Computers and Chemical Engineering 64 (2014) 24–40

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Performance comparison of parameter estimation techniques for unidentifiable models

J.E. Graciano, D.F. Mendoza [1], G.A.C. Le Roux *

Universidade de São Paulo, Escola Politécnica, Departamento de Engenharia Química, Av. Prof. Luciano Gualberto, trav. 3, 380 São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 25 March 2013
Received in revised form 10 January 2014
Accepted 14 January 2014
Available online 28 January 2014

Keywords:
Parameter estimation
Identifiability
Biased estimators
Chemical kinetics

ABSTRACT

Four different estimation approaches exploiting sensitivities, eigenvalue analysis (rotational discrimination and automatic parameter selection and estimation), reparameterization via differential geometry and the classical nonlinear least squares are assessed in terms of predictivity, robustness and speed. A Monte Carlo methodology is adopted to evaluate the statistical information required to quantify the inherent uncertainty of each approach. The results show that the rotational discrimination method presents the best characteristics among the evaluated methods, since it requires less a priori information than the reparameterization via differential geometry, uses simpler stop criteria than the automatic selection, reduces the overfitting caused by the nonlinear least squares solution and because it estimates parameters with the best predictivity among the methods tested. Additionally, results suggest that assessing the goodness of the estimated parameters solely in the calibration set can be misleading, and that the statistical information obtained from a validation set is more valuable.

1. Introduction

The increasing understanding of the mechanisms involved in physico-chemical systems together with the growing computational power have promoted the intensive use of mathematical modeling to predict the behavior of different processes under diverse conditions spanning from molecular to plant-wide scales (Chung, Jhon, & Biegler, 2011; Senkan, 1992). Mathematical models can be classified according to the relative amount of knowledge of the internal mechanisms used to describe a specific process as (Hangos & Cameron, 2001): empirical (black box), semi-empirical (gray box) and mechanistic (white box) models. Empirical models are entirely based on experimental input/output information without taking into consideration any information about the internal mechanisms of the system while mechanistic models are derived from the knowledge of the basic principles governing a specific process. Semi-empirical models are in between, since they include both basic principles and experimental information (used to fit purely mathematical correlations). The degree of complexity of a model increases as it includes more basic principles, i.e., as it incorporates more mechanistic description.

Theoretically while a model becomes mathematically more complex and more mechanistic, it would potentially allow a broader representation and prediction of a system behavior. However, the main disadvantage associated with complex models is the amount of information (theoretical and experimental) on the internal mechanism, hindered by the noise of the available measurements, which widens the possible sources of uncertainties. In this situation, identifiability problems are prone to take place.

A model is said locally (globally) identifiable when the objective function (OF) of the parameter estimation problem, e.g. least squares, has a local (global) minimum at an isolated point (Nguyen & Wood, 1982). The identifiability of a model can be analyzed from structural and practical points of view. The first one assesses if the functional form of the model (model structure) permits the determination of a unique parameter set of the parameters from noise-free measurements while the second evaluates if the quality (e.g., measurement noise) and quantity (statistical degrees of freedom e.g., few measured states in a bio/chemical reactor) of the available measurements allow such a determination in practice (Bellman & Åström, 1970; Raue et al., 2009).

Lack of structural identifiability implies lack of practical identifiability, but the opposite is not true (Miao, Xia, Perelson, & Wu, 2011), since structurally identifiable models might not be identifiable in practice due to limitations imposed by the quality and quantity of the available measurements and by the numerical difficulty to find the local minimum.

* Corresponding author. Tel.: +55 11 3091 1170; fax: +55 11 3813 2380.
E-mail addresses: galoroux@usp.br, galo@mail.pqi.ep.usp.br (G.A.C. Le Roux).
[1] Current address: Universidad Autónoma del Caribe, Departamento de Ingeniería Mecánica, Calle 90 No. 46-12, Barranquilla, Colombia.

0098-1354/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2014.01.009

Even if a model is structurally identifiable model predictions can be extremely insensitive to individual parameters or to parameter combinations. One of the main consequences of this is the ill-conditioning of the Hessian in the optimization problem used to find the estimate. The ill-conditioning can make the solution of the optimization problem impossible to evaluate (McLean & McAuley, 2012).

However, even if the solution could be calculated accurately (in a numerical sense, i.e., if the computational precision could be increased adequately in order to obtain precise results in spite of the very ill-conditioning) the solution is contaminated by overfitting. As a consequence, the regression coefficient vector estimated by least squares is expected to be far from the true parameter vector and negligible changes in the data can cause the least squares solution coefficients to assume very large absolute values and even to reverse signs (Marquardt, 1970).

In order to tackle the practical identifiability problem two approaches can be conceived, namely: obtaining more experimental information by generating more data points spanning different experimental conditions or modifying the model estimation procedure, applying mathematical strategies without adding new experimental data or modifying the experimental procedure.

The first approach tends to be costly, time consuming and sometimes physically infeasible, therefore a great deal of effort has been devoted to find methods to tackle the practical identifiability problem using the least amount of experimental data.

An approach to successfully reduce uncertainty in parameters can be obtained by providing a priori information about the physical process, model parameters or the estimator. This information can be used to modify the model structure (e.g. model reduction, reparameterization) or to identify a subset of identifiable parameters (e.g., via sensitivity analysis) (Miao et al., 2011).

Model reduction methods aim to reduce the model complexity using simplifying assumptions to decrease the number of equations, consequently, the number of parameters that should be estimated (Keesman, Spanjers, & Straten, 1998; Nikerel, van Winden, Verheijen, & Heijnen, 2009; Tjärnström & Ljung, 2002). It can be shown that this approach can be equivalent to introducing false a priori information, but it is common practice (Le Roux, 1995). A consequence of using model simplifications is that the reduced model might not be used to represent a wide range of conditions as the original one would.

In the reparameterization approach the original model is rearranged grouping some parameters in order to reduce its number. It is said that this approach, as well as the model reduction case, requires expertise to obtain a suitable transformation (Biegler, Damiano, & Blau, 1986; Surisetty, Hoz Siegler, McCaffrey, & Ben-Zvi, 2010). However, this problem has major analogies with the problem of finding what vectors to select as a base in a rank deficient linear algebra problem.

Ben-zvi (2008) proposed a reparameterization method for unidentifiable models via differential geometry, where the reparameterization is implemented by developing a transformation which partitions the parameter space into an estimable and an inestimable part. The estimable part of the parameter space is chosen based on a priori information about the system. This method does not require sensitivity calculations and is applicable over a wide variety of experimental conditions.

Sensitivity-based methods tackle the identifiability problem by determining the influence of the model parameters on the model outputs. In this way, it is possible to select which parameters should be fitted from the available information. Sensitivities can be classified as global and local. Global sensitivities serve to quantify the parameter influence over the whole search space, whereas local sensitivities show parametric influence locally (Chu, Huang, & Hahn, 2011; Haaker & Verheijen, 2004). The main drawback using local sensitivities is that they may change from point to point in the search space, therefore there is no certitude about the real importance of a parameter in the model. This issue becomes critical when sensitivities are used as the only criterion to decide which parameter subset should be adjusted using the available information. On the other hand, global sensitivities provide an unambiguous picture of the importance of a parameter in the model, at the expense of a high computational cost, thus, global sensitivities are employed to study the general behavior of mathematical models rather than to determine a specific solution (Sobol′, 2001), which is the aim of the present work.

Miao and coworkers (2011) report four typical local-sensitivity-based methods: correlation method, principal component analysis (PCA) method, eigenvalue method and orthogonal method. They point out that the last two methods outperform the two firsts, because they are better designed to evaluate and compare the influence of parameters values on the system outputs (Quaiser & Mönnigmann, 2009).

A simple and useful eigenvalue-based method is the rotational discrimination algorithm (Fariss & Law, 1979). This method performs a decomposition of the search space, such that it projects the least-squares direction onto a reduced space where the objective function decreases the more, using the spectral decomposition to deal with Hessian matrix near singularity, typical of unidentifiable systems. Thus, the search direction restricted to a principal-component projection helps to reduce model overfitting, compared to an unbiased parameter estimator.

The combination of orthogonality and eigenvalue analyses gives rise to a family of methods that automatically adjust a subset of model parameters while keeping the other ones at some nominal values. The challenge in this approach is to choose a parameter subset to fit the model, since the available data must be used to adjust the most relevant parameters. The objective of these methods is to determine how many and which parameters should be chosen to compose the subset of adjustable parameters.

Estimation methods based on automatic selection of parameters have been an intensive research field. Initially, Weijers and Vanrolleghem (1997) suggested a method to evaluate all possible permutations of model parameters, for this, they used the determinant and condition number of the Fisher Information Matrix (FIM) to choose the best parameter subset to fit the model. Li, Henson, & Kurtz (2004) presented a parameter ranking methodology using eigenvalue and orthogonality criteria, this algorithm starts with a PCA of FIM to find the most sensitive parameters, then, it continues choosing the parameters with less linearity index in relation to the parameters already chosen (using an orthogonality analysis), until completing the parameter subset to adjust the model. The number of elements in this subset was heuristically selected. Later, Lund and Foss (2008) proposed a method to determine the ideal number of elements in the parameter subset employing variance contribution analysis. Secchi, Cardozo, Neto, and Finkler (2006), on the other hand, improved the algorithm created by Li et al. (2004) employing predictability degradation and parameter correlation indexes, that gave rise to an algorithm for automatic selection of the parameter subset used to adjust the model.

In this work the performance of four methods representing different parameter estimation approaches (Rotational discrimination RD (Fariss & Law, 1979), Automatic parameter selection and estimation APS (Secchi et al., 2006), reparameterization via differential geometry RDG (Ben-zvi, 2008) and the classical nonlinear least squares LSq) are assessed in terms of quality of the parameters obtained (understood as the prediction capacity of the model on a validation set), robustness and speed, using a Monte Carlo (MC) strategy. The outcome of this study will be useful to evaluate the suitability of these methods to handle unidentifiable models, such as the ones encountered in real time optimization problems