 20, 80 ng/ml
$y_1$	RecSurf_obs	IL-13 receptor at cell membrane
$y_2$	IL13_cell_obs	IL-13 at cell membrane or intracellular
$y_3$	pIL4Ra_obs	Phosphorylated IL-13 receptor at membrane or intracellular
$y_4$	pJAK2_obs	Phosphorylated Janus kinase 2
$y_5$	SOCS3mRNA_obs	Suppressor of cytokine signaling 3 mRNA
$y_6$	CD274mRNA_obs	Cluster of Differentiation 274 mRNA
$y_7$	SOCS3_obs	Suppressor of cytokine signaling 3 protein
$y_8$	pSTAT5_obs	phosphorylated Signal Transducer and Activate of Transcription 5

**Table 3.** List of model parameters, their biological short names, identifiability and confidence intervals

	Biological short name	Identifiability	MLE point	Likelihood-based confidence interval
$\theta_1$	Kon_IL13Rec	Identifiable	-3.087	[-3.261, -2.882]
$\theta_2$	Rec_phosphorylation	Identifiable	-1.185	[-1.383, -0.946]
$\theta_3$	pRec_intern	Practically non-identifiable	+2.236	[+0.037, $+\infty$ ]
$\theta_4$	pRec_degradation	Practically non-identifiable	+1.268	[-0.203, $+\infty$ ]
$\theta_5$	Rec_intern	Identifiable	-0.995	[-1.193, -0.783]
$\theta_6$	Rec_recycle	Identifiable	-2.225	[-2.700, -1.920]
$\theta_7$	JAK2_phosphorylation	Identifiable	+0.172	[-0.067, +0.442]
$\theta_8$	pJAK2_dephosphorylation	Identifiable	-2.788	[-3.283, -2.352]
$\theta_9$	STAT5_phosphorylation	Identifiable	-1.678	[-1.835, -1.472]
$\theta_{10}$	pSTAT5_dephosphorylation	Identifiable	-3.568	[-3.777, -3.393]
$\theta_{11}$	SOCS3mRNA_production	Structurally non-identifiable	-1.260	[- $\infty$ , $+\infty$ ]
$\theta_{12}$	DecoyR_binding	Practically non-identifiable	-5.000	[- $\infty$ , -4.131]
$\theta_{13}$	JAK2_p_inhibition	Identifiable	-2.042	[-2.483, -1.552]
$\theta_{15}$	SOCS3_accumulation	Structurally non-identifiable	+2.059	[- $\infty$ , $+\infty$ ]
$\theta_{14}$	SOCS3_translation	Practically non-identifiable	+1.081	[+0.392, $+\infty$ ]
$\theta_{16}$	SOCS3_degradation	Practically non-identifiable	-1.442	[- $\infty$ , $+\infty$ ]
$\theta_{17}$	CD274mRNA_production	Structurally non-identifiable	-1.786	[- $\infty$ , $+\infty$ ]
$\theta_{18}$	scale_IL13_cell_obs	Identifiable	+1.259	[+1.083, +1.416]
$\theta_{19}$	scale_pIL4Ra_obs	Practically non-identifiable	+3.000	[+1.491, $+\infty$ ]
$\theta_{20}$	scale_pJAK2_obs	Identifiable	-0.026	[-0.143, +0.145]
$\theta_{21}$	scale_SOCS3mRNA_obs	Structurally non-identifiable	+0.194	[- $\infty$ , $+\infty$ ]
$\theta_{22}$	scale_CD274mRNA_obs	Structurally non-identifiable	-1.965	[- $\infty$ , $+\infty$ ]
$\theta_{23}$	init_Rec_i	Identifiable	+1.384	[+1.156, +1.766]

Note: Parameter values are given on a  $\log_{10}$ -scale and are allowed to vary between -5 and +3.



To speed up the algorithm, it may be advisable to add derivatives of the actual output functions to the system equations. This is legitimate, as Ritt algorithm is based on differentiation, besides the usual algebraic operations.

After a suitable normalization, the input–output polynomials can be rendered monic, and their coefficients provide a set of rational functions of the unknown parameter  $\theta$ , which form the so-called *exhaustive summary* of the model.

Identifiability is tested by checking injectivity of the exhaustive summary with respect to the parameter  $\theta$ . We should, in principle, calculate the range set of these functions. This could be done in symbolic language evaluating them at a symbolic parameter value. In practice, but only to speed up the process, instead of choosing a symbolic parameter value, we can use a set of randomly chosen numerical points in the range set. Solution of these algebraic equations is done by computing a Gröbner basis, by applying the Buchberger algorithm. The results show that all the parameters are uniquely identifiable except for  $\theta_{11}, \theta_{15}, \theta_{17}, \theta_{21}$  and  $\theta_{22}$ , which have an infinite number of solutions. Thus, the model is non-identifiable.

The analysis actually provides some hint on how to simplify the model to make it globally identifiable. For example, by assigning known values to parameters  $\theta_{21}$  and  $\theta_{22}$ , the model would become globally identifiable.

Further information about DAISY and instructions on how to obtain it can be found at: <http://www.dei.unipd.it/wdyn/?IDsezione=4364>.

### 3.2 EAR approach

The identifiability of a dynamic model is closely related to the properties of the Jacobian matrix containing the derivatives of signals assumed to be measured (i.e. model outputs) and their time derivatives, with respect to the parameters. Furthermore, structural identifiability is a generic property of the symbolic form of the system and measurement equations, and hence it is sufficient to analyze this property for a specific (generic) point in parameter space. In the EAR approach, this is utilized using exact modular integer arithmetics for fast computations, i.e. the specialization of parameter values to random integers only serves the purpose of fast computation of structural properties and has nothing to do with biological feasibility. The direct approach of first deriving the entries of the Jacobian matrix in symbolic form, inserting integer values and then computing the matrix rank is not feasible for anything but very small systems, because of extensive swell of the size of symbolic expressions. Instead, it can be shown that the numerical values of the entries of the Jacobian matrix can be computed efficiently by computing power series solutions to the original ODEs augmented by their corresponding parametric sensitivity differential equations, followed by insertion into the output sensitivity expressions the obtained truncated power series solutions of the state and state sensitivities. To prevent the need for computation with rational numbers and the inherent swell in size of numerators and denominators, all computations are carried out modulo a large prime. To summarize, the above approach is based on exact (modular) arithmetics for obtaining the entries in the Jacobian as well as for the subsequent rank computation; hence, the name EAR. It is implemented in terms of a fully documented

Mathematica application and is completely automatic once a specific system description has been provided.

Here is an outline of the steps of the algorithm

- (1) Identifiability is a generic property of the symbolic form of the system and measurement equations. Therefore, any generic point can be analyzed. First, specializations of parameters and initial conditions are generated.
- (2) Specialize inputs to truncated random integer coefficient power series.
- (3) Truncated power series solutions of  $x$ ,  $\frac{\partial x}{\partial x(0)}$  and  $\frac{\partial x}{\partial \theta}$  are computed, using the system and sensitivity system equations.
- (4) The power series are inserted in the expressions for the derivatives of the outputs with respect to the initial state and with respect to the parameters

$$\frac{d}{dx(0)}y = \frac{d}{dx(0)}h = \frac{\partial h}{\partial x} \frac{\partial x}{\partial x(0)} \quad (7)$$

$$\frac{d}{d\theta}y = \frac{d}{d\theta}h = \frac{\partial h}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial h}{\partial \theta} \quad (8)$$

resulting in power series representations of the output sensitivity derivatives.

- (5) Identification of the coefficients of the truncated power series of the output derivatives with the coefficients of a general Taylor expansion of the output sensitivity derivatives gives the higher order time-derivatives of the output sensitivity derivatives, i.e. the entries of the specialized Jacobian matrix.
- (6) Calculate the rank.
- (7) If the matrix is rank-deficient, the non-identifiable parameters are found using the fact that removing the corresponding columns from the matrix do not change the rank.

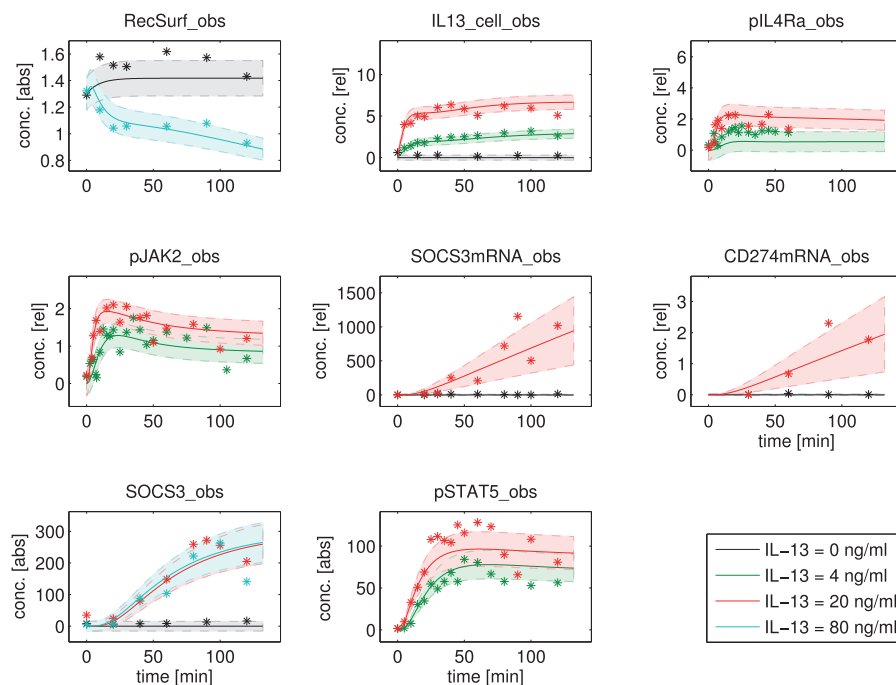
The analysis using the Mathematica package is fully automatic. Further information about the package and instructions on how to obtain it can be found at: <http://www.fcc.chalmers.se/sys/products/identifiabilityanalysis>.

Included in the package is, apart from the identifiability test demonstrated above, also functionality for automatically finding minimal sets of output expressions that guarantee a structurally identifiable model. This functionality is described in Anguelova *et al.* (2012).

### 3.3 PL approach

The PL approach determines the identifiability of the model parameters by posing a parameter estimation problem. Here, we use the original data of Raia *et al.* (2011) and investigate both the structural and practical identifiability of the model parameters. For parameter estimation, maximum likelihood estimation (MLE) is applied. The likelihood  $L(\theta)$  describes the probability of the data given certain parameter values  $\theta$ . The MLE fit of the dynamic model to the experimental data for MedB-1 cell is shown in Figure 1, and the MLE parameter values are given in Table 3. Likelihood profiles were calculated as described in Raue *et al.* (2009) by

$$PL(\theta_i) = \max_{j \neq i} [L(\theta_j)] \quad (9)$$



**Fig. 1.** Fit of the dynamic model given in Section 2.4 to the original data of Raia *et al.* (2011) for lymphoma-derived MedB-1 cells. The cells were treated with four different doses of IL-13. The data were obtained by Immunoblot and qRT-PCR measurements. Some measurements are on a relative concentration scale given in arbitrary units [au] and some on an absolute concentration scale [abs] that corresponds to molecules per cell ( $\times 1000$ )

where for each fixed value of parameter  $\theta_i$ , all other parameters  $\theta_j \neq \theta_i$  are reoptimized. Figure 2 shows the results of the analysis. A perfectly flat profile indicates a structural non-identifiable parameter. Perfectly flat profiles reveal the five structural non-identifiable parameters. The change of the parameters  $\theta_j$  along a profile of a structurally non-identifiable parameter  $\theta_i$  can be used to determine functionally related groups between the structurally non-identifiable parameters (Fig. 3). Here, the five structurally non-identifiable parameters are functionally related in two groups. In the first group,

$$\theta_{17} \sim \frac{1}{\theta_{22}}$$

indicates that the concentration scale of CD274 mRNA ( $x_{14}$ ) is not fixed by measurements. In the second group,

$$\theta_{15} \sim \frac{1}{\theta_{21}} \quad \text{and} \quad \theta_{15} \sim \theta_{11}$$

indicates that the concentration scale of SOCS3 mRNA ( $x_{10}$ ) is not fixed by measurements. Using this information, two new experiments that determine the respective concentration scales can be used to resolve these structural non-identifiabilities.

The likelihood profiles can also be used to assess practical identifiability and to calculate confidence intervals of the model parameters. A threshold in the likelihood, measured from the MLE point, can be used to compute likelihood-based confidence intervals [for details on the statistics, see in Raue *et al.* (2009)]. Profiles that have a unique minimum, but do not cross the confidence threshold, reveal six practically non-identifiable parameters:  $\theta_3$ ,  $\theta_4$ ,  $\theta_{12}$ ,  $\theta_{14}$ ,  $\theta_{16}$  and  $\theta_{19}$ . The likelihood profiles

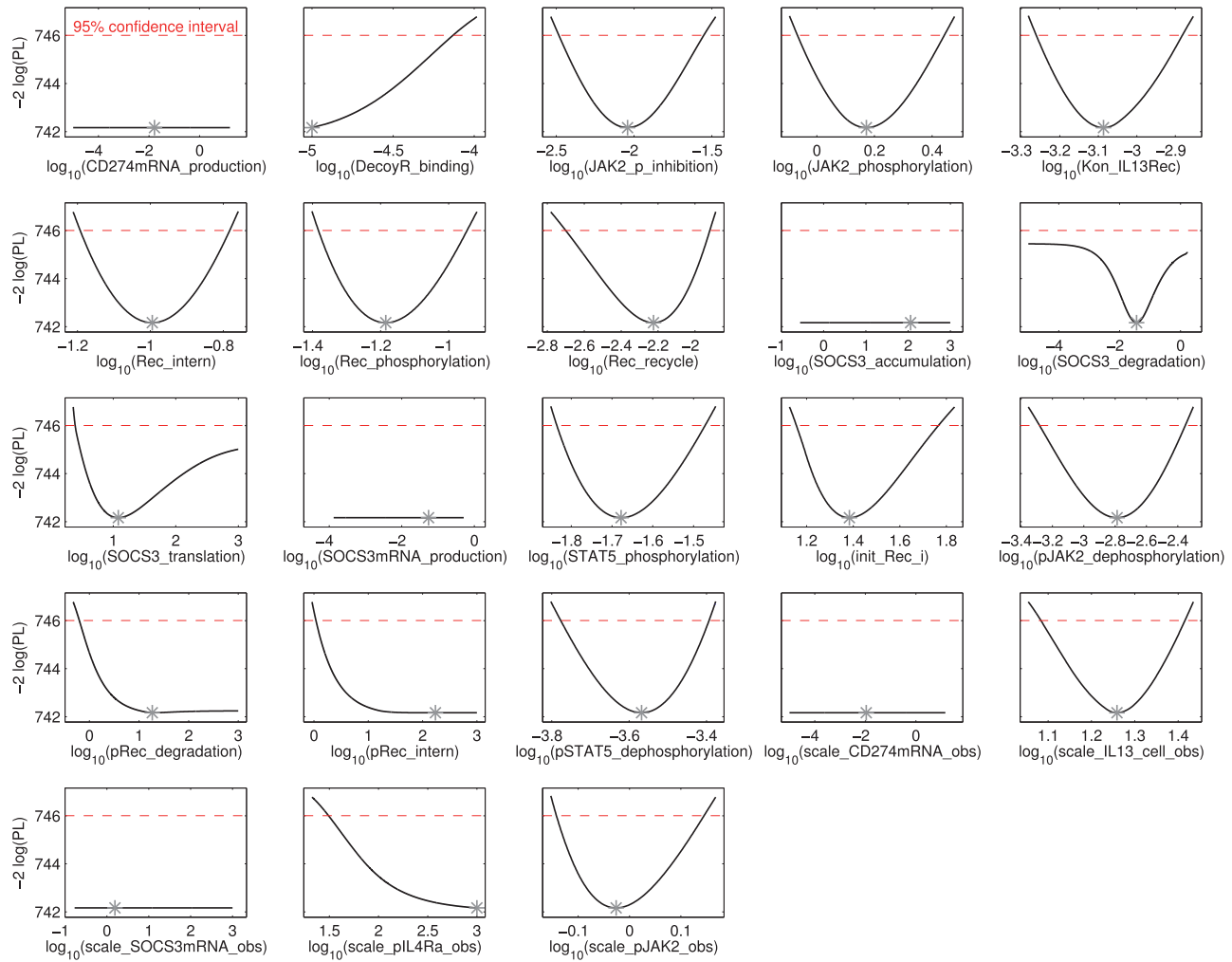
can also be used to design experiment that resolve practical non-identifiabilities [for an illustrative example, see Raue *et al.* (2010)]. The remaining parameters are both structurally and practically identifiable, and have finite confidence intervals; see Table 3 for values.

The PL approach is implemented in the freely available MATLAB software packages Data2Dynamics (Raue *et al.*, 2013) and PottersWheel (Maiwald and Timmer, 2008). For the Data2Dynamics software packages, the Raia *et al.* (2011) model and data are included in the software as an example application. The software package is open source and freely available on the Web site: <https://bitbucket.org/d2d-development/d2d-software>.

## 4 DISCUSSION

The results of all three approaches are in good agreement for the benchmark application considered here. Five of 23 parameters are consistently classified as structurally non-identifiable. The procedure to reproduce and interpret the results obtained by each of the three approaches was presented and can serve as reference for further application.

The strength of the DAISY approach is to check for global identifiability, i.e. it checks the uniqueness of the parameter solution. Thus, it is able to distinguish between global and local identifiability. Being based on differential algebra methods, DAISY can directly deal only with polynomial or rational functions  $\mathbf{f}$  and  $\mathbf{h}$ . However, the method can be generalized to deal with some non-polynomial functions, e.g. exponential functions. Although the program is usually very fast, in the order of few seconds, for complex models the algorithm may not successfully



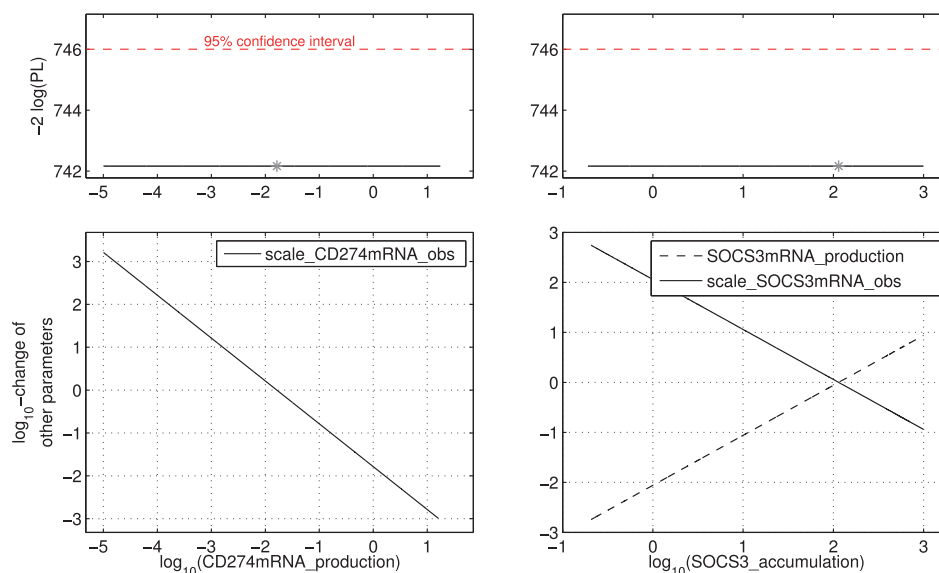
**Fig. 2.** Likelihood profiles for all 23 model parameters. The parameters are allowed to vary between  $-5$  and  $+3$  on a  $\log_{10}$ -scale. The MLE point is indicated by asterisks. The red dashed line corresponds to a threshold that indicates a 95% confidence level. The points of pass-over of profile and threshold determine likelihood-based confidence intervals. Perfectly flat profiles reveal five structural non-identifiable parameters: CD274mRNA\_production, SOCS3\_accumulation, SOCS3mRNA\_production, scale\_CD274mRNA\_obs and scale\_SOCS3mRNA\_obs. In addition, profiles that have a unique minimum but do not cross the confidence threshold reveal six practically non-identifiable parameters: DecoyR\_binding, SOCS3\_degradation, SOCS3\_translation, pRec\_degradation, pRec\_intern and scale\_pIL4Ra\_obs. The remaining parameters are both structurally and practically identifiable and have finite confidence intervals

terminate because of a lack of memory of the system. In most cases, however, it is possible to simplify the calculations required by the algorithm by eliminating redundant model equations by hand. This was done for the benchmark application considered here.

The main advantage of the EAR approach is that it is fast and can handle large and complex systems. The system analyzed in this article is considered small and simple for this approach. Systems on the scale of 100 states and 100 parameters can be handled. One limitation of the EAR approach is that it requires the vector valued functions  $\mathbf{f}$ ,  $\mathbf{g}$  and  $\mathbf{h}$  to be rational functions of their arguments. This limitation is not as restrictive as it may first sound, however, as it can be shown that any function, which in itself is the solution to an equation such as (1), can be handled through an extended state space approach (Lindskog 1996).

Even if-statements can be closely approximated using rational functions.

The strength of the PL approach is that it does not pose any restrictions on the algebraic form of the model equations. Even non-algebraic constructs such as if-statements or constraints on the parameters or model dynamics can be handled. It also allows for statements on practical identifiability and confidence intervals. Besides the structurally non-identifiable parameters, six parameters are practically non-identifiable considering the data available from Raia *et al.* (2011). However, owing to the underlying parameter estimation problem, issues such as local minima have to be handled with care. In such case, it might be necessary to repeat profile calculations for multiple minima detected in the objective function to enhance robustness of the results (Raue *et al.*, 2013).



**Fig. 3.** Likelihood profiles and corresponding  $\log_{10}$ -change of the related parameters along the profile for structurally non-identifiable parameters CD274mRNA\_production (left) and SOCS3\_accumulation (right). The lower panels reveal that the five structurally non-identifiable parameters are functionally related in two groups. In the first group,  $\text{CD274mRNA\_production} \sim 1/\text{scale\_CD274mRNA\_obs}$  indicates that the concentration scale of CD274 mRNA is not fixed by measurements. In the second group,  $\text{SOCS3\_accumulation} \sim 1/\text{scale\_SOCS3mRNA\_obs}$  and  $\text{SOCS3\_accumulation} \sim \text{SOCS3mRNA\_production}$  indicate that the concentration scale of SOCS3 mRNA is not fixed by measurements

## 5 CONCLUSION

The presented approaches allow for comparable and reliable conclusions about structural identifiability and can readily be used in Systems Biology applications. Software implementation of all approaches is freely available. It is important and good practice to double check results with different, but comparable, approaches. Here, we provide a case study that can serve as reference and hand-on guide to apply and interpret the results of three current approaches to structural identifiability analysis. The results of identifiability analysis can be helpful to provide guidelines on how to simplify the model structure or design additional experiments that enhance the predictive power of a mathematical model.

All three approaches examined in this paper are useful for real application examples. In many cases, all three approaches work equally well, but in some cases one of the three is preferred. If the system is very large and/or if the analysis must be fast, then EAR is the preferred approach. If it is of importance to get truly global identifiability, DAISY is the preferred approach. If practical identifiability is important, or if the equations include nonrational expressions like if-statements, then PL should be used. Using combinations of approaches can sometimes give the strengths of all approaches to the analysis.

## ACKNOWLEDGEMENTS

The authors thank the group of Ursula Klingmüller for providing the application and for devoting their experimental work to quantitative dynamical modeling.

**Funding:** German Federal Ministry of Education and Research [Virtual Liver (Grant No. 0315766), LungSys II (Grant

No. 0316042G)], and the Excellence Initiative of the German Federal and State Governments (EXC 294). This work was also supported by grants from the European Commission 7th Framework Programme [UNICELLSYS, (grant No 201142) and Cancersys, (grant No 223188)] and the Swedish Foundation for Strategic Research through the Gothenburg Mathematical Modelling Centre.

**Conflict of Interest:** none declared.

## REFERENCES

- Anguelova, M. *et al.* (2012) Minimal output sets for identifiability. *Math. Biosci.*, **239**, 139–153.
- Bachmann, J. *et al.* (2011) Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Mol. Syst. Biol.*, **7**, 516.
- Becker, V. *et al.* (2010) Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science*, **328**, 1404–1408.
- Bellman, R. and Åström, K. (1970) On structural identifiability. *Math. Biosci.*, **7**, 329–339.
- Bellu, G. *et al.* (2007) DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.*, **88**, 52–61.
- Karlsson, J. *et al.* (2012) An efficient method for structural identifiability analysis of large dynamic systems. In: *Proceedings of the 16th IFAC Symposium on System Identification*. 2012.
- Kreutz, C. *et al.* (2012) Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Syst. Biol.*, **6**, 120.
- Lindskog, P. (1996) *Methods, algorithms and tools for system identification based on prior knowledge*. PhD Thesis no 436, Linköping University, Sweden.
- Maiwald, T. and Timmer, J. (2008) Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics*, **24**, 2037–2043.
- Pohjanpelto, H. (1978) System identifiability based on the power series expansion of the solution. *Math. Biosci.*, **41**, 21–33.



- Raia, V. et al. (2011) Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets. *Cancer Res.*, **71**, 1–12.
- Raue, A. et al. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**, 1923–1929.
- Raue, A. et al. (2010) Identifiability and observability analysis for experimental design in non-linear dynamical models. *Chaos*, **20**, 045105.
- Raue, A. et al. (2013) Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One*, **8**, e74335.
- Saccomani, M. et al. (2003) Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, **39**, 619–632.
- Sedoglavic, A. (2002) A probabilistic algorithm to test local algebraic observability in polynomial time. *J. Symbolic Comput.*, **33**, 735–755.
- Walter, E. (1987) *Identifiability of Parametric Models*. Pergamon Press, Oxford, UK.