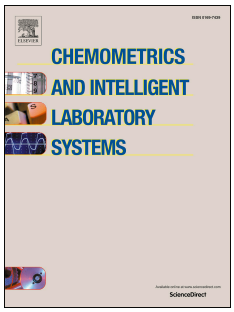
Accepted Manuscript



A model-based data mining approach for determining the domain of validity of approximated models

Marco Quaglio, Eric S. Fraga, Enhong Cao, Asterios Gavriilidis, Federico Galvanin

PII: S0169-7439(17)30448-3

DOI: 10.1016/j.chemolab.2017.11.010

Reference: CHEMOM 3543

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 5 July 2017

Revised Date: 13 October 2017

Accepted Date: 11 November 2017

Please cite this article as: M. Quaglio, E.S. Fraga, E. Cao, A. Gavriilidis, F. Galvanin, A model-based data mining approach for determining the domain of validity of approximated models, *Chemometrics and Intelligent Laboratory Systems* (2017), doi: 10.1016/j.chemolab.2017.11.010.

A model-based data mining approach for determining the domain of validity of approximated models

Marco Quaglio[a], Eric S. Fraga[a], Enhong Cao[a], Asterios Gavriilidis[a], Federico Galvanin[a,*]

[a] *Department of Chemical Engineering, University College London (UCL), Torrington Place, WC1E 7JE London, United Kingdom*

Abstract

Parametric models derived from simplifying modelling assumptions give an approximated description of the physical system under study. The value of an approximated model depends on the consciousness of its descriptive limits and on the precise estimation of its parameters. In this manuscript, a framework for identifying the model domain of validity for the simplifying model hypotheses is presented. A model-based data mining method for parameter estimation is proposed as central block to classify the observed experimental conditions as compatible or incompatible with the approximated model. A nonlinear support vector classifier is then trained on the classified (observed) experimental conditions to identify a decision function for quantifying the expected model reliability in unexplored regions of the experimental design space. The proposed approach is employed for determining the domain of reliability for a simplified kinetic model of methanol oxidation on silver catalyst.

Keywords: model identification, maximum likelihood, data mining, machine learning, model diagnosis

*Corresponding author
Email address: f.galvanin@ucl.ac.uk (Federico Galvanin)