Vol. 30 no. 10 2014, pages 1440–1448 BIOINFORMATICS ORIGINAL PAPER doi:10.1093/bioinformatics/btu006

Systems biology Advance Access publication January 23, 2014 Comparison of approaches for parameter identifiability analysis of biological systems

Andreas Raue1,2,\*, Johan Karlsson3, Maria Pia Saccomani4, Mats Jirstrand3 and

D

o

w

Jens Timmer1,5,6

n

l

o

a

d

1University of Freiburg, Institute for Physics, 79104 Freiburg, Germany, 2Merrimack Pharmaceuticals Inc., 02139

e

d

Cambridge, MA, USA, 3Fraunhofer-Chalmers Research Centre for Industrial Mathematics, Chalmers Science Park,

f

r

o

SE-412 88 Go¨ teborg, Sweden, 4Department of Information Engineering, University of Padova, 35131 Padova, Italy,

m

h

5BIOSS Centre for Biological Signalling Studies and 6Zentrum fu¨ r Biosystemanalyse (ZBSA), University of Freiburg, 79104

t

t

p

s

Freiburg, Germany

:

/

/

a

c

Associate Editor: Alfonso Valencia

a

d

e

m

i

c

experimental treatments; in this case it represents a constant

.

ABSTRACT

o

u

treatment with the cytokine IL13. The vector h contains all par

p

Motivation: Modeling of dynamical systems using ordinary differential

.

c

ameters of the dynamical model, such as the reaction rate con

o

equations is a popular approach in the field of Systems Biology. The

m

stants. For a specific cell type or biological context, the

/

b

amount of experimental data that are used to build and calibrate these

i

o

parameters h are often not available from literature and have

i

models is often limited. In this setting, the model parameters may not

n

f

o

be uniquely determinable. Structural or a priori identifiability is a prop erty 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 inde pendently. 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 dif ferential equations (ODEs). We used a model of IL13-induced JAK/STAT signaling (Raia et al., 2011) to present a comprehen sive comparison of three approaches for structural identifiability analysis. Mathematically, such dynamical models can be charac terized by the ODE

x\_ðtÞ ¼ fðxðtÞ, uðtÞ, hÞ with xð0Þ ¼ x0 ¼ gðhÞ ð1Þ In this case, the vector of state variables xðtÞ describes

to be estimated from experimental data. The initial concentra

r

m

tions x0 can be known or unknown; in the latter case they have

a

t

i

c

to be estimated from experimental data as well. Each possible

s

/

a

measurement is mathematically represented by a functional

r

t

i

c

mapping

l

e

/

3

0

yðtÞ ¼ hðxðtÞ, uðtÞ, hÞ ð2Þ

/

1

0

/

1

that might include additional parameters and thus increase the

4

4

dimension of h, such as scaling parameter for relative data

0

/

2

obtained by immunoblotting in our case.

6

6

1

The inference problem of concern is the determination of the

8

8

model parameters h from the measurements y and inputs u

b

y

described by Equation (2). In general, this may not be possible.

S

a

Structural or a priori identifiability (Bellman and A˚stro¨m, 1970) is

n

D

a property of the systems (1) and (2) that guarantees that the

i

e

g

unknown model parameters can, in principle, be determined

o

S

from generic input and output functions of the model,

t

a

t

e

provided they satisfy certain minimal conditions called ‘persist

U

n

ence of excitation’ [for a broad introduction and classical

i

v

e

approaches Walter (1987)]. Structural identifiability is a

r

s

i

t

prerequisite and necessary condition for any estimation procedure

y

L

to render the recovery of h from input–output measurements as a

i

b

r

a

well-posed problem and to return meaningful results about h.

r

y

u

Well-known approaches to structural identifiability analysis

s

e

often have problems with the realistically sized models

r

o

n

(Karlsson et al., 2012; Sedoglavic, 2002), but there are approaches

2

8

that can handle them.

J

u

Let us consider a simple model for the illustration of structural

n

e

identifiability. The model describes a transition between two

2

0

2

components

2

the dynamics of 15 molecular components involved in the

JAK/STAT signaling pathway (for details see equation (3)).

The function uðtÞ represents possibly time-dependent

x\_1ðtÞ¼ 1x1 x\_2ðtÞ¼þ 1x1

We assume that x1ð0Þ ¼ 2 and x2ð0Þ ¼ 0, and only the measure-

\*To whom correspondence should be addressed. ment y1ðtÞ ¼ 3 x2ðtÞ, are available. In this case, we can solve for 1440 The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Identifiability analysis

Table 1. Classification of the investigated approaches

Approach Identifiability Scope Type Runtime Main reference

DAISY Structural Global A priori 120 min Saccomani et al. (2003) EAR Structural Local (around a generic point) A priori 30 s Karlsson et al. (2012) PL Structural and practical Local (region covered by profile) Data-based 30 min Raue et al. (2009)

D

o

w

n

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

l

o

a

d

e

d

f

r

x2ðtÞ analytically and substitute into the measurement equation. The result

y1ðtÞ ¼ 3  2 expð 1 tÞ ðexpð 1 tÞ 1Þ

shows that 3 and 2 are structurally non-identifiable. 1 and only the product of 3 and 2 are structurally identifiable. Such redun dant 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 sev eral 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 identifiabil ity, 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 depend ing also on derivatives of the variable. Ritt’s algorithm permits to elim inate 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

o

m

linear equations in the unknown . By applying a computer algebra al

h

t

gorithm, i.e. the Buchberger algorithm, it is possible to check whether

t

p

s

there is one or multiple solutions and hence distinguish between global, or

:

/

/

a

local identifiability or non-identifiability of the original dynamic system.

c

a

An additional advantage of using this computer algebra tool is that

d

e

m

it does not require expertise on mathematical modeling by the

i

c

experimenter.

.

o

u

p

.

c

2.2 EAR approach

o

m

/

b

This approach is based on applying the inverse function theorem to the

i

o

i

system of algebraic equations relating higher order derivatives of the

n

f

o

output y with respect to time at the initial time with the initial state

r

m

and parameters (Pohjanpalo (1978). Using a differential algebra ap

a

t

i

proach, an upper bound of the order of differentiation can be given,

c

s

/

resulting in a non-linear algebraic system of equations in the parameters.

a

r

t

The rank of the Jacobian matrix for this system of equations gives infor

i

c

l

e

mation about its solvability, and in case of a rank-deficient matrix, a

/

3

0

more detailed analysis of the Jacobian provides information about

/

1

0

which parameters are involved in relations rendering the system non

/

1

identifiable. The EAR method provides means to efficiently compute

4

4

0

the generic rank of the Jacobian matrix to return a conclusive result if

/

2

the system is structurally identifiable. It considers local identifiability but

6

6

around a generic point, i.e. the computations are carried out for a

1

8

8

random specialization of the unknown parameters, and initial state to

b

integer values and a random input in terms of a truncated integer coef

y

S

ficient power series. Local structural identifiability is an almost every

a

n

where property by definition, i.e. it holds everywhere apart from

D

i

possibly on a set of measure zero. The EAR approach is based on a

e

g

o

method for local algebraic observability (Sedoglavic, 2002). The EAR

S

analysis is implemented as a fully automatic Mathematica function.

t

a

t

The user simply inputs the equations and gets the answer. Based on the

e

U

EAR identifiability analysis, it is also possible to find minimal sets of

n

i

v

outputs giving identifiability (Anguelova et al., 2012).

e

r

s

i

t

y

2.3 PL approach

L

i

b

r

a

This approach checks for non-identifiability by posing a parameter esti

r

y

mation problem using real or simulated data. The central idea is that non

u

s

identifiability manifests as a flat manifold in the parameter space of the

e

r

estimation problem, e.g. the likelihood function. A profile can be calcu

o

n

lated for each parameter i individually by repeated optimization of all

2

8

parameters f jj8j 6¼ ig for a series of fixed values of the parameter i. A

J

u

flat profile indicates a structurally non-identifiable parameter. For detect

n

e

ing structural non-identifiability, simulated data are sufficient. In case

2

0

real experimental data are available, practical non-identifiability can

2

2

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)

1441

A.Raue et al.

extended the approach to detect non-observability of the dynamics dir ectly. 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

8>>>>>>>>>>>>>>>>>>>>>>>>>><

x\_1ðtÞ ¼ 1c1 u1 x1  5x1 þ 6x2

x\_2ðtÞ ¼þ 5x1  6x2

x\_3ðtÞ ¼þ 1c1 u1 x1  2x3x7

x\_4ðtÞ ¼þ 2x3x7  3x4

x\_5ðtÞ ¼þ 3x4  4x5

x\_6ðtÞ ¼ 7x3x6=ð1 þ 13x14Þ

7x4x6=ð1 þ 13x14Þ þ c2  8x7

x\_7ðtÞ ¼þ 7x3x6=ð1 þ 13x14Þþ

>>>>>>>>>>>>>>>>>>>>>>>>>>:ð3Þ

þ 7x4x6=ð1 þ 13x14Þ c2  8x7

x\_8ðtÞ ¼ 9x8x7 þ c2  10x9

x\_9ðtÞ ¼þ 9x8x7  c2  10x9

x\_10ðtÞ¼þ 11x9

x\_11ðtÞ¼ 12c1 u1 x11

x\_12ðtÞ¼þ 12c1 u1 x11

x\_13ðtÞ¼þ 14x10=ð 15 þ x10Þ 16x13

x\_14ðtÞ¼þ 17x9

where x is the 14D state vector, the dot denotes the time derivative, u1 is an input functions and c1, 2 are constants. The initial conditions of the state variables are

xð0Þ¼½1:3, 23, 0, 0, 0, 2:8, 0, 165, 0, 0, 0:34, 0, 0, 0 ð4Þ

the values are given in units of molecules per cell ( 1000). The set of measurement equations is defined by the following equation 8>>>>>>>>>><

y1ðtÞ ¼ x1 þ x3 þ x4

y2ðtÞ ¼ 18ðx3 þ x4 þ x5 þ x12Þ

y3ðtÞ ¼ 19ðx4 þ x5Þ

y4ðtÞ ¼ 20x7

>>>>>>>>>>:ð5Þ

y5ðtÞ ¼ 21x10

y6ðtÞ ¼ 22x14

y7ðtÞ ¼ x13

y8ðtÞ ¼ x9:

The components x, u and c of Equation (3) and y of Equation (5) are described in Table 2. The model Equations (3–5) contain 23 unknown parameters h 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: 11, 15, 17, 21, 22.

1442

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

D

model is more suitable for further mathematical investigations. o

w

Thus, before checking model identifiability from the designed

n

l

o

experiment, it is convenient to check its minimality. In this

a

d

case, just by visual inspection, it is easy to see that some of the

e

d

14 model equations defining the model are redundant.

f

r

o

m

In particular, the equations for x6, x8, x11 are the same as those

h

for x7, x9, x12 with the opposite sign. For example, from the

t

t

p

s

eighth and ninth equations, one obtains x\_9 ¼ x\_8. By integrat

:

/

/

a

ing and using the known initial conditions, one arrives at

c

a

x9 ¼ x8 þ 165, thus eliminating one differential equation.

d

e

m

Same procedure is followed for x7, x12. Finally, x14 can be ex

i

c

.

pressed as x14 ¼ x10 17= 11, and the last equation is also redun

o

u

p

dant. The model can then be rewritten in a simplified form

.

c

o

involving only 10 state variables. This is done not only for the m

/

sake of mathematical simplification, but also to satisfy some

b

i

o

important structural property, such as minimality and accessibil

i

n

f

o

ity (Saccomani et al., 2003). This is important, because the

r

m

lack of minimality of the model may lead to spurious non

a

t

i

c

identifiability results for some parameters that may not occur

s

/

a

in a minimal model. Also, if the model has two or more differ

r

t

i

c

ential equations dependent on each other, as in this case, the

l

e

/

3

model would not be accessible, making more difficult the iden

0

/

1

tifiability check of the model from the given initial conditions.

0

/

1

One thus arrives at the following simplified 10D model:

4

4

8>>>>>>>>>>>>>>>><

0

x\_1ðtÞ ¼ 1x1c1 u1  5x1 þ 6x2

/

2

6

x\_2ðtÞ ¼ 5x1  6x2

6

1

8

x\_3ðtÞ ¼ 1c1 u1x1  2x3ð x6 þ 2:8Þ

8

b

x\_4ðtÞ ¼ 2x3ð x6 þ 2:8Þ 3x4

y

S

x\_5ðtÞ ¼ 3x4  4x5

a

n

x\_6ðtÞ ¼ 7x3x6=ð1 þ 13x13Þ

>>>>>>>>>>>>>>>>:ð6Þ

D

i

7x4x6=ð1 þ 13x13Þ þ 8ð x6 þ 2:8Þc2

e

g

o

x\_8ðtÞ ¼ 9x8ð x6 þ 2:8Þ þ 10ð x8 þ 165Þc2

S

x\_10ðtÞ ¼ 11ð x8 þ 165Þ

t

a

t

e

x\_11ðtÞ¼ 12c1 u1x11

U

x\_13ðtÞ ¼ x10 14=ð 15 þ x10Þ 16x13

n

i

v

e

r

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

s

i

t

y

To check global identifiability of this simplified 10D model

L

i

b

with DAISY, the user has to write in the input file the ordered

r

a

r

list of the output and state variables, the list of the unknown

y

u

parameters, the model equations and the known initial condi

s

e

r

tions. Later in the text, the results will be illustrated. For explan

o

n

ations of the technical terms, one may consult Bellu et al. (2007).

2

8

DAISY automatically ranks the input, output, state variables

J

u

and their derivatives, starts the pseudodivision algorithm, i.e. the

n

e

Ritt algorithm, and calculates the characteristic set of the model.

2

0

2

This is a minimal set of differential polynomials, which provides 2

an equivalent description of the model. The subset made of the first eight (i.e. the number of model outputs) differential polyno mials does not depend on the state variable 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

x1 Rec Receptor for IL-13

x2 Rec\_i Internalized receptor for IL-13

x3 IL13\_Rec IL-13 receptor complex

x4 p\_IL13\_Rec Phosphorylated IL-13 receptor complex

x5 p\_IL13\_Rec\_i Internalized phosphorylated IL-13 receptor complex x6 JAK2 Janus kinase 2

x7 pJAK2 Phosphorylated Janus kinase 2

x8 STAT5 Signal Transducer and Activate of Transcription 5

x9 pSTAT5 phosphorylated Signal Transducer and Activate of Transcription 5 x10 SOCS3Mrna Suppressor of cytokine signaling 3 mRNA

x11 DecoyR Decoy receptor for IL-13

x12 IL13\_DecoyR Il-13 decoy receptor complex

x13 SOCS3 Suppressor of cytokine signaling 3 protein

x14 CD274Mrna Cluster of Differentiation 274 mRNA

c1 Conversion factor from ng/ml to molecules per cell ( 1000) with value 2.265

Identifiability analysis

D

o

w

n

l

o

a

d

e

d

f

r

o

m

h

t

t

p

s

:

/

/

a

c

a

d

e

m

c2 SHP1 Constant intracellular concentration of protein tyrosine phosphatase SHP-1 with value 91 molecules

i

c

per cell ( 1000)

.

o

u

p

.

u1 IL13 Constant extracellular concentration of Interleukine-13 0, 4, 20, 80 ng/ml

c

o

m

/

y1 RecSurf\_obs IL-13 receptor at cell membrane

b

i

o

y2 IL13\_cell\_obs IL-13 at cell membrane or intracellular

i

n

f

y3 pIL4Ra\_obs Phosphorylated IL-13 receptor at membrane or intracellular

o

r

m

y4 pJAK2\_obs Phosphorylated Janus kinase 2

a

t

y5 SOCS3mRNA\_obs Suppressor of cytokine signaling 3 mRNA

i

c

s

y6 CD274mRNA\_obs Cluster of Differentiation 274 mRNA

/

a

r

y7 SOCS3\_obs Suppressor of cytokine signaling 3 protein

t

i

c

l

e

y8 pSTAT5\_obs phosphorylated Signal Transducer and Activate of Transcription 5

/

3

0

/

1

0

/

1

4

4

0

/

2

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

6

6

1

8

8

b

Biological short name Identifiability MLE point Likelihood-based confidence interval

y

S

a

n

1 Kon\_IL13Rec Identifiable –3.087 ½ 3:261, 2:882

D

i

2 Rec\_phosphorylation Identifiable –1.185 ½ 1:383, 0:946

e

g

3 pRec\_intern Practically non-identifiable þ2.236 ½þ0:037, þ 1

o

S

4 pRec\_degradation Practically non-identifiable þ1.268 ½ 0:203, þ 1

t

a

t

5 Rec\_intern Identifiable –0.995 ½ 1:193, 0:783

e

U

6 Rec\_recycle Identifiable –2.225 ½ 2:700, 1:920

n

i

7 JAK2\_phosphorylation Identifiable þ0.172 ½ 0:067, þ 0:442

v

e

r

8 pJAK2\_dephosphorylation Identifiable –2.788 ½ 3:283, 2:352

s

i

t

y

9 STAT5\_phosphorylation Identifiable –1.678 ½ 1:835, 1:472

L

10 pSTAT5\_dephosphorylation Identifiable –3.568 ½ 3:777, 3:393

i

b

r

a

11 SOCS3mRNA\_production Structurally non-identifiable –1.260 ½ 1, þ 1

r

y

12 DecoyR\_binding Practically non-identifiable –5.000 ½ 1, 4:131

u

s

13 JAK2\_p\_inhibition Identifiable –2.042 ½ 2:483, 1:552

e

r

15 SOCS3\_accumulation Structurally non-identifiable þ2.059 ½ 1, þ 1

o

n

14 SOCS3\_translation Practically non-identifiable þ1.081 ½þ0:392, þ 1

2

8

16 SOCS3\_degradation Practically non-identifiable –1.442 ½ 1, þ 1

J

u

17 CD274mRNA\_production Structurally non-identifiable –1.786 ½ 1, þ 1

n

e

18 scale\_IL13\_cell\_obs Identifiable þ1.259 ½þ1:083, þ 1:416

2

0

19 scale\_pIL4Ra\_obs Practically non-identifiable þ3.000 ½þ1:491, þ 1

2

2

20 scale\_pJAK2\_obs Identifiable –0.026 ½ 0:143, þ 0:145 21 scale\_SOCS3mRNA\_obs Structurally non-identifiable þ0.194 ½ 1, þ 1 22 scale\_CD274mRNA\_obs Structurally non-identifiable –1.965 ½ 1, þ 1 23 init\_Rec\_i Identifiable þ1.384 [þ1.156, þ1.766]

Note: Parameter values are given on a log10-scale and are allowed to vary between 5 and þ3.

1443

A.Raue et al.

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 h, 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 h. We should, in prin ciple, calculate the range set of these functions. This could be done in symbolic language evaluating them at a symbolic par ameter value. In practice, but only to speed up the process, in stead 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 Gro¨bner

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 par D

ameters and initial conditions are generated.

o

w

n

(2) Specialize inputs to truncated random integer coefficient

l

o

a

power series.

d

e

d

(3) Truncated power series solutions of x, @x

@xð0Þ and @x@ are com puted, using the system and sensitivity system equations.

f

r

o

m

h

(4) The power series are inserted in the expressions for the

t

t

p

derivatives of the outputs with respect to the initial state

s

:

/

/

and with respect to the parameters

a

c

a

d

dxð0Þy ¼ d

d

basis, by applying the Buchberger algorithm. The results show

that all the parameters are uniquely identifiable except for

11, 15, 17, 21 and 22, which have an infinite number of solu

dxð0Þh ¼ @h@x@x

@xð0Þ ð7Þ

e

m

ic

.

o

u

p

tions. 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 as signing known values to parameters 21 and 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/?IDsezione¼4364.

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 ap proach 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, fol lowed 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 de nominators, 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

1444

d y ¼ dd h ¼ @h@x@x@ þ@h@ ð8Þ

d

.

c

o

m

/

b

i

o

resulting in power series representations of the output sen

i

n

f

sitivity derivatives.

o

r

m

(5) Identification of the coefficients of the truncated power

a

t

i

c

series of the output derivatives with the coefficients of a

s

/

a

general Taylor expansion of the output sensitivity deriva

r

t

i

c

tives gives the higher order time-derivatives of the output

l

e

/

3

sensitivity derivatives, i.e. the entries of the specialized

0

/

1

Jacobian matrix.

0

/

1

(6) Calculate the rank.

4

4

0

/

(7) If the matrix is rank-deficient, the non-identifiable param

2

6

6

eters are found using the fact that removing the corres

1

8

ponding columns from the matrix do not change the rank.

8

b

y

The analysis using the Mathematica package is fully auto

S

a

matic. Further information about the package and instructions

n

D

on how to obtain it can be found at: http://www.fcc.chalmers.se/

i

e

g

sys/products/identifiabilityanalysis.

o

S

Included in the package is, apart from the identifiability test

t

a

t

demonstrated above, also functionality for automatically finding

e

U

minimal sets of output expressions that guarantee a structurally

n

i

v

identifiable model. This functionality is described in Anguelova

e

r

s

et al. (2012).

i

t

y

L

i

b

r

a

3.3 PL approach

r

y

u

The PL approach determines the identifiability of the model par

s

e

r

ameters by posing a parameter estimation problem. Here, we use the

o

n

original data of Raia et al.(2011) and investigate both the structural

2

8

and practical identifiability of the model parameters. For parameter

J

u

estimation, maximum likelihood estimation (MLE) is applied. The

n

e

likelihood LðhÞ describes the probability of the data given certain

2

0

2

parameter values h. The MLE fit of the dynamic model to the ex 2

perimental 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

PLð iÞ ¼ max

j6¼i½Lð jÞ ð9Þ

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

Identifiability analysis

D

o

w

n

l

o

a

d

e

d

f

r

o

m

h

t

t

p

s

:

/

/

a

c

a

d

e

m

i

c

.

o

u

p

.

c

o

m

/

b

i

o

i

n

f

o

r

m

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 a

with four different doses of IL-13. The data were obtained by Immunoblot and qRT-PCR measurements. Some measurements are on a relative

t

i

c

s

concentration scale given in arbitrary units [au] and some on an absolute concentration scale [abs] that corresponds to molecules per cell ( 1000)

/

a

r

t

i

c

l

e

/

3

0

/

where for each fixed value of parameter i, all other parameters j 6¼ 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 j along a profile of a structurally non-identifiable parameter 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,

17 1 22

indicates that the concentration scale of CD274 mRNA (x14) is not fixed by measurements. In the second group,

15 1 21and 15  11

indicates that the concentration scale of SOCS3 mRNA (x10) 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: 3, 4, 12, 14, 16 and 19. The likelihood profiles

can also be used to design experiment that resolve practical non

1

0

/

identifiabilities [for an illustrative example, see Raue et al.

1

4

4

(2010)]. The remaining parameters are both structurally and

0

/

2

practically identifiable, and have finite confidence intervals; see

6

6

Table 3 for values.

1

8

8

The PL approach is implemented in the freely available

b

y

MATLAB software packages Data2Dynamics (Raue et al.,

S

2013) and PottersWheel (Maiwald and Timmer, 2008). For the

a

n

Data2Dynamics software packages, the Raia et al. (2011) model

D

i

e

and data are included in the software as an example application.

g

o

The software package is open source and freely available on the

S

t

a

Web site: https://bitbucket.org/d2d-development/d2d-software.

t

e

U

n

i

v

e

r

4 DISCUSSION

s

i

t

y

The results of all three approaches are in good agreement for the

L

i

b

benchmark application considered here. Five of 23 parameters

r

a

r

are consistently classified as structurally non-identifiable. The

y

u

s

procedure to reproduce and interpret the results obtained by

e

r

each of the three approaches was presented and can serve as

o

n

reference for further application.

2

8

The strength of the DAISY approach is to check for global

J

u

n

identifiability, i.e. it checks the uniqueness of the parameter so

e

2

lution. Thus, it is able to distinguish between global and local

0

2

identifiability. Being based on differential algebra methods, 2

DAISY can directly deal only with polynomial or rational func tions f and 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

1445

A.Raue et al.

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

|  |
| --- |

D

o

w

n

lo

a

de

d

fr

o

m

h

ttp

s

:

//a

c

a

de

m

ic

.

o

u

p

.

c

o

m

/b

io

in

fo

r

m

a

t

ic

s

/a

r

t

ic

le

/3

0

/1

0

/1

440

/2

66188

by

Fig. 2. Likelihood profiles for all 23 model parameters. The parameters are allowed to vary between 5 and þ3 on a log10-scale. The MLE point is

S

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

a

n

threshold determine likelihood-based confidence intervals. Perfectly flat profiles reveal five structural non-identifiable parameters:

D

i

e

CD274mRNA\_production, SOCS3\_accumulation, SOCS3mRNA\_production, scale\_CD274mRNA\_obs and scale\_SOCS3mRNA\_obs. In addition, g

o

profiles that have a unique minimum but do not cross the confidence threshold reveal six practically non-identifiable parameters: DecoyR\_binding,

S

SOCS3\_degradation, SOCS3\_translation, pRec\_degradation, pRec\_intern and scale\_pIL4Ra\_obs. The remaining parameters are both structurally and

t

a

t

e

practically identifiable and have finite confidence intervals

U

n

i

v

e

r

s

i

t

y

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 f, g and 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).

1446

Even if-statements can be closely approximated using rational

L

i

b

functions.

r

a

r

The strength of the PL approach is that it does not pose any

y

u

restrictions on the algebraic form of the model equations. Even

s

e

r

non-algebraic constructs such as if-statements or constraints on

o

n

the parameters or model dynamics can be handled. It also allows

2

8

for statements on practical identifiability and confidence inter

J

u

vals. Besides the structurally non-identifiable parameters, six par

n

e

ameters are practically non-identifiable considering the data

2

0

2

available from Raia et al. (2011). However, owing to the under 2

lying 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).

|  |
| --- |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Identifiability analysis

|  |
| --- |

D

o

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

w

n

l

o

a

d

e

d

f

r

o

m

h

t

t

p

s

:

/

/

a

c

a

d

e

m

i

c

.

o

u

p

.

c

Fig. 3. Likelihood profiles and corresponding log10-change of the related parameters along the profile for structurally non-identifiable

o

m

parameters CD274mRNA\_production (left) and SOCS3\_accumulation (right). The lower panels reveal that the five structurally non-identifiable

/

b

i

parameters are functionally related in two groups. In the first group, CD274mRNA production 1=scale CD274mRNA obs indicates that the con

o

i

n

centration scale of CD274 mRNA is not fixed by measurements. In the second group, SOCS3 accumulation 1=scale SOCS3mRNA obs and

f

o

r

SOCS3 accumulation SOCS3mRNA production indicate that the concentration scale of SOCS3 mRNA is not fixed by measurements

m

a

t

i

c

s

/

a

r

t

i

c

l

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 nonra tional 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 Klingmu¨ller for provid ing 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

e

/

3

Federal and State Governments (EXC 294). This work was

0

/

1

also supported by grants from the European Commission 7th

0

/

1

Framework Programme [UNICELLSYS, (grant No 201142)

4

4

0

and CANCERSYS, (grant No 223188)] and the Swedish

/

2

6

Foundation for Strategic Research through the Gothenburg

6

1

Mathematical Modelling Centre.

8

8

b

y

Conflict of Interest: none declared.

S

a

n

D

i

e

g

o

REFERENCES

S

t

a

t

Anguelova,M. et al. (2012) Minimal output sets for identifability. Math. Biosci.,

e

U

239, 139–153.

n

Bachmann,J. et al. (2011) Division of labor by dual feedback regulators

i

v

e

controls JAK2/STAT5 signaling over broad ligand range. Mol. Syst. Biol., 7,

r

s

i

516.

t

y

Becker,V. et al. (2010) Covering a broad dynamic range: information processing at

L

i

b

the erythropoietin receptor. Science, 328, 1404–1408.

r

a

Bellman,R. and A˚ stro¨m,K. (1970) On structural identifiability. Math. Biosci., 7,

r

y

u

329–339.

s

e

Bellu,G. et al. (2007) DAISY: a new software tool to test global identifiability of

r

o

biological and physiological systems. Comput. Methods Programs Biomed., 88, n

52–61.

2

8

Karlsson,J. et al. (2012) An efficient method for structural identifiability analysis of

J

u

large dynamic systems. In: Proceedings of the 16th IFAC Symposium on System

n

e

Identification. 2012.

2

0

Kreutz,C. et al. (2012) Likelihood based observability analysis and confidence inter 2

2

vals 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, Linko¨ping University, Sweden.

Maiwald,T. and Timmer,J. (2008) Dynamical modeling and multi-experiment fit ting with PottersWheel. Bioinformatics, 24, 2037–2043.

Pohjanpalo,H. (1978) System identifiability based on the power series expansion of the solution. Math. Biosci., 41, 21–33.

1447

A.Raue et al.

Raia,V. et al. (2011) Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of thera peutic 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.

D

o

w

n

l

o

a

d

e

d

f

r

o

m

h

t

t

p

s

:

/

/

a

c

a

d

e

m

i

c

.

o

u

p

.

c

o

m

/

b

i

o

i

n

f

o

r

m

a

t

i

c

s

/

a

r

t

i

c

l

e

/

3

0

/

1

0

/

1

4

4

0

/

2

6

6

1

8

8

b

y

S

a

n

D

i

e

g

o

S

t

a

t

e

U

n

i

v

e

r

s

i

t

y

L

i

b

r

a

r

y

u

s

e

r

o

n

2

8

J

u

n

e

2

0

2

2

1448