Contents lists available at SciVerse ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Easy parameter identifiability analysis with *COPASI*

Jörg Schaber*

Institute for Experimental Internal Medicine, Medical Faculty, Otto von Guericke University, Magdeburg, Germany

ARTICLE INFO

ABSTRACT

Background and scope: Differential equation systems modeling biochemical reaction networks can only give quantitative predictions, when they are in accordance with experimental data. However, even if a model can well recapitulate given data, it is often the case that some of its kinetic parameters can be arbitrarily chosen without significantly affecting the simulation results. This indicates a lack of appropriate data to determine those parameters. In this case, the parameter is called to be practically non-identifiable. Well-identified parameters are paramount for reliable quantitative predictions and, therefore, identifiability analysis is an important topic in modeling of biochemical reaction networks. Here, we describe a hidden feature of the free modeling software *COPASI*, which can be exploited to easily and quickly conduct a parameter identifiability analysis of differential equation systems by calculating likelihood profiles. The proposed combination of an established method for parameter identifiability analysis with the user-friendly features of *COPASI* offers an easy and rapid access to parameter identifiability analysis even for non-experts.

Availability: COPASI is freely available for academic use at http://www.copasi.org.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The most favored formalism to describe the dynamics of biochemical reaction networks is ordinary differential equations (ODEs) (Klipp et al., 2005a). Accordingly, there exists a vast number of tools to implement, simulate and analyze ODEs (Klipp et al., 2007; Schaber and Klipp, 2011). One of the most popular tools is the software *COPASI* (Hoops et al., 2006; Klipp et al., 2007). *COPASI* is freely available, user-friendly and offers a number of useful tools to analyze ODEs. Specifically, *COPASI* has an exceptionally user-friendly interface to conduct parameter estimation of kinetic constants using experimental data.

Estimating parameters of a dynamic biochemical reaction networks is mandatory, if experimental data are to be described in a quantitative way. Estimated parameters can give valuable information about biological properties that are not observable. In addition, parameterized models can give quantitative predictions for a range of perturbations, which may be of interest or are difficult or expensive to measure (Bhalla and Iyengar, 1999). In *COPASI*, parameters are estimated by minimizing the weighted sum of squared residuals (*SSR*), which can be considered the log-likelihood function for normally distributed measurement noise (Seber and Wild, 2003). *COPASI* provides a number of algorithms to minimize the *SSR* and estimate parameters (Hoops et al., 2006).

Often it is the case, that the number of parameters in a model exceeds the number of data points used to fit those parameters (Hao et al., 2007; Klipp et al., 2005b; Kuhn and Klipp, 2012; Swameye et al., 2003). This may lead to the situation that a parameter can be chosen arbitrarily, because of the lack of data to determine it (Raue et al., 2009; Schaber et al., 2011). Apparently, in this situation, the parameter does not contain any information about the data. Such a parameter is called non-identifiable (Raue et al., 2009) and may be removed from the model, in case it also does not affect predictions. However, parameter identifiability is usually not checked and not reported in publications. This may partly be because of the lack of methods and tools to conduct a parameter identifiability analysis.

Recently, the implementation of a method for identifiability analysis was reported, which exploits one-dimensional likelihood profiles (Raue et al., 2009). This method is attractive, because it has a sound statistical basis (Murphy and Van der Vaart, 2000; Seber and Wild, 2003) and is easy to use and computationally feasible for modern simulation tools. This method was implemented in POTTERSWEEL (Maiwald and Timmer, 2008), which itself is freely available, but it is based on MATLAB, which is commercial software and, thus, might not be available for everyone. More importantly, the usage of MATLAB requires special training and skills and might, therefore, prevent the application of a parameter identifiability analysis for people not familiar with MATLAB.

Here, we demonstrate that the method for parameter identifiability based on one-dimensional likelihood profiles can easily be applied using the free software *COPASI*. The *COPASI*-approach for parameter identifiability analysis exploits a hidden feature of

* Tel.: +49 391 67 14453; fax: +49 391 67 13312.
  E-mail address: schaber@med.ovgu.de

COPASI, which has not been documented nor used in the literature before.

2. Software description

2.1. Algorithm

The method for parameter identifiability analysis using one-dimensional likelihood profiles is described in detail elsewhere (Raue et al., 2009). Shortly, to assess identifiability for each fitted parameter its corresponding likelihood profile is calculated. The likelihood profile $LP(p_i)$ for each fitted parameter $\hat{p}_i$ ($i = 1, \ldots, m$) is defined as:

$$LP(p_i) = \min_{p_{j \neq i}}(SSR(p_j))$$

i.e. re-optimizing the objective function value $SSR(p_j)$ with respect to all parameters $p_{j \neq i}$ for defined values of p_i in a neighborhood of the original estimated parameter value $\hat{p}_i$. The parameter $\hat{p}_i$ is called identifiable, if the re-optimized $SSR(p_j)$ exceeds a certain confidence limit within the tested interval. This confidence limit can approximately be calculated. According to asymptotic theory, approximate $100(1-\alpha)\%$ confidence regions based on likelihood contours C_{LC} or likelihood ratios C_{LR} are given by

$$C_{LC} = \left\{ p : SSR(p) \leq SSR(\hat{p}) \left(1 + \frac{m}{n-m} F^{\alpha}_{m,n-m} \right) \right\}$$

or

$$C_{LR} = \left\{ p : SSR(p) \leq SSR(\hat{p}) e^{\chi^2_\alpha / n} \right\}$$

where m is the number of parameters, n is the number of data points and $F^{\alpha}_{m,n-m}$ and χ^2_α are the upper α-critical values for the F-ratio distribution and the Chi-squared distribution, respectively (Seber and Wild, 2003).

2.2. Implementation

Once a model and a corresponding parameter estimation task is defined in *COPASI*, it is easy to calculate $LP(p_i)$ for each fitted parameter $\hat{p}_i$ using the built-in parameter scan task. The parameter scan task varies one or more parameters on a predefined scale and for each parameter value a certain subtask is performed. *COPASI* allows the parameter estimation task to be performed as a subtask in the parameter scan. Clearly, the parameter estimation subtask for each scanned value of the parameter p_i only has to report the actual value p_i and the corresponding value of $LP(p_i)$. With this information we can analyze whether the SSR exceeds the above defined confidence limits. This way we can perform a quick and easy parameter identifiability analysis on a discretized interval. Here, we shortly describe how $LP(p_i)$ can be reported. Please refer to the Supplementary Material for a detailed worked example.

Starting from the implemented model with defined parameter estimation task, the steps are the following:

(1) Define a new report and/or plot in the 'Output Specification' section, selecting the following items:
    a. The parameter of interest using the standard pop-up interface.
    b. The objective function value $LP(p_i)$ using the 'expert-mode' from the pop-up interface, selecting *COPASI* → ModelList[] → Root → TaskList[] → Parameter Estimation → Parameter Estimation → Best Value.
(2) Delete the parameter of interest from the parameter estimation task.

(3) Create a new 'Parameter Scan' task selecting the parameter of interest from the standard pop-up interface and define the interval and its discretization to be tested.
(4) Select 'Parameter Estimation' as a subtask.
(5) Select the above defined new report as the report in the 'Parameter Scan' interface.
(6) Save the model in a new file and run the 'Parameter Scan' task.

As a result of these steps a new tab-delimited text file and/or plot is created with the scanned values of p_i and the corresponding $LP(p_i)$ values. This file can be used to check whether within the scanned parameter range $LP(p_i)$ exceeds the above defined confidence limit and, therefore, is identifiable or not.

3. Application

The proposed combination of the established method of calculating likelihood profiles (Raue et al., 2009) for parameter identifiability analysis with the user-friendly features of *COPASI* offers an easy way to quickly test whether a parameter is identifiable or not, even for non-experts and people not familiar with complex software like MATLAB.

The described method gives a first impression about the identifiability of a parameter. Due to the discretization of the likelihood profile in the parameter scan, confidence limits can be missed or might only roughly be approximated. Therefore, different discretization schemes of the interval of interest should be tested. The likelihood profile is a smooth function by theory: therefore, a rough looking profile might also be due to improperly working parameter estimation. Reliably working parameter estimation is indeed a necessary condition for calculating the likelihood profile correctly and different parameter estimation methods should be tested as well. This is also easy to accomplish, because *COPASI* already has a range of methods implemented.

Given the fact that many biochemical models are SBML-compliant and can therefore readily be imported into *COPASI*, the described method offers a quick and easy way for a rough check of parameter identifiability. This might help the important concept of parameter identifiability to be used and applied more widely. Using this method parameter identifiability might also be reported on a more regular basis in publications of dynamic biochemical reaction networks.

A detailed application using a worked example is provided in the Supplementary Information.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biosystems.2012.09.003.

References

Bhalla, U.S., Iyengar, R., 1999. Emergent properties of networks of biological signaling pathways. Science 283, 381–387.

Hao, N., et al., 2007. A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway. Curr. Biol. 17, 659–667.

Hoops, S., et al., 2006. COPASI—a COmplex PAthway SImulator. Bioinformatics 22, 3067–3074.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H., 2005a. System Biology in Practice: Concepts, Implementation and Application. Wiley-VCH, Weinheim.

Klipp, E., et al., 2007. Systems biology standards—the community speaks. Nat. Biotechnol. 25, 390–391.

Klipp, E., et al., 2005b. Integrative model of the response of yeast to osmotic shock. Nat. Biotechnol. 23, 975–982.

Kuhn, C., Klipp, E., 2012. Zooming in on yeast osmoadaptation. Adv. Exp. Med. Biol. 736, 293–310.

Maiwald, T., Timmer, J., 2008. Dynamical modeling and multi-experiment fitting with POTTERWHELL. Bioinformatics 24, 2037–2043.

Murphy, S.A., Van der Vaart, A.W., 2000. On profile likelihood. J. Am. Stat. Assoc. 95, 449–465.

Raue, A., et al., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929.

Schaber, J., et al., 2011. Automated ensemble modeling with modelMaGe: analyzing feedback mechanisms in the Sho1 branch of the HOG pathway. PLoS One 6, e14791.

Schaber, J., Klipp, E., 2011. Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. Curr. Opin. Biotechnol. 22, 109–116.

Seber, G.A.F., Wild, C.J., 2003. Nonlinear Regression. Wiley Interscience.

Swameye, I., et al., 2003. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proc. Natl. Acad. Sci. U.S.A. 100, 1028–1033 (Epub 2003 January 1027).