Bioinformatics, 2019, 1–2
doi: 10.1093/bioinformatics/bty1069
Advance Access Publication Date: 2 January 2019
Applications Note

Systems biology

SIAN: software for structural identifiability analysis of ODE models

Hoon Hong[1,†], Alexey Ovchinnikov[2,†], Gleb Pogudin[3,*,†] and Chee Yap[3,†]

[1]Department of Mathematics, North Carolina State University, Raleigh, NC, USA, [2]Department of Mathematics, CUNY Queens College and Ph.D. Programs in Mathematics and Computer Science, CUNY Graduate Center, New York, NY, USA and [3]Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

*To whom correspondence should be addressed.
[†]The authors are ordered alphabetically and share the same credit.

Associate Editor: Jonathan Wren

Received on October 30, 2018; revised on December 17, 2018; editorial decision on December 21, 2018; accepted on December 28, 2018

Abstract

Summary: Biological processes are often modeled by ordinary differential equations with unknown parameters. The unknown parameters are usually estimated from experimental data. In some cases, due to the structure of the model, this estimation problem does not have a unique solution even in the case of continuous noise-free data. It is therefore desirable to check the uniqueness a priori before carrying out actual experiments. We present a new software SIAN (Structural Identifiability ANalyser) that does this. Our software can tackle problems that could not be tackled by previously developed packages.
Availability and implementation: SIAN is open-source software written in Maple and is available at https://github.com/pogudingleb/SIAN.
Contact: pogudin.gleb@gmail.com
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Ordinary differential equations (ODEs) with unknown parameters are widely used for modeling biological processes and phenomena. One is often interested in the values of these parameters due to their importance, as, e.g. they may represent key biological mechanisms or targets for intervention. A standard way to find the values of the parameters from experimental data is to find the parameter values that fit the data with minimal error, typically framed from a statistical perspective as maximum likelihood or Bayesian inference.

However, it might happen that, due to the structure of the model, it is impossible to recover the value of a parameter of interest from the data even assuming the ideal case of continuous noise-free data. If this is the case, then regardless of the chosen data fitting approach, it is impossible to guarantee that it will find the correct parameter value. As we will see, this structural property can be assessed *a priori* without conducting (often costly) experiments. Thus, a crucial first step to any parameter estimation problem is to check whether the parameter of interest is *structurally globally identifiable*, i.e. the parameter value can be recovered uniquely from the data under the assumption that the data is continuous and noise-free. We explain the notion of global identifiability in more detail in Section 3. For a formal definition and illustrating examples, we refer to Hong *et al.* (2018, Section 2).

We present SIAN (Structural Identifiability ANalyser), our new software for assessing identifiability for ODE models, based on the algorithm developed and rigorously justified in Hong *et al.* (2018).

2 Existing software for structural identifiability

Assessing global identifiability is a challenging problem. Hence a weaker notion called 'local identifiability' was introduced and tackled first. 'Local' indicates that a parameter can be identified locally (in some neighborhood). For a polynomial system, it is the same as saying that a parameter can be identified up to finitely many options. Modern software packages for assessing local identifiability include ObservabilityTest (Sedoglavic, 2002), EAR (Karlsson *et al.*, 2012) and STRIKE_GOLDD (Villaverde *et al.*, 2016).

However, even relatively simple real-life systems can involve locally but not globally identifiable parameters [see Thomaseth and Saccomani

(2018, Section 4), Norton (1982) and Supplementary Materials A.1]. Thus, it is highly desirable to have software that could assess global identifiability. There has been significant progress in this direction:

- packages DAISY (Bellu *et al.*, 2007) and COMBOS (Meshkat *et al.*, 2014) are based on the approach via input-output equations and can check global identifiability for systems with the 'solvability' property [see Hong *et al.* (2018, Example 6) for a discussion].
- GenSSI 2.0 package (Ligon *et al.*, 2018) is based on the generating series approach and checks global identifiability conditionally on extra input, the truncation order [for a discussion how the truncation order affects the output of the algorithm, see Hong *et al.* (2018, Example 7)].

3 Features

We present SIAN, software written in Maple, that has the following input-output specification.

Input. A system Σ of the form

$$\begin{cases} \dot{x}(t) = f(x(t), \mu, u(t)), \\ y(t) = g(x(t), \mu, u(t)), \\ x(0) = x^*, \end{cases} \quad (1)$$

- x is a vector of state variables,
- u is a vector of input (control) variables to be chosen by an experimenter,
- y is a vector of output variables,
- μ and x^* are vectors of unknown scalar parameters and unknown initial conditions, respectively,
- f and g are vectors of rational functions in x, μ and u with complex coefficients (other types of functions can also be handled, see Supplementary Material A.2)

and **a real number** $0 < P < 1$, the user-specified probability of correctness of the result. That is, SIAN is a Monte Carlo randomized algorithm, see Motwani and Raghavan (1995, Chapter 1.2).

Output. For every $\theta \in \mu \cup x^*$, SIAN assigns one of the following labels:

- **Globally identifiable**: for almost every solution of (1), every solution of (1) with the same u-component and y-component has the same value of θ.
- **Locally but not globally identifiable**: for almost every solution of (1), among the solutions of (1) with the same u-component and y-component, there are only finitely many possible values of θ.
- **Not identifiable**: for almost every solution of (1), among the solutions of (1) with the same u-component and y-component, there are infinitely many possible values of θ.

The assigned labels are correct with probability at least P.
We would like to emphasize the following **extra features**:

- SIAN is parallellizable and can take advantage of a multicore computing environment.
- SIAN assesses not only the identifiability of the model, but checks individual identifiability of every parameter.
- SIAN can assess identifiability of the parameters appearing in the system and the initial conditions of the state variables.

4 Performance and applications

In this section, we compare our software with the existing software tools for assessing global identifiability, namely COMBOS, DAISY and GenSSI (see Section 2). All of the benchmark problems are listed in Supplementary Material B and are available at https://github.com/pogudingleb/SIAN/tree/master/examples. The source code of the benchmark problems for COMBOS, DAISY and GenSSI used for the comparison is included into the Supplementary Data.

We use a computer with 96 CPUs, 2.4 GHz and CentOS 6.9 (Linux). The runtimes in Table 1 are the elapsed time. SIAN was run on Maple 2017 with the probability of correctness $P = 0.99$, GenSSI 2.0 was run on Matlab R2017a, and we used DAISY 1.9.

Table 1. Runtimes (in minutes) on benchmark problems

Example	GenSSI 2.0	COMBOS	DAISY	SIAN
Chemical Reaction	*	**	>6000	<1
HIV	>12 000	**	>6600	<1
SIRS w/forcing	>12 000	**	>6600	<1
Cholera	*	85	30	3
Protein complex	>12 000	**	>6600	47
Pharmacokinetics	>12 000	**	>7800	962

*GenSSI 2.0 returns '*Warning: Unable to find explicit solution*'.

**COMBOS returns '*Model may have been entered incorrectly or cannot be solved with COMBOS algorithms*'.

Acknowledgements

The authors are grateful to the CCiS at Queens College and CIMS NYU for the computational resources, to Julio Banga, Marisa Eisenberg, Nikki Meshkat and Maria Pia Saccomani for useful discussions, and to the referees for valuable comments.

Funding

This work was supported by the National Science Foundation [CCF-1563942, CCF-1564132, CCF-1319632, DMS-1760448, CCF-1708884]; National Security Agency [#H98230-18-1-0016]; and City University of New York [PSC-CUNY #69827-00 47, #60098-00 48].

Conflict of Interest: none declared.

References

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.

Hong,H. *et al.* (2018) Global identifiability of differential models. preprint, https://arxiv.org/abs/1801.08112.

Karlsson,J. *et al.* (2012) An efficient method for structural identifiability analysis of large dynamic systems. *IFAC Proc.* **45**, 941–946.

Ligon,T. *et al.* (2018) GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models. *Bioinformatics*, **34**, 1421–1423.

Meshkat,N. *et al.* (2014) On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: a novel web implementation. *PLoS One*, **9**, 1–14.

Motwani,R. and Raghavan,P. (1995) *Randomized Algorithms*. Cambridge University Press, New York.

Norton,J. (1982) An investigation of the sources of nonuniqueness in deterministic identifiability. *Math. Biosci.*, **60**, 89–108.

Sedoglavic,A. (2002) A probabilistic algorithm to test local algebraic observability in polynomial time. *J. Symbolic Comput.*, **33**, 735–755.

Thomaseth,K. and Saccomani,M.P. (2018) Local identifiability analysis of nonlinear ODE models: how to determine all candidate solutions. *IFAC-PapersOnLine*, **51**, 529–534.

Villaverde,A.F. *et al.* (2016) Structural identifiability of dynamic systems biology models. *PLoS Comput. Biol.*, **12**, 1–22.