A MINLP approach to model-based data mining for the quick development of nonlinear dynamic models

M. Quaglio[1,2], F. Bezzo[2], A. Gavriilidis[1], E. Cao[1] and F. Galvanin[1]

[1]Department of Chemical Engineering, University College London (UCL),
Gower St., WC1E 6BT London, United Kingdom

[2]Department of Industrial Engineering, University of Padua (UNIPD)
via Marzolo 9, 35131 Padua, Italy

Simple and reliable phenomenological models always represent an attractive and powerful instrument in several chemical and biological industrial processes. A trustworthy model can potentially predict the response of a system outside the investigated range of experimental conditions and can be fruitfully exploited for the purposes of process design and non-empirical process optimization. Following the seminal work by Box and Lucas [1] a number of works appeared in the scientific literature regarding the discrimination among candidate models and the precise identification of the model parameters through model-based design of experiments (MBDoE) techniques for model discrimination and parameter precision (PP) in nonlinear dynamic systems [2,3]. However, the application of these techniques always starts from the availability of an existing set of candidate models while direct guidelines for subsequent model improvement are not obvious. A well-established systematic technique for quick model development and enhancement has not been proposed yet.

A model may be affected by two types of weaknesses: *i*) a structural weakness, intrinsically associated to the mathematical structure of the equations and to model identifiability [4]; *ii*) a descriptive weakness (i.e. the model can be weak on representing the system under certain experimental conditions because of an incorrect or incomplete set of equations). In this work a new approach is presented to guide the improvement of the model through the introduction of a method for the automated detection of the second-type weaknesses. The proposed method is based on the solution of a MINLP problem whose aim is the maximization of the likelihood function [5] acting on the following variables: *i*) the set of parameters of the candidate model (parameters are treated as continuous variables); *ii*) a set of user-defined binary variables. The user-defined binary variables act like switchers, including or removing experimental data from the parameter estimation problem, on single or groups of measurements (e.g. one could be interested in evaluating the model's capability in the prediction of certain experimental conditions grouping the measurements gathered in the same experiment under the same switcher). The method works as a model-based data mining (MBDM) filter for parameter estimation (PE) simultaneously removing the experimental results that the model is not able to fit taking into account the measurement uncertainty. It allows for a quick, automated mapping in the space of experimental conditions in terms of good and bad model predictive capabilities, preventing the thoughtless use of fake optimal process points located in regions where the model is not reliable. The presented MBDM-PE technique is proposed as part of a wider framework for a systematic approach to model identification which is synthetically shown in Figure 1a. The procedure starts with a candidate model and preliminary statistical information on measurement errors. The distribution of the residuals associated to the measurements removed by the MBDM-PE filter identifies the main weaknesses of the proposed model, highlighting both critical experimental conditions and specific measured variables whose associated prediction is very poor. This provides a feedback to the gradual improvement of the model itself. Once the enhanced model is known to give a satisfactory description of the phenomenon in the investigated experimental conditions, statistically unsatisfactory parameter estimates are amended performing additional experiments designed through known MBDoE methodologies for improving the parameter precision [1,2,3].
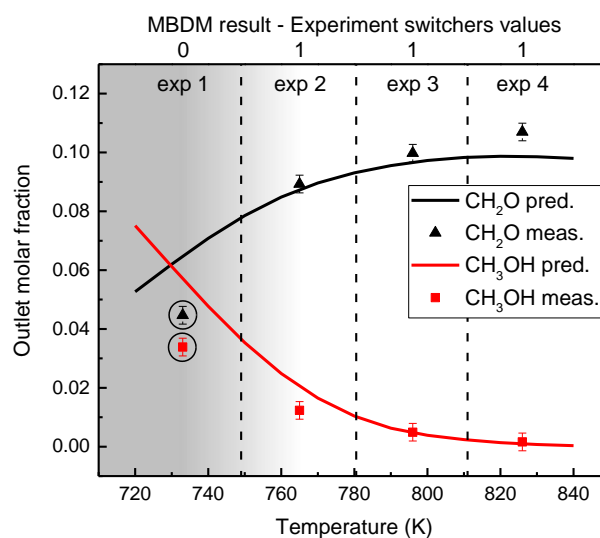
The MBDM-PE filtering technique has been successfully applied in a case study on the identification of a simplified reaction mechanism for the partial oxidation of methanol to formaldehyde over a silver catalyst in micro-reactor devices [6]. The investigation has been carried out adopting the optimization tool implemented in gPROMS Model Builder. The filter has automatically identified the poor predictive capabilities offered by a candidate kinetic model in the low temperature region of the experimental design space as shown in Figure 1b where the filtered (removed) experimental data have been highlighted with black circles. The underestimation of the conversion of methanol and the overestimation

of selectivity for the formaldehyde indicate a parallel reaction of complete oxidation for methanol, non-negligible at low temperatures.

Future works will focus on the development of meaningful methods for assessing the model reliability in non-investigated intermediate experimental conditions and systematic approaches for model building, given the model structure, for the quick, automated identification of phenomenological models.
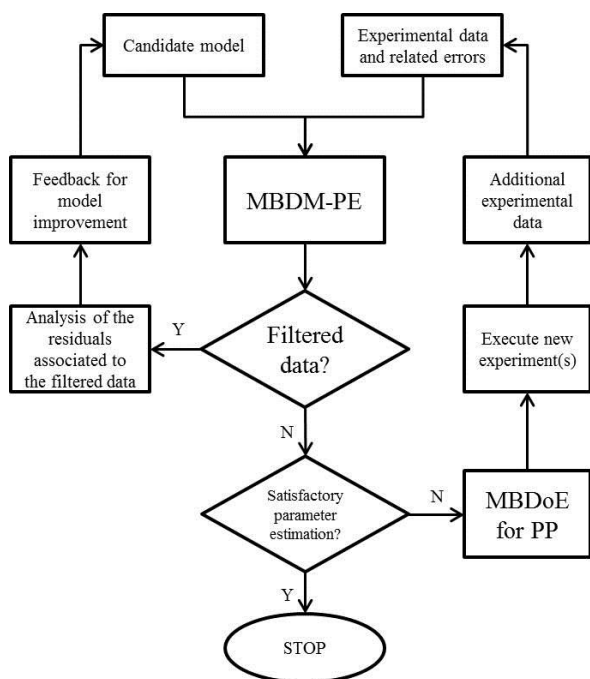
(a)

(b)

Figure 1. (a) Systematic approach for model identification implementing the MBDM-PE filtering technique; (b) Results given by the MBDM filter applied to a simplified kinetic model for partial methanol oxidation. Solid lines represent model predictions; scattered triangles and squares represent measurements collected in four experiments at different temperature. The experiment switchers in the upper part of the graph represent the final values evaluated by the solver for the user-defined binary variables indicating which experiments have been included in the PE problem (1) and which have been removed (0). Removed experimental data are also highlighted with black circles. The grey-coloured side in the plot indicates the region of low model reliability.

References

[1]   G.E.P. Box, H.L. Lucas (1959). Design of experiments in non-linear situations, Biometrika, 46, 77-90.

[2]   D. Espie, S. Macchietto (1989). The optimal design of dynamic experiments. AIChE Journal , 35(2), 223-229.

[3]   F. Galvanin, S. Macchietto and F. Bezzo (2007). Model-based design of parallel experiments. Ind. Eng. Chem. Res., 46, 871-882.

[4]   F. Galvanin, C.C. Ballan, M. Barolo and F. Bezzo (2013). A general model-based design of experiments approach to achieve practical identifiability of pharmacokinetic and pharmacodynamic models. J. Pharmacokinet. Pharmacodyn., 40, 451-467.

[5]   Yonathan Bard (1974), *Nonlinear parameter estimation*, ACADEMIC PRESS, New York and London.

[6]   F. Galvanin, E. Cao, N. Al-Rifai, V. Dua and A. Gavriilidis (2015). Optimal design of experiments for the identification of kinetic models of methanol oxidation over silver catalyst. Chemistry Today, 33(3), 51-56.