

Manuscript Details

Manuscript number	CACE_2018_779_R1
Title	An online reparametrisation approach for robust parameter estimation in automated model identification platforms
Article type	Full Length Article

Abstract

Automated model identification platforms were recently employed to identify parametric models online in the course of unmanned experimental campaigns. The algorithms controlling these platforms include two computational elements: i) a tool for parameter estimation; ii) a tool for model-based experimental design. Both tools require the solution of complex optimisation problems and their effective outcome relies on their respective objective functions being well-conditioned. Ill-conditioned objective functions may arise when the model is characterised by a weak parametrisation, i.e. the model parameters are practically non-identifiable and/or extremely correlated. In this work, a robust reparametrisation technique is proposed and tested both in-silico and in an automated model identification platform. The benefit of reparametrisation is demonstrated on a case study for the identification of a kinetic model of catalytic esterification of benzoic acid with ethanol in a flow microreactor.

Keywords	online; identification; information; parametrization; design; experiment;
Manuscript category	Modeling, numerical analysis and simulation
Corresponding Author	Federico Galvanin
Corresponding Author's Institution	Chemical Engineering Department, University College London
Order of Authors	Marco Quaglio, Conor Waldron, Arun Pankajakshan, Enhong Cao, Asterios Gavriilidis, Eric Fraga, Federico Galvanin
Suggested reviewers	Filip Logist, Eva Balsa-Canto, Alexei Lapkin, Flavio Manenti, Jose Carlos Pinto

Submission Files Included in this PDF

File Name [File Type]

cover letter.pdf [Cover Letter]
CACE_2018_Response_to_reviewers.pdf [Response to Reviewers]
highlights.docx [Highlights]
abstract.docx [Abstract]
quaglio_online_rp.pdf [Manuscript File]
methodology.tif [Figure]
non_rp_condition_number.tif [Figure]
platform.tif [Figure]
real_case_ellipsoids.tif [Figure]
real_confidence_intervals.tif [Figure]
real_estimates.tif [Figure]
rp_condition_number.tif [Figure]
simulated_case_ellipsoids.tif [Figure]
simulated_confidence_intervals.tif [Figure]
simulated_estimates.tif [Figure]
quaglio_online_rp_markup.pdf [Supporting File]

Submission Files Not Included in this PDF

File Name [File Type]

elsart-harv-mod2.bst [LaTeX Source File]
quaglio_online_rp.tex [LaTeX Source File]
mybibfile.bib [LaTeX Source File]
mybibfile_custom.bib [LaTeX Source File]
methodology.png [Figure]
non_rp_condition_number.png [Figure]
platform.png [Figure]
real_case_ellipsoids.png [Figure]
real_confidence_intervals.png [Figure]
real_estimates.png [Figure]
rp_condition_number.png [Figure]
simulated_case_ellipsoids.png [Figure]
simulated_confidence_intervals.png [Figure]
simulated_estimates.png [Figure]

An online reparametrisation approach for robust parameter estimation in automated model identification platforms

Marco Quaglio, Conor Waldron, Arun Pankajakshan, Enhong Cao, Asterios Gavriilidis,
Eric S. Fraga and Federico Galvanin

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

corresponding author e-mail: f.galvanin@ucl.ac.uk

HIGHLIGHT LIST

- A technique for automated online model reparametrisation (RP) is introduced;
- Online RP is based on the eigendecomposition of the Fisher information matrix;
- RP improves the numerical robustness of online parameter estimation algorithms;
- The benefit of online RP is demonstrated on an experimental automated setup;

An online reparametrisation approach for robust parameter estimation in automated model identification platforms

Marco Quaglio, Conor Waldron, Arun Pankajakshan, Enhong Cao, Asterios Gavriilidis,
Eric S. Fraga and Federico Galvanin

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

corresponding author e-mail: f.galvanin@ucl.ac.uk

ABSTRACT

Automated model identification platforms were recently employed to identify parametric models online in the course of unmanned experimental campaigns. The algorithms controlling these platforms include two computational elements: i) a tool for parameter estimation; ii) a tool for model-based experimental design. Both tools require the solution of complex optimisation problems and their effective outcome relies on their respective objective functions being well-conditioned. Ill-conditioned objective functions may arise when the model is characterised by a weak parametrisation, i.e. the model parameters are practically non-identifiable and/or extremely correlated. In this work, a robust reparametrisation technique is proposed and tested both in-silico and in an automated model identification platform. The benefit of reparametrisation is demonstrated on a case study for the identification of a kinetic model of catalytic esterification of benzoic acid with ethanol in a flow microreactor.

An online reparametrisation approach for robust parameter estimation in automated model identification platforms

Marco Quaglio^a, Conor Waldron^a, Arun Pankajakshan^a, Enhong Cao^a,
5 Asterios Gavriilidis^a, Eric S. Fraga^a, and Federico Galvanin^{a,*}

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

*corresponding author e-mail: f.galvanin@ucl.ac.uk

Abstract

10 Automated model identification platforms were recently employed to identify parametric models online in the course of unmanned experimental campaigns. The algorithms controlling these platforms include two computational elements: *i*) a tool for parameter estimation; *ii*) a tool for model-based experimental design. Both tools require the solution of complex optimisation problems and their effective outcome relies
15 on their respective objective functions being well-conditioned. Ill-conditioned objective functions may arise when the model is characterised by a weak parametrisation, i.e. the model parameters are practically non-identifiable and/or extremely correlated. In this work, a robust reparametrisation technique is proposed and tested both in-silico and in an automated model identification platform. The benefit of reparametrisation
20 is demonstrated on a case study for the identification of a kinetic model of catalytic esterification of benzoic acid with ethanol in a flow microreactor.

keywords: online, identification, information, parametrization, design, experiment

1 Introduction

25 The kinetic modelling of chemical phenomena through the identification of an appropriate set of model equations is an important step in many research domains related to chemical engineering. Reliable kinetic models (i.e. models that accurately quantify the kinetic behaviour of the physical system) are regarded as key tools for supporting the design and intensification of chemical processes, performing non-empirical process optimisation and understanding which degrees of freedom in the physical system ultimately determine its ob-
30 servable behaviour (Berger et al., 2001). The identification of reliable models requires *i*) the determination of an opportune structure for the model equations and *ii*) the precise estimation of the model parameters. Both aspects typically require extensive amounts of time and resources for performing kinetic experiments. In the last decades, much effort

has been devoted by the scientific community to reducing the experimental burden required
35 to identify and validate kinetic models (Bonvin et al., 2016). Important steps towards the
reduction in the cost of kinetic studies are 1) the coupling of automated, small-scale flow
reactor technologies with online analysis equipment for the quick collection of experimental
data (Goodell et al., 2009) and 2) the employment of model-based design of experiments
(MBCoE) techniques for planning optimal experiments, minimise the cost, time and amount
40 of resources required for the experimentation (Asprey and Macchietto, 2000; Prasad and
Vlachos, 2008; Chakrabarty et al., 2013; Galvanin et al., 2013; Stamati et al., 2016).

Automated flow reactors have been employed in a wide variety of situations from process
monitoring (Malig et al., 2017) to screening of operating conditions (Walsh et al., 2005).
Automated flow reactors were also successfully coupled to algorithms for online sequential
45 design of experiments (McMullen and Jensen, 2010; Moore and Jensen, 2012; Fabry et al.,
2014; Holmes et al., 2016). After every experiment is terminated and new data are collected
by these platforms, algorithms construct black-box representations of the physical system
(e.g. response surfaces) for designing the following experiment with the aim of optimising
the reaction performance (e.g. the conversion or the yield). These self-optimising reactors
50 demonstrated the possibility for an automated platform of conducting experimental cam-
paigns with minimum human intervention. However, these platforms do not exploit the
collected data for the online development and identification of *physics-based* models. A ma-
jor consequence of this is that optimised reaction conditions identified through a black-box
approach in the lab-scale equipment are not necessarily transferable to the design, optimi-
55 sation and control of equipment at the industrial scale.

Only few works are available in the literature in which algorithms for online kinetic
model identification were coupled to automated reactor systems (McMullen and Jensen, 2011;
Bournazou et al., 2016; Echtermeyer et al., 2017). In these works, algorithms for parameter
estimation and optimal MBCoE were employed online to drive experimental campaigns
60 with the aim of selecting the best model among a set of given model structures (i.e. model
discrimination) (McMullen and Jensen, 2011) and/or improving the statistical quality of the
parameter estimates for a given model structure (McMullen and Jensen, 2011; Bournazou
et al., 2016; Echtermeyer et al., 2017). Automated model identification systems have the
potential of dramatically speeding up the modelling of kinetic phenomena and, consequently,
65 the discovery and the study of new chemical processes. However, the diffusion of these
promising systems in research laboratories is hampered by the high chance of numerical
failures whenever model identification algorithms are invoked.

The mathematical structure of kinetic models is frequently affected by problems of prac-
tical *identifiability*, i.e., the fitting quality of the data may be insensitive to a change in some
70 parameters and/or model parameters may be affected by extreme correlation. Whenever
the kinetic model exhibits this type of behaviour it is called *sloppy* (alternatively called ill-
conditioned model or poorly constrained model) (Chis et al., 2014) and its identification may
pose significant challenges even to state-of-the-art model identification algorithms (Asprey
and Naka, 1999; Transtrum et al., 2010, 2015; White et al., 2016). Parameter estimation
75 and optimal MBCoE problems are normally recast as optimisation problems and solved
numerically. In the presence of a sloppy parametrisation, the objective functions of the
aforementioned optimisation problems are ill-posed. The optimisation of ill-posed functions
may lead to significant numerical failures in the course of an unmanned experimental cam-

80 paign with the concomitant waste of experimental resources. Improving the robustness of automated model identification platforms towards model sloppiness is key to further promote their employment in the discovery and study of kinetic phenomena.

The main contribution of this manuscript is a computational strategy for online model reparametrisation (RP), i.e. a tool for transforming automatically the model parameter space in the course of the online model identification process. The presented tool is introduced to 85 enhance the robustness of unmanned platforms for model identification towards *numerical failures* derived by model sloppiness. Throughout this work, it is assumed that an opportune set of kinetic model equations is provided by the user to the model identification algorithm from the beginning of the unmanned experimental campaign. The benefit of the online RP is demonstrated experimentally on a case study where the objective is the identification of 90 a kinetic model of catalytic esterification of benzoic acid in a microreactor system.

2 Methods

2.1 Problem statement

An automated platform is available for performing experiments on a physical system of interest. An array \mathbf{y} of N_y physical quantities can be sampled by an online measurement 95 system. The kinetic behaviour of the physical system is described by a system of differential and algebraic equations as follows:

$$\begin{aligned} \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) &= \mathbf{0} \\ \hat{\mathbf{y}} &= \mathbf{h}(\mathbf{x}) \end{aligned} \tag{1}$$

In (1), \mathbf{f} is a $N_f \times 1$ array of model functions, \mathbf{x} is a $N_x \times 1$ array of state variables, $\dot{\mathbf{x}}$ is a $N_x \times 1$ array of time derivatives for the state variables¹, $\mathbf{u} \in U$ is a $N_u \times 1$ array of manipulable system inputs, t is time and $\boldsymbol{\theta} \in \Theta$ is a $N_\theta \times 1$ column array of model 100 parameters $\theta_1, \dots, \theta_{N_\theta}$. In (1), $\hat{\mathbf{y}}$ is a $N_y \times 1$ array of model predictions for the N_y measurable system states, expressed as a $N_y \times 1$ array of functions \mathbf{h} . It is assumed that the model (1) satisfies the requirements for structural identifiability, i.e., in principle, values of the model parameters $\boldsymbol{\theta}$ can be uniquely determined from the fitting of experimental data (Raue et al., 2009). The objective of the scientist is estimating the set of model parameters as precisely as 105 possible through an unmanned experimental campaign conducted on the automated platform given that the experimental budget allows for the collection of N_{MAX} samples of \mathbf{y} .

Whenever new data become available from the automated reactor system, the model identification algorithm is required to solve sequentially 1) a parameter estimation problem given the available dataset (Bard, 1974) and 2) a model-based design of experiments (MB- 110 DoE) problem to design the following experiment with the aim of minimising the predicted confidence region of parameter estimates (Franceschini and Macchietto, 2008b). The solution of both problems requires the employment of optimisation algorithms and their effectiveness

¹Only the derivatives of the states with respect to time are made explicit in the general model equations for simplicity of notation. However, in general, the model equations may be defined not only in the time domain, but also in the space domain and reactor model equations may involve functional relationships among partial derivatives of states with respect to time and space coordinates.

requires their respective objective function to be well-conditioned (Wilson et al., 2015; White et al., 2016). Ill-conditioned objective functions derive from the attempt of identifying models whose parametrisation is sloppy given the available dataset and the level of noise present in the system (Chis et al., 2014; White et al., 2016). Sloppiness arises when measured model responses are poorly sensitive to the change of some parameters and/or measurements do not carry sufficient information to bring parameter correlation below a critical threshold (typically considered as high as 95%). Whenever these circumstances occur, the eigenvalues of the covariance matrix of the parameter estimates span over a wide range of orders of magnitude, i.e. the condition number of the covariance matrix is very high.

Numerical failures may occur in the course of the model identification problem in the presence of a sloppy parametrisation. These may be classified as follows:

- *False convergence.* Ill-conditioned objective functions both in the parameter estimation and in the optimal MBDoE problem may cause numerical optimisation routines to fail in converging to the optimal solution (Higham, 1996).
- *Inaccuracy in the computation of gradients.* The calculation of the sensitivities (i.e. partial derivatives in the parameter space) using direct differential methods is frequently impractical. As a consequence, numerical differentiation routines are regularly employed in model building practice (Saltelli et al., 2000). The numerical computation of sensitivities requires a perturbation of the model parameter values. The computed sensitivities are *sensitive* to the choice of the perturbation. In the presence of a sloppy parametrisation, the applied perturbation may not be appropriate to accurately quantify the gradient in the parameter space (Higham, 1996). As a consequence, the Hessian and covariance matrix computed as functions of the parameter sensitivities may be inaccurate, affecting the model validation process and the design of following experiments (Pukelsheim, 2006).
- *Inaccuracy in the inversion of matrices.* In the presence of a sloppy parametrisation, the covariance matrix of the parameter estimates is ill-conditioned (White et al., 2016). The solution of an optimal MBDoE problem requires the inversion of an ill-conditioned covariance matrix if the parametrisation is sloppy (Franceschini and Macchietto, 2008b).

Different approaches have been proposed in the literature to address the identifiability problem of sloppy models (Dovi et al., 1994):

1. *Experimental-design-based (ED) methods.* These methods are based on the design of optimal experiments for *reshaping* the covariance matrix of the parameter estimates and improve the condition number. For more information on these approaches, the reader is referred to the relevant literature on design criteria for relaxing model sloppiness and reducing parameter correlation (Hosten, 1974; Pritchard and Bacon, 1978; Versyck and Van Impe, 1997; Galvanin et al., 2007; Franceschini and Macchietto, 2008a,d,c; Maheshwari et al., 2013; Chis et al., 2014; Wilson et al., 2015).

2. *Regularisation-based (RG) methods.* Regularisation involves the introduction of a bias in the parameter estimates with the aim of constraining their variance and, concomitantly, reducing the condition number associated to the parameter estimation problem (Barz et al., 2016). Popular regularisation techniques are *i*) Tikhonov regularization (Johansen, 1997; Hansen, 2005; Bardow, 2008) *ii*) truncated singular value decomposition (Hansen, 2005; Lopez C. et al., 2015) and *iii*) parameter subset selection (Barz et al., 2013; Lopez C. et al., 2015).
3. *Reparametrisation-based (RP) methods.* The aim of reparametrisation is transforming the original parameter space Θ into a robust parameter space Ω where both parameter estimation and MBDoE can be performed more effectively on well-conditioned objective functions (Agarwal and Brisk, 1985b,a). Although there is no theoretical advantage in the use of a reparametrised model (Rimensberger and Rippin, 1986; Dovi et al., 1994), the performance of model identification algorithms is sensitive to the type of parametrisation used (Espie and Macchietto, 1988). The effectiveness of RP-based methods has been recognised in many kinetic studies in the literature (Espie and Macchietto, 1988; Asprey and Naka, 1999; Benabbas et al., 2005; Schwaab and Pinto, 2007; Schwaab et al., 2008; Buzzi-Ferraris and Manenti, 2009).

These methods present strengths and weaknesses. ED-based methods are systematic. Optimal ED criteria to relax model sloppiness can be easily implemented into a computer program. However, even optimally designed experiments may not be sufficient to bring the condition number below critical levels. This weakness of ED-based methods is typically associated to either a too narrow range of explorable experimental conditions and/or an insufficient experimental budget to perform these optimal experiments. Furthermore, optimally designed experiments to reduce the condition number may not carry optimal amounts of information for the estimation of the model parameters. This limitation is typically overcome by designing experiments that represent a compromise between improving the parameter statistics and reducing the condition number (Franceschini and Macchietto, 2008c; Maheshwari et al., 2013).

An advantage of RG-based and RP-based methods is that they do not require the execution of experiments for improving the condition number and one can devote the entire experimental budget on improving the statistics of the parameter estimates. In RG-based approaches, the condition number is controlled through the introduction of prior information on the model parameter values. Systematic approaches, e.g. approaches based on Bayesian inference (MacKay, 1992), are available in the literature for supporting the selection of appropriate priors (Hansen, 2005). The introduction of prior information in the parameter estimation problem generally results in the computation of biased parameter estimates.

In contrast to RG-based approaches, RP-based methods do not involve the introduction of any bias in the model identification problem. Ad hoc strategies to reparametrise sloppy models were suggested for very specific kinetic model structures, e.g. Arrhenius-type reaction rates (Asprey and Naka, 1999; Schwaab and Pinto, 2007; Schwaab et al., 2008; Buzzi-Ferraris and Manenti, 2009). However, only few systematic approaches to the reparametrisation of sloppy models are available in the literature (Espie and Macchietto, 1988). An additional feature of RP-based methods is that whenever a model is reparametrised, the parametrisation

195 is fixed until the end of the experimental campaign. However, sloppiness is a consequence
of the combination of both the model parametrisation and the dataset available to identify
the model. There is no theoretical guarantee that the reparametrised model will not become
sloppy after the collection of new data (Wilson et al., 2015). The arising of sloppiness may be
200 i.e. by reparametrising the model every time new data are collected and included in the
parameter estimation problem. Nonetheless, online applications of RP-based methods seem
to be missing in the scientific literature.

In the following section, a RP-based framework for the identification of sloppy models
in automated model identification platforms is proposed. In the framework, a systematic
205 approach to model reparametrisation is introduced and applied online to maintain a small
condition number even when new data are collected by the automated system and included
in the parameter estimation problem.

2.2 Proposed methodology

The original set of equations (1) is initially extended including a linear system of equations
210 to transform the parameter space.

$$\begin{aligned}
\mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) &= \mathbf{0} \\
\hat{\mathbf{y}} &= \mathbf{h}(\mathbf{x}) \\
\boldsymbol{\theta} &= \mathbf{G}\boldsymbol{\omega}
\end{aligned} \tag{2}$$

In (2), $\boldsymbol{\omega} \in \Omega$ represents the $N_\theta \times 1$ array of model parameters in the transformed parameter space Ω , \mathbf{G} is a $N_\theta \times N_\theta$ matrix which transforms the parameter space Ω to the original model parameter space Θ . An online approach to model reparametrisation in automated
215 model identification platforms is now introduced with the aim of effectively estimating the
original parameter set $\boldsymbol{\theta} \in \Theta$. A block diagram showing the proposed procedure is given in
Figure 1. The procedure starts from the availability of preliminary experimental data and
the model structure (1). The parameter transformation \mathbf{G} is initially set equal to \mathbf{I}_θ , where
 \mathbf{I}_θ is the $N_\theta \times N_\theta$ identity matrix, i.e. the parameter spaces Θ and Ω are initially coincident.
220 The model identification algorithm is then called providing the available dataset as input.
The fundamental steps in the algorithm are now illustrated:

1. *A primary parameter estimation step.* At this stage, the set of transformed parameters $\boldsymbol{\omega}$ is estimated fitting the available dataset using a maximum likelihood approach (Bard, 1974). The Hessian of the likelihood function is then computed to characterise
225 the geometry of the parameter space and quantify its *sloppiness*.
2. *A parametrisation update step.* The Hessian matrix computed at the *primary parameter estimation* step is employed to compute and update the transformation matrix \mathbf{G} with the aim of minimising the condition number (i.e. eliminating the sloppiness) given the available dataset.

- 230 3. *A secondary parameter estimation step.* The model parameters $\boldsymbol{\omega} \in \Omega$ are estimated after the *parametrisation update* step and their statistical quality is quantified computing their covariance matrix \mathbf{V}_ω . Parameter estimates and related covariance computed in the transformed parameter space Ω are then transformed to the original parameter space Θ and returned as output.
- 235 4. *An optimal MBDoE for parameter precision step.* If parameter statistics in Θ are unsatisfactory and the experimental budget allows for additional samples to be collected, the experimental activity shall proceed. Optimal experimental conditions are identified at this stage through MBDoE techniques for parameter precision (Franceschini and Macchietto, 2008b) and transmitted to the automated platform for collecting the
- 240 next sample. Notice that in the proposed procedure the optimal MBDoE step occurs in the transformed parameter space Ω .

The illustrated steps constitute an iteration in the presented online framework. These are further detailed in the following subsections. The computational burden associated with the application of the proposed methodology is comparable with standard parameter estimation

245 algorithms based on parameter fitting. The procedure shows how it is possible to achieve an effective estimation of parameters in a (potentially) sloppy parameter space Θ by invoking the parameter estimation and the MBDoE algorithms in a conveniently transformed, non-sloppy, parameter space Ω . The values of the estimates and the related covariance obtained in the robust space Ω are transformed to the original parameter space Θ by applying linear

250 transformations, which are computationally more robust operations than optimisations.

2.2.1 Primary parameter estimation

The available dataset Y is provided to the model identification algorithm (see Figure 1). The dataset Y consists N samples of \mathbf{y} , i.e. $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$. It is assumed that the measurements for \mathbf{y} are affected by Gaussian noise with zero mean and covariance $\boldsymbol{\Sigma}$. The

255 transformation matrix \mathbf{G} is set equal to the *primary* transformation matrix \mathbf{G}_P . At the beginning of the model identification procedure \mathbf{G}_P is initialised as the identity matrix \mathbf{I}_θ . A primary estimation of the model parameters $\hat{\boldsymbol{\omega}}_P$ is performed as in (4) maximising the log-likelihood function (3).

$$\begin{aligned} \Phi(\boldsymbol{\omega}|Y)|_{\mathbf{G}=\mathbf{G}_P} = & -\frac{N}{2}[N_y \ln(2\pi) + \ln(\det(\boldsymbol{\Sigma}))] \\ & -\frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \hat{\mathbf{y}}_i(\boldsymbol{\omega})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\boldsymbol{\omega})] |_{\mathbf{G}=\mathbf{G}_P} \end{aligned} \quad (3)$$

$$\hat{\boldsymbol{\omega}}_P = \arg \max_{\boldsymbol{\omega} \in \Omega} \Phi(\boldsymbol{\omega}|Y)|_{\mathbf{G}=\mathbf{G}_P} \quad (4)$$

In (3), the quantity $\hat{\mathbf{y}}_i$ represents the model prediction for the sample \mathbf{y}_i . The negative Hessian \mathbf{H} of the log-likelihood function is then computed to evaluate the geometrical

260 properties of the log-likelihood profile in proximity of the maximum likelihood estimate as in (5).

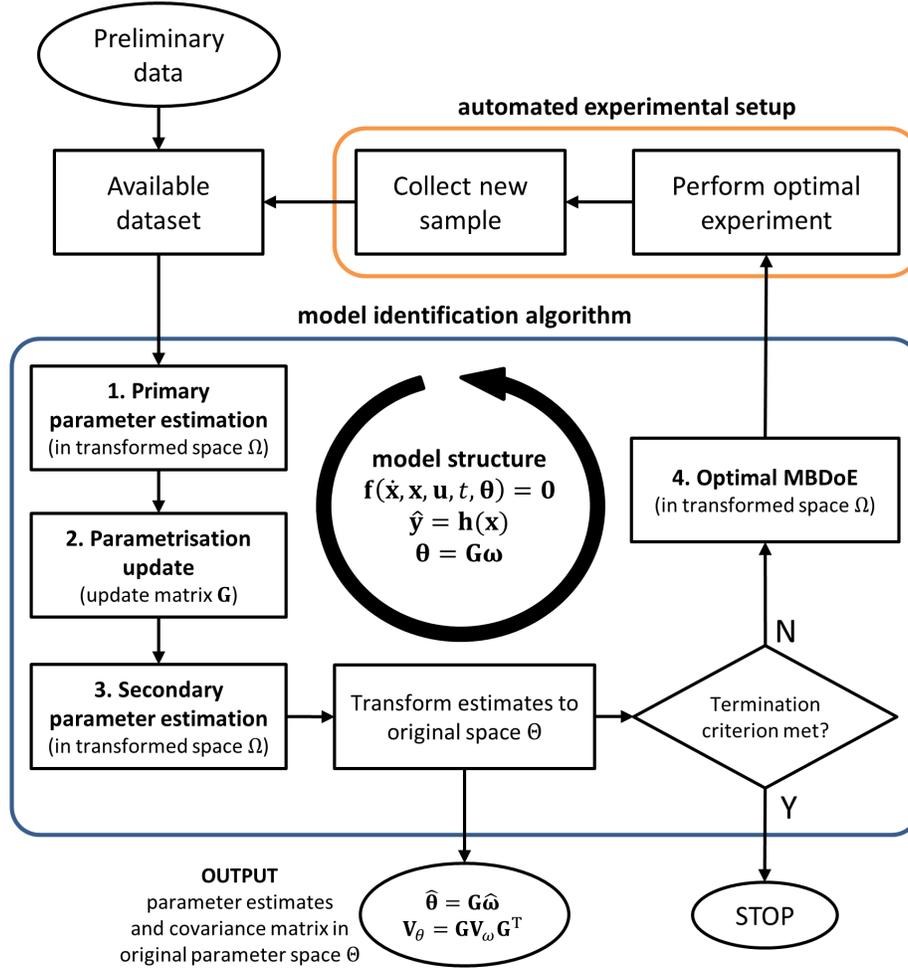


Figure 1: Proposed framework for the online identification of models in automated model identification platforms. Fundamental step in the procedure is the update of the parametrisation matrix \mathbf{G} after the collection and fitting of each sample. The online modification of the model parametrisation is performed to maintain a high computational performance at the parameter estimation and optimal MBDoe stages in the procedure.

$$\mathbf{H}(\hat{\boldsymbol{\omega}}_P)|_{\mathbf{G}=\mathbf{G}_P} = -\nabla\nabla^T\Phi(\hat{\boldsymbol{\omega}}_P|Y)|_{\mathbf{G}=\mathbf{G}_P} \quad (5)$$

In (5), the symbol ∇ defines the gradient operator in the parameter space Ω . Matrix \mathbf{H} is also known as the observed Fisher information matrix and its inverse quantifies the covariance matrix of the parameter estimates (Pukelsheim, 2006).

Notice that the model may be sloppy at the *primary parameter estimation* stage and the condition number may be very high, leading to numerical inaccuracy in the computation of the primary parameter estimate in (4) and in the computation of the Hessian in (5). Numerical results in Section 4 show that the performance of the online RP approach is not affected significantly by this aspect, but further analysis is required. Assessing the sensitivity of the proposed approach to numerical inaccuracies at the *primary parameter estimation* stage is going to be object of future research activities.

2.2.2 Parametrisation update

An eigendecomposition of the matrix (5) is performed at this stage with the aim of diagnosing the structure of the log-likelihood function in proximity of the maximum likelihood estimate and compute an opportune update to the transformation matrix \mathbf{G} . Let $\boldsymbol{\Lambda}$ be the diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \dots, \lambda_{N_\theta}$ of the observed Fisher information matrix (5). The eigenvalues of the observed Fisher information matrix represent the inverse eigenvalues of the covariance of the parameter estimates and the ratio between the maximum and the minimum eigenvalue represents the condition number κ .

$$\kappa = \frac{\max_i \lambda_i}{\min_i \lambda_i} \quad (6)$$

Let matrix \mathbf{U} be the matrix whose columns represent the right normalised eigenvectors of the observed Fisher information matrix (5). Matrix $\boldsymbol{\Lambda}$ and matrix \mathbf{U} quantify the sloppiness of the model in a more readable format. In fact, the eigenvalues and eigenvectors of the negative Hessian (5) respectively quantify the extent of the sloppiness and the directions in the parameter space which are associated to the sloppy behaviour of the model (Lopez C. et al., 2015). A family of *secondary* transformations \mathbf{G}_S can be constructed from \mathbf{G}_P , \mathbf{U} and $\boldsymbol{\Lambda}$ as in (7) with the aim of minimising the condition number of the problem (i.e. making $\kappa = 1.0$).

$$\mathbf{G}_S = d \mathbf{G}_P \mathbf{U} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{R} \quad (7)$$

The family of transformations given in (7) is parametrised by the scalar $d > 0$ and by the matrix \mathbf{R} , which represent respectively a scaling factor and a rotation matrix in the parameter space. The condition number κ is not influenced by the choice of d and \mathbf{R} . However, the omission of d and \mathbf{R} from (7) (the omission is equivalent to setting $d = 1.0$ and $\mathbf{R} = \mathbf{I}_\theta$ in (7)) may result in a transformation to a new parameter space in which there is significant discrepancy in the orders of magnitude of the model parameters. Model identification algorithms are influenced by the relative scale of parameters, e.g. in the computation of the gradients and, consequently, in the computation of the covariance of parameter estimates (Saltelli et al., 2000). Working with parameters sharing the same order of magnitude is

therefore desirable to avoid discrepancies on how the model identification algorithm handles different directions of the parameter space. In this work, the scaling factor d and the matrix \mathbf{R} are computed to map the *primary* parameter estimate $\hat{\boldsymbol{\omega}}_P$ into the parameter vector whose entries are all equal to 100.0 (this value was chosen arbitrarily) (Zhelezov, 2017). More specifically, the rotation applied by \mathbf{R} and the scaling factor d are computed to satisfy the equality $\mathbf{G}_P \hat{\boldsymbol{\omega}}_P = 100.0 \cdot \mathbf{G}_S \mathbf{1}_\theta$ where the vector $\mathbf{1}_\theta$ is the $N_\theta \times 1$ array whose entries are all equal to 1.0.

The secondary transformation matrix \mathbf{G}_S , computed as in (7), is then used to *update* the primary transformation matrix \mathbf{G}_P that will be used at the following iteration in the procedure of Figure 1.

2.2.3 Secondary parameter estimation

The aim at the *secondary parameter estimation* stage is obtaining a more accurate estimate for the parameters in the transformed space Ω . This is done by repeating the estimation of the parameters $\boldsymbol{\omega}$ after the *parametrisation update* stage, i.e. after the transformation of the (possibly) sloppy parameter space in a more robust, non sloppy parameter space. The log-likelihood function of the model is optimised as in (8) with $\mathbf{G} = \mathbf{G}_S$ obtaining the *secondary* parameter estimate $\hat{\boldsymbol{\omega}}_S$.

$$\hat{\boldsymbol{\omega}}_S = \arg \max_{\boldsymbol{\omega} \in \Omega} \Phi(\boldsymbol{\omega} | Y) |_{\mathbf{G}=\mathbf{G}_S} \quad (8)$$

In principle, the *primary* and the *secondary* parameter estimates satisfy the equality $\mathbf{G}_P \hat{\boldsymbol{\omega}}_P = \mathbf{G}_S \hat{\boldsymbol{\omega}}_S$. However, numerical algorithms for parameter estimation are sensitive to the model parametrisation (Rimensberger and Rippin, 1986; Dovi et al., 1994). More specifically, the convergence rate of numerical optimisation routines to the maximum likelihood estimate is sensitive to the choice of the transformation matrix \mathbf{G} and the aforementioned equality may not be satisfied in practice (Higham, 1996). The covariance \mathbf{V}_ω is then computed for the *secondary* parameter estimates as the inverse of the observed Fisher information matrix (Bard, 1974).

$$\mathbf{V}_\omega = [\mathbf{H}(\hat{\boldsymbol{\omega}}_S) |_{\mathbf{G}=\mathbf{G}_S}]^{-1} \quad (9)$$

The parameter estimates $\hat{\boldsymbol{\theta}}$ and their associated covariance matrix \mathbf{V}_θ in the original parameter space Θ are then computed by applying the *secondary* transformation to the estimates $\hat{\boldsymbol{\omega}}_S$ and covariance \mathbf{V}_ω computed in the transformed space Ω .

$$\hat{\boldsymbol{\theta}} = \mathbf{G}_S \hat{\boldsymbol{\omega}}_S \quad (10)$$

$$\mathbf{V}_\theta = \mathbf{G}_S \mathbf{V}_\omega \mathbf{G}_S^T \quad (11)$$

In standard parameter estimation algorithms, the computation of the covariance \mathbf{V}_θ requires the inversion of the information matrix in the original parameter space Θ (Bard, 1974). However, in the presence of a sloppy parametrisation, the information matrix in Θ may be ill-conditioned. Notice that, in the proposed framework, the inversion of ill-conditioned matrices is avoided. In fact, matrix inversion is performed in a conveniently transformed

parameter space Ω , as in (9), where the information matrix is well-conditioned. The covariance in the original parameter space \mathbf{V}_θ is then computed as in (11) by applying algebraic transformations, which are numerically more robust operations than matrix inversions.

From the covariance \mathbf{V}_θ , it is possible to derive the confidence intervals for the estimates $\hat{\boldsymbol{\theta}} \in \Theta$ and the correlation coefficient c_{ij} between any estimated parameter pair $\hat{\theta}_i$ and $\hat{\theta}_j$ (Bard, 1974). Let $v_{\theta,ij}$ be the ij -th element of the covariance matrix \mathbf{V}_θ . The confidence interval with significance α for the i -th parameter estimate $\hat{\theta}_i$ can be computed as $\hat{\theta}_i \pm z_{\alpha/2} \sqrt{v_{\theta,ii}}$ where $z_{\alpha/2}$ represents a two-tailed value computed from a standard normal distribution with significance α . The correlation coefficient between any parameter pair $\hat{\theta}_i$ and $\hat{\theta}_j$ can be computed according to (12).

$$c_{ij} = \frac{v_{\theta,ij}}{\sqrt{v_{\theta,ii}v_{\theta,jj}}} \quad \forall i, j \quad (12)$$

The statistical quality of the parameter estimates $\hat{\boldsymbol{\theta}}$ in the original parameter space Θ can be checked through a statistical test (e.g. a t -test) for assessing parameter precision (Walpole et al., 2011).

2.2.4 Optimal MBDoE for parameter precision

If some parameter statistics are not satisfactory and the experimental budget allows for the collection of additional data then the experimental activity will continue with the collection of an additional sample from the automated experimental setup. The following sample will be collected with the aim of minimising the size of the confidence region of the parameter estimates $\hat{\boldsymbol{\theta}} \in \Theta$. Popular measures of the size of the confidence region are (Galvanin et al., 2007; Franceschini and Macchietto, 2008b) *i*) the determinant of the covariance matrix $\det(\mathbf{V}_\theta)$ (i.e. the D-criterion), which quantifies the volume of the confidence ellipsoid of the parameter estimates and *ii*) the trace of the covariance matrix $\text{Tr}(\mathbf{V}_\theta)$ (i.e. the A-criterion), which quantifies the volume of the hyperbox containing the confidence ellipsoid.

Optimal MBDoE problems for parameter precision may be ill-conditioned in the presence of a sloppy parametrisation (White et al., 2016). In fact, the solution of an optimal MBDoE problem requires the inversion of an ill-conditioned matrix if the parametrisation is sloppy. In this work it is proposed to solve the MBDoE problem in the robust parameter space Ω with the aim of minimising the size of the confidence region in the original parameter space Θ . In general, the optimal experimental conditions depend on the type of criterion adopted for the design and on the model parametrisation. In this study, the D-criterion is employed because it is invariant under transformations of the parameter space (Fedorov, 1972; Rimensberger and Rippin, 1986). In fact, the following equality holds:

$$\det(\mathbf{V}_\theta) = \det(\mathbf{G}_S)^2 \det(\mathbf{V}_\omega) \quad (13)$$

It is sufficient to notice that matrix \mathbf{G}_S is not modified at the *optimal MBDoE* stage of the procedure (see Figure 1), i.e. $\det(\mathbf{G}_S)$ represents a constant in the MBDoE problem. Therefore, minimising the determinant of the covariance $\det(\mathbf{V}_\omega)$ in the transformed parameter space Ω is equivalent to minimising the determinant of the covariance $\det(\mathbf{V}_\theta)$ in the original parameter space Θ .

The optimal MBDoe problem in the robust space Ω requires the computation of a prediction for the parameter covariance $\hat{\mathbf{V}}_\omega$ (i.e. the posterior covariance matrix) after the collection of the new sample.

$$\hat{\mathbf{V}}_\omega = [\mathbf{V}_\omega^{-1} + \nabla \hat{\mathbf{y}}(\hat{\boldsymbol{\omega}}_s) \boldsymbol{\Sigma}^{-1} \nabla \hat{\mathbf{y}}(\hat{\boldsymbol{\omega}}_s)^T |_{\mathbf{G}=\mathbf{G}_s}]^{-1} \quad (14)$$

In (14), the second addend in the bracket represents the expected Fisher information matrix of the sample to be designed, which is a function of the experimental design vector $\boldsymbol{\varphi}$. The inverse of the prior covariance matrix \mathbf{V}_ω is also included in (14) to quantify the preliminary information that is available from previously fitted samples. The prior covariance is updated at every iteration of the procedure in Figure 1, i.e. after the collection of each sample, according to (9). The D-optimal experimental conditions $\boldsymbol{\varphi}^*$ for the collection of the following sample are computed solving the following optimisation problem:

$$\boldsymbol{\varphi}^* = \arg \min_{\boldsymbol{\varphi}} \det(\hat{\mathbf{V}}_\omega) \quad (15)$$

The optimised conditions computed as in (15) are then transmitted to the automated experimental setup for the collection of the following sample (see Figure 1).

3 Case study

The proposed algorithm presented in Section 2.2 is tested on a case study. The objective is the identification in an automated platform of a kinetic model of benzoic acid esterification with ethanol (Pipus et al., 2000). The reaction is homogeneous and it is catalysed by sulphuric acid. A description of the automated model identification platform is given in Section 3.1. The modelling assumptions are presented in Section 3.2. The proposed online RP methodology is tested both in-silico (Section 4.1) and experimentally on an automated model identification platform (Section 4.2). For both the simulated and the real cases two experimental campaigns are performed:

- a campaign where the parametrisation matrix is not modified;
- a campaign where the parametrisation matrix is updated online.

The two campaigns are performed to assess the influence of the online RP on the model identification process. The methods adopted for the conduction of the experimental campaigns are detailed in Section 3.3.

3.1 Automated model identification platform

A simplified diagram for the online model identification platform is given in Figure 2. The esterification of benzoic acid with ethanol catalysed by sulphuric acid occurs in a flow microreactor. The microreactor is a 2 m long PEEK tube with a diameter of 250 μm . It is placed in a stirred oil bath whose temperature is controlled by a rope heater. The reactants and the catalyst are injected through the flow reactor by three syringe pumps. Syringe 1 and

400 syringe 2 are filled with two different mixtures of benzoic acid and ethanol. The feed concentration of benzoic acid in the reactor is manipulated by modifying the relative flowrates of syringe 1 and syringe 2. Syringe pump 3 is filled with a 160 g L^{-1} sulphuric acid solution. The flowrate of syringe 3 is kept at 10% of the overall flowrate to maintain a constant concentration of sulphuric acid at 16 g L^{-1} at the inlet of the flow reactor. The mixture
405 at the outlet of the reactor is analysed online by a Jasco HPLC using a 250 mm long, 4.6 mm internal diameter ODS hypersil column with a particle size of $5 \text{ }\mu\text{m}$ (Thermo Fisher Scientific). The HPLC method uses 1.25 mL min^{-1} of a 40% water and 60% acetonitrile mobile phase (percentages refer to volume fractions). The oven is maintained at 303 K and a UV detection at 274 nm is used to detect the composition of the outlet mixture. Samples
410 are diluted using an online auto-sampler device (Syrris Asia) applying a dilution factor of 250.

The experimental conditions which can be manipulated by the automated system are:

- the inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$ in the range 0.9 - 1.55 mol L^{-1} ;
- the flowrate F of the feed mixture to the reactor in the range 7.5 - $30.0 \text{ }\mu\text{L min}^{-1}$;
- 415 • the temperature of the oil bath T in the range 343.0 - 413.0 K.

These constitute independent directions of the explorable space of experimental conditions $\varphi = [C_{\text{BA}}^{\text{IN}}, F, T]$. The experimental setup is controlled through a LabVIEW interface (Elliott et al., 2007) implemented in a 32-bit Windows machine with Intel Core i7-3770 3.40 GHz processor and 4.0 GB of RAM. A script written in Python 2.7 implementing the
420 model identification algorithm presented in Section 2.2 is integrated with LabVIEW for the purposes of online parameter estimation and sample design. The main Python packages employed in the script are NumPy 1.13 (Oliphant, 2015) for the manipulation of algebraic objects and SciPy 1.1 (Jones et al., 2001) for integrating the kinetic model equations and solving the optimisation problems associated with parameter estimation and MBD_{oE}.
425 Parameter estimation problems are solved using the *Nelder-Mead* method. MBD_{oE} problems are solved employing the *SLSQP* solver.

The *parametrisation update* stage of the algorithm (see Figure 1) was implemented in the Python script as an option that can be activated or deactivated from LabVIEW. This option was implemented to give more flexibility to the user in testing the model identification
430 algorithm both in the presence and in the absence of the online RP method.

3.2 Modelling assumptions

The catalytic esterification of benzoic acid and ethanol is modelled as a single reaction system where benzoic acid (BA) and ethanol (Et) react to produce ethyl benzoate (EB) and water (W) (Pipus et al., 2000).



435 Available studies in the literature report that the reaction is reversible. However, if a large excess of ethanol in the reactor is maintained (as in this work), the reverse reaction can be

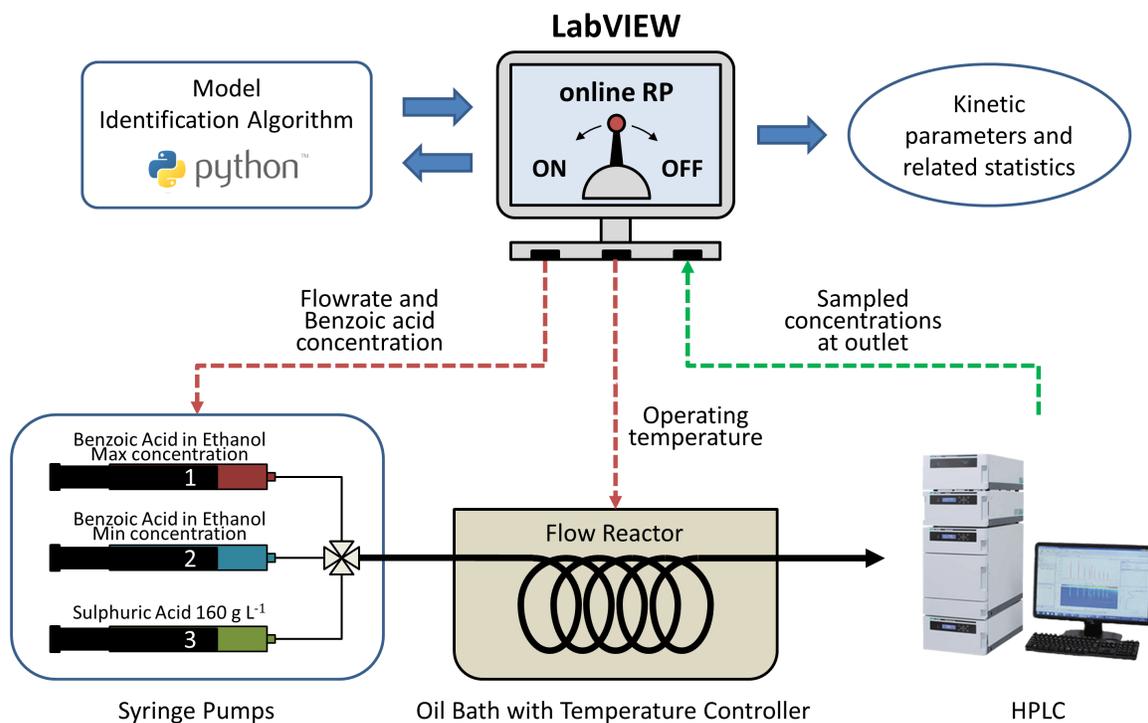


Figure 2: Simplified diagram representing the online model identification platform.

neglected (Pipus et al., 2000). The tubular reactor is modelled as an ideal plug flow reactor operated at isothermal conditions, i.e. thermal and mass transfer resistances are neglected. The validity of plug flow behaviour was checked by evaluating the vessel dispersion number (Levenspiel, 1998; Rossi et al., 2017). A maximum vessel dispersion number of $6.8 \cdot 10^{-4}$ was computed for the flowrate range considered in the study. The computed value is significantly smaller than $1.28 \cdot 10^{-2}$, i.e. the maximum vessel dispersion number recommended in the literature for the validity of the plug flow assumption (Levenspiel, 1998).

The reaction rate is assumed as first order with respect to benzoic acid. Following from the aforementioned assumptions, the steady-state kinetic behaviour of the system is modelled through the following set of ordinary differential equations (17):

$$v \frac{dC_i}{dz} = \nu_i k C_{BA}(z) \quad \forall i = \text{BA, Et, EB, W} \quad (17)$$

In (17), C_i is the concentration of the i -th component in the mixture expressed in molL^{-1} ; z represents the axial spatial coordinate of the tubular reactor expressed in m; v is the axial velocity of the liquid bulk expressed in m s^{-1} ; ν_i is the stoichiometric coefficient of the i -th component in the mixture; k is the rate constant expressed in s^{-1} .

An Arrhenius-type kinetic constant involving a set of two kinetic parameters $\theta = [\theta_1, \theta_2]$ is assumed with the following mathematical structure:

$$k = e^{\theta_1 - \frac{10^4 \theta_2}{RT}} \quad (18)$$

In (18), R is the ideal gas constant expressed in $\text{J mol}^{-1} \text{K}^{-1}$. As one can see from (18), the pre-exponential factor is included as exponent in the rate constant and the activation energy is multiplied by a scaling factor. The above structure for the kinetic rate constant was selected because it is generally recognised as robust within the literature on kinetic parameter estimation (Asprey and Naka, 1999; Buzzi-Ferraris and Manenti, 2009). In other words, parametrisation (18) generally leads to an improvement of the condition number with respect to the original form of the Arrhenius constant, i.e. $k = Ae^{-E_a/RT}$, parametrised by pre-exponential factor A and activation energy E_a .

3.3 Objective and methods

The objective of the study is the estimation of the kinetic parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$ with the smallest volume confidence region of $\hat{\boldsymbol{\theta}}$ by conducting an experimental campaign on the online model identification platform with an available budget of 9 samples. A sample is constituted by the single measurement of ethyl benzoate concentration at the outlet of the reactor, i.e. $\mathbf{y} = [C_{\text{EB}}^{\text{OUT}}]$ [mol L^{-1}]. The measurement error is modelled as Gaussian noise with covariance matrix $\boldsymbol{\Sigma} = [2.5 \cdot 10^{-5}]$, i.e. a standard deviation of 0.0165 molL^{-1} is assumed to model the Gaussian measurement noise for $C_{\text{EB}}^{\text{OUT}}$. The experimental conditions for the collection of samples 1, 2 and 3 are fixed to the values reported in Table 1. The following samples, i.e. samples from 4 to 9, are designed by the model identification algorithm by employing a D-optimal criterion, i.e. by solving an MBDoe problem in the form (15).

Two cases are proposed to test the model identification algorithm implemented in the online model identification platform:

1. *Simulated case: samples generated in-silico.* Samples are generated simulating the experiments with the kinetic model (17) setting the kinetic parameters equal to the value $\boldsymbol{\theta}^* = [15.27, 7.60]$ and adding Gaussian noise with covariance $\boldsymbol{\Sigma}$.
2. *Real case: samples collected from the experimental platform.* In this case, samples are collected from the experimental platform described in Section 3.1. An interval of 65 min is allowed between the collection of samples to let the system reach steady-state conditions.

For both the *Simulated* and the *Real* case, two experimental campaigns are performed: 1) a non-RP campaign in which the online reparametrisation is not activated; 2) a RP campaign in which the online reparametrisation is activated. This is done to provide a comparison of the performance of the model identification algorithm both in the presence and in the absence of the online RP method. In the *Simulated* case, the effect of the online RP is assessed comparing statistically the parameter estimates $\hat{\boldsymbol{\theta}}$ computed in the two campaigns with the target parameter value $\boldsymbol{\theta}^* = [15.27, 7.60]$. This is done by means of a χ^2 -test in the parameter space Θ . This involves testing the null hypothesis that the following statistic $\chi_{\hat{\boldsymbol{\theta}}}^2$ is distributed as a χ^2 distribution with degree of freedom $N_{\boldsymbol{\theta}} = 2$.

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \mathbf{V}_{\hat{\boldsymbol{\theta}}}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \chi_{\hat{\boldsymbol{\theta}}}^2 \sim \chi^2 \quad (19)$$

490 A small p -value associated to the statistic χ_{θ}^2 (e.g. smaller than 1.0%) is interpreted as an index of failure of the model identification algorithm in estimating the target parameter values. In the *Real* case, the target parameter value θ^* is unknown. Furthermore, a discrepancy in the parameter estimates between the RP and the non-RP campaigns is not only caused by numerical reasons, but also by problems of experimental repeatability caused by external
 495 disturbances (Alberton et al., 2009). The presence of disturbances can lead to changes in the parameters of the population from which experimental data are sampled and the concomitant inclusion of outliers in the dataset (Huber, 2004). It is recognised that, in the presence of such uncertainty sources, a statistical analysis to validate the models identified in the two campaigns would not be significant and it is therefore omitted.

500 Confidence intervals and correlation coefficient for the parameter estimates (see Section 2.2.3) are recorded in the course of the experimental campaigns and they are reported in the Results section. The condition number κ is also recorded in the course of the experimental campaigns and it is reported to demonstrate the performance of the online RP in improving and maintaining the well-posedness of the model identification problem.

Table 1: Experimental conditions φ adopted for the collection of samples 1 to 3 in the experimental campaigns: inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$; flowrate F ; temperature of the oil bath T .

Sample number	$C_{\text{BA}}^{\text{IN}}$ [mol L ⁻¹]	F [$\mu\text{L min}^{-1}$]	T [K]
1	1.50	20.0	413.0
2	1.00	10.0	393.0
3	1.25	15.0	403.0

505 4 Results

4.1 Simulated case: samples generated in-silico

The estimates for the kinetic parameters θ_1 and θ_2 for the non-RP campaign are reported in Table 2 together with information on their statistical quality. More specifically, the 95% confidence intervals and the correlation coefficient c_{12} between the kinetic parameters θ_1
 510 and θ_2 are reported. One can see from Table 2 that the correlation coefficient c_{12} remains above 99.96% in the course of the campaign. The parameter estimation and the MBDoE problems are solved in the original parameter space Θ where the condition number of the log-likelihood function remains above $6.1 \cdot 10^3$ throughout the whole experimental campaign. The χ^2 -test was conducted to compare statistically the computed parameter distribution
 515 with the target parameter value θ^* (see Section 3.3 for information on how the test statistic is computed). As one can see from Table 2, a p -value of 0.00% in the course of the non-RP campaign suggests that the parameter estimates computed by the algorithm are statistically inconsistent with the target parameter values.

Table 2: Simulated case: non-RP campaign. Parameter estimates are reported together with their respective 95% confidence intervals and correlation coefficient in the course of the experimental campaign. Parameter estimation and MBDoe problems are solved in the original parameter space Θ . The condition number of the log-likelihood function in Θ is reported in the table.

Simulated case - non-RP campaign					
Samples collected	Estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals		Correlation coefficient c_{12}	p -value of target parameters θ^*	Condition number κ in Θ
1	[- , -]		-	-	-
2	[- , -]		-	-	-
3	[12.15 \pm 2.14 , 6.56 \pm 1.35]		0.9998	0.00%	$1.4 \cdot 10^4$
4	[14.83 \pm 1.22 , 7.47 \pm 0.81]		0.9996	0.00%	$6.1 \cdot 10^3$
5	[15.99 \pm 1.01 , 7.85 \pm 0.70]		0.9998	0.00%	$1.0 \cdot 10^4$
6	[15.06 \pm 0.79 , 7.53 \pm 0.53]		0.9997	0.00%	$7.2 \cdot 10^3$
7	[14.90 \pm 0.74 , 7.47 \pm 0.50]		0.9997	0.00%	$9.2 \cdot 10^3$
8	[14.84 \pm 0.66 , 7.45 \pm 0.44]		0.9997	0.00%	$8.2 \cdot 10^3$
9	[14.94 \pm 0.63 , 7.49 \pm 0.42]		0.9998	0.00%	$9.6 \cdot 10^3$

Table 3: Simulated case: RP campaign. Parameter estimates in the course of the experimental campaign are reported together with their respective 95% confidence intervals and correlation coefficient. Parameter estimation and MBDoe problems are solved in the transformed parameter space Ω . The condition number of the log-likelihood function in Ω is reported in the table.

Simulated case - RP campaign					
Samples collected	Estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals		Correlation coefficient c_{12}	p -value of target parameters θ^*	Condition number κ in Ω
1	[- , -]		-	-	-
2	[- , -]		-	-	-
3	[16.44 \pm 64.52 , 8.01 \pm 25.05]		0.9999	0.00%	$5.5 \cdot 10^8$
4	[16.61 \pm 3.55 , 8.06 \pm 1.21]		0.9999	68.26%	$3.8 \cdot 10^2$
5	[15.60 \pm 2.01 , 7.72 \pm 0.68]		0.9998	86.41%	$1.2 \cdot 10^0$
6	[15.72 \pm 1.62 , 7.76 \pm 0.55]		0.9997	70.47%	$1.0 \cdot 10^0$
7	[15.72 \pm 1.50 , 7.76 \pm 0.51]		0.9998	69.84%	$1.0 \cdot 10^0$
8	[15.59 \pm 1.44 , 7.71 \pm 0.49]		0.9998	56.62%	$1.0 \cdot 10^0$
9	[15.39 \pm 1.24 , 7.64 \pm 0.42]		0.9998	64.74%	$1.0 \cdot 10^0$

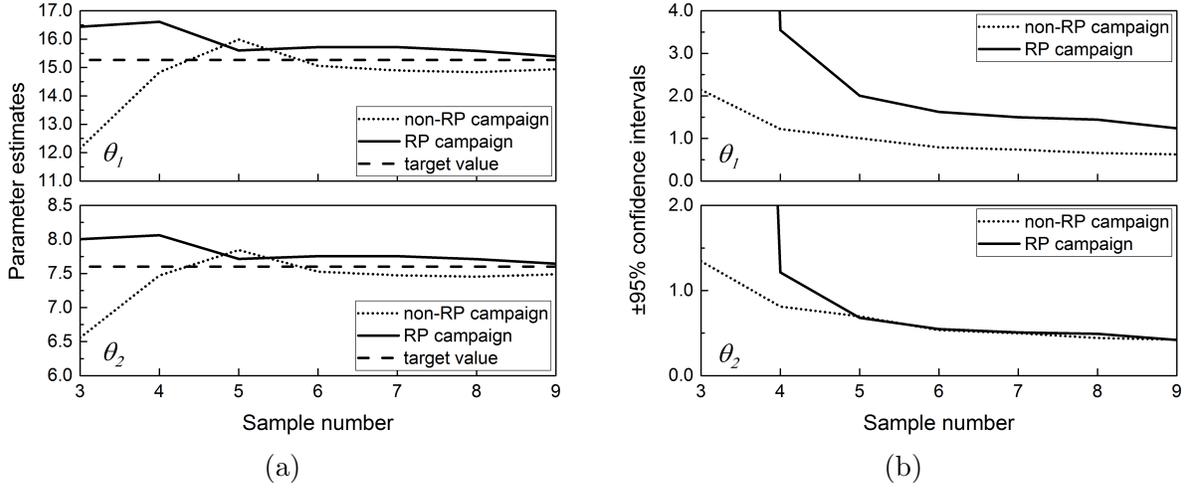


Figure 3: Simulated case: (a) parameter estimates and (b) $\pm 95\%$ confidence intervals throughout the non-RP campaign (dotted) and the RP campaign (solid). In subfigure (a), the target parameters are indicated by a dashed line.

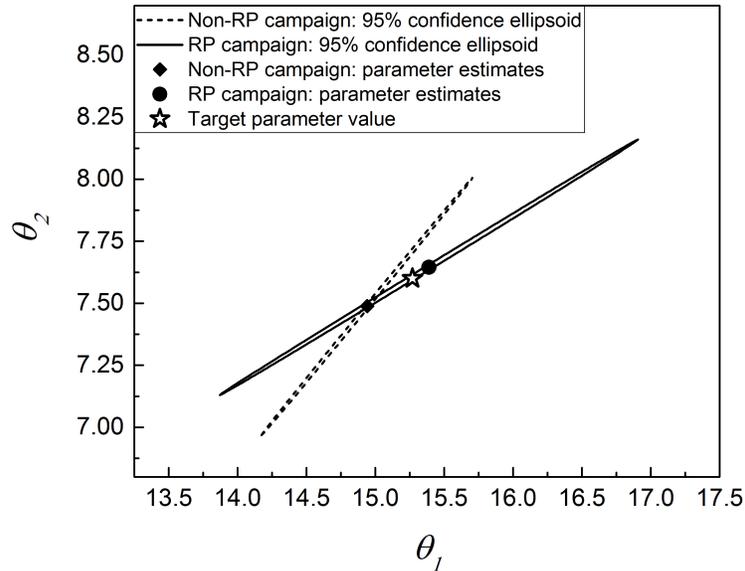


Figure 4: Simulated case: parameter estimates and related 95% confidence ellipsoids at the end of the non-RP campaign (dotted) and at the end of the RP campaign (solid). The target parameter value is highlighted in the graph by a star-shaped symbol.

Parameter estimates and related information on their statistical quality are given in
 520 Table 3 for the RP campaign. In the course of the RP campaign, the correlation coefficient
 c_{12} remains above 99.97%. In the RP campaign, the parameter estimation problem and the
 MBDoe problem are solved in the transformed parameter space Ω , where the transformation
 matrix \mathbf{G} is refined after the collection of each sample. The condition number of the log-
 likelihood function in Ω starts from a value of $5.5 \cdot 10^8$ at the first iteration of the model
 525 identification algorithm (i.e. after the collection of 3 samples) and it is reduced to 1.0 at
 the fourth iteration (i.e. after the collection of 6 samples). The benefit derived from the
 application of the online RP is validated by the χ^2 -test. The p -value of the target value θ^*
 given the computed covariance at the end of the model identification campaign is 64.74%.
 This confirms that the algorithm computed estimates that are statistically consistent with
 530 the target parameter value θ^* .

The parameter estimates and related 95% confidence intervals obtained in the non-RP
 campaign and in the RP campaign are compared graphically in Figure 3a and Figure 3b.
 In Figure 3a, one can see that both the methods present a similar convergence to the target
 parameter values, highlighted with dashed lines in the plot. In Figure 3b, one can see
 535 that the 95% confidence intervals for the parameters are significantly different between the
 non-RP and the RP campaign. In particular the confidence interval of parameter $\hat{\theta}_1$ is
 significantly larger in the RP case than in the non-RP case. The discrepancy is interpreted
 as a consequence of an inaccurate computation of the log-likelihood gradient in the non-RP
 case, which results in an underestimation in the variance of the estimate $\hat{\theta}_1$ (see Section 2.1
 540 for more details).

The final parameter estimates obtained in the non-RP and in the RP campaigns in
 the simulated case are compared graphically in Figure 4. In Figure 4 the final parameter
 estimates are plotted with their respective 95% confidence ellipsoids for the non-RP campaign
 (dotted) and for the RP campaign (solid). The target parameter value is highlighted in Figure
 545 4 by a star-shaped symbol. As one can see from Figure 4 the target value lies within the
 solid ellipsoid of the RP campaign, while it lies outside the dotted ellipsoid of the non-RP
 campaign. The graph shows that the non-RP campaign leads to the misleading conclusion
 that the target parameter values are not the parameters values of the physical system. The
 RP campaign led to a more reliable estimate of the kinetic parameter values.

550 Additional campaigns were performed in-silico to demonstrate that the performance of
 the model identification algorithm is insensitive to a change in the dataset, i.e. it is insensitive
 to a change in the random seed used to generate the data in-silico. The results obtained from
 20 simulated campaigns are reported in Appendix A. Both in RP and non-RP campaigns,
 each algorithm iteration required only few seconds of CPU time.

555 4.2 Real case: samples collected from the experimental platform

Two campaigns of experiments, i.e. a non-RP campaign and a RP campaign, were per-
 formed on the automated system. Experimental conditions investigated in the course of the
 campaign and the associated sampled concentrations are given in Appendix B. Parameter
 estimates $\hat{\theta}$ with associated confidence intervals and correlation coefficient are reported in
 560 Table 4 for the non-RP campaign and in Table 5 for the RP campaign. Numerical estimates
 in terms of pre-exponential factor and activation energy were also computed from $\hat{\theta}$. These

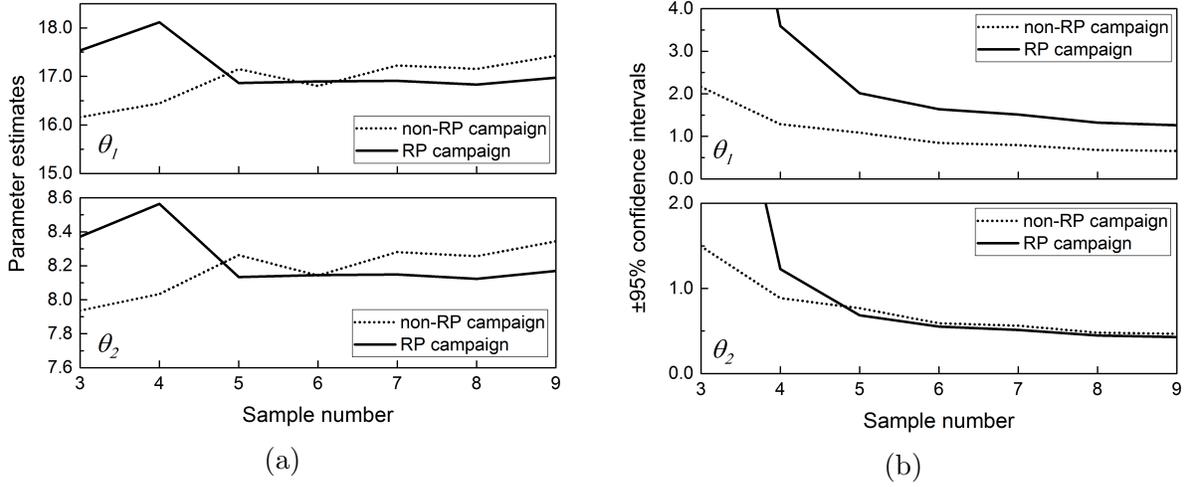


Figure 5: Real case: (a) parameter estimates and (b) $\pm 95\%$ confidence intervals throughout the non-RP campaign (dotted) and the RP campaign (solid).

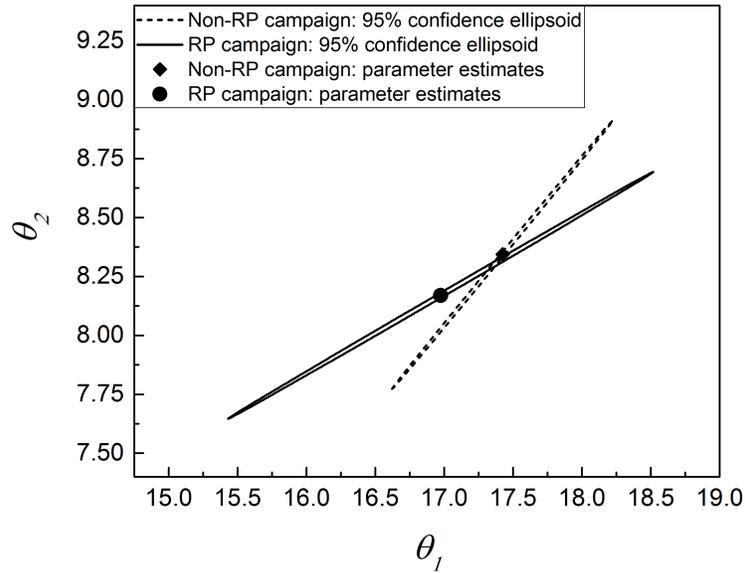


Figure 6: Real case: parameter estimates and related 95% confidence ellipsoids at the end of the non-RP campaign (dotted) and at the end of the RP campaign (solid).

Table 4: Real case: non-RP campaign. Parameter estimates in the course of the experimental campaign with 95% confidence intervals and correlation coefficient. Parameter estimation and MBDoE problems are solved in the original parameter space Θ . The condition number of the log-likelihood function in Θ is reported in the table.

Real case - non-RP campaign					
Samples collected	Estimates $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals			Correlation coefficient c_{12}	Condition number κ in Θ
1	-	,	-	-	-
2	-	,	-	-	-
3	16.16 \pm 2.16	,	7.94 \pm 1.49	0.9998	$1.5 \cdot 10^4$
4	16.44 \pm 1.29	,	8.03 \pm 0.89	0.9996	$6.1 \cdot 10^3$
5	17.15 \pm 1.09	,	8.26 \pm 0.77	0.9998	$1.1 \cdot 10^4$
6	16.80 \pm 0.85	,	8.14 \pm 0.59	0.9997	$7.8 \cdot 10^3$
7	17.23 \pm 0.79	,	8.28 \pm 0.56	0.9998	$1.1 \cdot 10^4$
8	17.15 \pm 0.68	,	8.26 \pm 0.48	0.9997	$8.4 \cdot 10^3$
9	17.42 \pm 0.66	,	8.34 \pm 0.47	0.9998	$1.0 \cdot 10^4$

Table 5: Real case: RP campaign. Parameter estimates in the course of the experimental campaign with 95% confidence intervals and correlation coefficient. Parameter estimation and MBDoE problems are solved in the transformed parameter space Ω . The condition number of the log-likelihood function in Ω is reported in the table.

Real case - RP campaign					
Samples collected	Estimates $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals			Correlation coefficient c_{12}	Condition number κ in Ω
1	-	,	-	-	-
2	-	,	-	-	-
3	17.54 \pm 13.41	,	8.37 \pm 5.38	0.9999	$2.6 \cdot 10^7$
4	18.12 \pm 3.59	,	8.56 \pm 1.23	0.9999	$8.0 \cdot 10^2$
5	16.86 \pm 2.01	,	8.13 \pm 0.68	0.9998	$1.3 \cdot 10^0$
6	16.90 \pm 1.64	,	8.15 \pm 0.55	0.9997	$1.0 \cdot 10^0$
7	16.91 \pm 1.51	,	8.15 \pm 0.51	0.9998	$1.0 \cdot 10^0$
8	16.83 \pm 1.32	,	8.12 \pm 0.45	0.9997	$1.0 \cdot 10^0$
9	16.98 \pm 1.26	,	8.17 \pm 0.43	0.9998	$1.0 \cdot 10^0$

are reported in Appendix B.

In the course of the non-RP campaign (see Table 4), the parameter correlation c_{12} between $\hat{\theta}_1$ and $\hat{\theta}_2$ remains above 99.96%. In the non-RP campaign the parameter estimation and MBDoe problems are solved in the original parameter space Θ . The condition number of the log-likelihood function in Θ remains above $6.1 \cdot 10^3$ in the course of the non-RP campaign.

The correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is above 99.97% throughout the whole RP campaign (see Table 5). However, in the RP campaign, parameter estimation and MBDoe problems are solved in the transformed parameter space Ω . The condition number in Ω is reduced by the algorithm from an initial value of $2.6 \cdot 10^7$ to the minimum value 1.0 in four iterations (i.e. after the collection of 6 samples). The transformation matrix \mathbf{G} is then adjusted after the collection of each sample to maintain a condition number $\kappa = 1.0$ until the end of the experimental campaign.

The parameter estimates and related 95% confidence intervals obtained in the non-RP and in the RP campaigns are plotted in Figure 5a and Figure 5b. The 95% confidence ellipsoids associated to the final parameter estimates achieved in the non-RP campaign and in the RP campaign are plotted in Figure 6.

Notice that in this case it is not possible to quantify and compare the performance of the two campaigns in retrieving the target parameter value. The target kinetic parameters are in fact unknown in the real case. One can observe from Figure 5a that the estimates achieved in the RP campaign exhibit a convergent behaviour around the values $\boldsymbol{\theta} = [16.90, 8.15]$. Estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ in the non-RP campaign do not exhibit a convergent behaviour, but they tend to increase in the course of the non-RP campaign (see Figure 5a). It is not possible to assess whether the absence of convergence in the non-RP campaign is the consequence of an unknown systematic disturbance in the system. However, it is possible to appreciate that the application of the online RP method led to the minimisation of the condition number (see Table 5) with the concomitant improvement in the numerical performance of the optimisation algorithms. Also in the real case, both in the RP and in the non-RP campaign, the CPU time required to complete each algorithm iteration was on the order of seconds.

A goodness-of-fit test was also conducted to demonstrate that, both in the RP and in the non-RP campaign, the postulated first order single-reaction mechanism (see Section 3.2) provided an accurate representation of the chemical system. Nonetheless, it was recognised that an analysis on the goodness-of-fit was not significant for demonstrating the online RP method. It was chosen to report in Appendix B the numerical details regarding the analysis on the fitting quality.

4.3 Results discussion

Both in the simulated and in the real case, the 95% confidence intervals of the estimates after 9 collected samples differ significantly between the non-RP and the RP campaign (see Figure 3b and Figure 5b). In the simulated case, a χ^2 -test was conducted to compare the final statistics on the parameters computed in the RP campaign with the final statistics obtained in the non-RP campaign. It was shown that the confidence region of the parameter estimates computed in the RP campaign *contains* the target parameter value $\boldsymbol{\theta}^*$ while the ellipsoid computed in the non-RP campaign does *not contain* the target value $\boldsymbol{\theta}^*$ (see Figure 4). Hence, it was possible to demonstrate statistically that the campaign with online RP

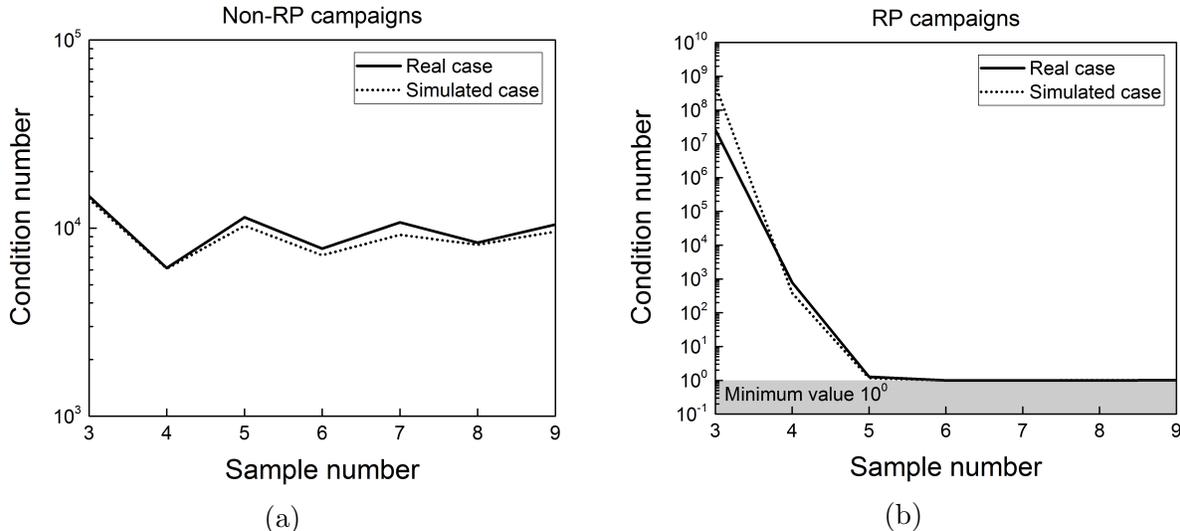


Figure 7: Condition number after each sample collected in the simulated case (dotted line) and in the real case (solid line): (a) non-RP campaigns; (b) RP campaigns.

605 led to a more accurate quantification of the uncertainty region associated to the computed parameter estimates.

Figure 7a and Figure 7b show the condition numbers in the course of the non-RP and RP campaigns respectively. In the non-RP campaigns (see Figure 7a), the condition number κ is around 10^4 and does not vary significantly in the course of the sample collection process. In the RP campaigns, both in the simulated and in the real case, the employment of the online RP method led to the minimisation of the condition number to $\kappa = 1.0$ in an initially ill-conditioned model identification problem (see Figure 7b). From Figure 7b, one can see that, both in the simulated and in the real case, the condition number is minimised to $\kappa = 1.0$ when sample 6 is collected, i.e. after 4 iterations in the model identification algorithm. This is explained by the fact that the update for the transformation matrix \mathbf{G} is evaluated as a function of the Hessian \mathbf{H} computed with the primary transformation matrix \mathbf{G}_P (see Section 2.2).

The condition number in the transformed space associated to \mathbf{G}_P may be very high at the first iteration of the algorithm. A high condition number at the *primary parameter estimation* step may lead to an inaccurate computation of the Hessian (i.e. an inaccurate quantification of the sloppiness) and, consequently, lead to the computation of an inappropriate update for \mathbf{G} . This does not appear to affect the performance of the online RP approach in the presented case study, but further analysis is required. It is object of future research activities to make the proposed algorithm insensitive towards numerical inaccuracies in the initial diagnosis of model sloppiness.

5 Conclusion

A model identification algorithm implementing a novel approach of online reparametrisation, i.e. an approach of online transformation of the model parameter space, is proposed

in this manuscript. The tool is designed specifically to deal with the problem of parameter estimation in the presence of *sloppy* model structures, i.e. models whose parameters are practically non-identifiable and/or highly correlated. The proposed approach to online reparametrisation is based on two fundamental steps: 1) a primary parameter estimation step, which is required to diagnose and quantify the sloppiness of the model parameter space; 2) a parametrisation update step in which the sloppy parameter space is transformed to a robust parameter space with the aim of reducing the sloppiness. Once the model parametrisation is updated, the parameter estimation is repeated solving an optimisation problem in the transformed, non-sloppy, parameter space. Additional samples are then designed by the algorithm employing techniques for optimal design of experiments with the aim of improving the statistical quality of parameter estimation. It is shown that optimisation algorithms benefit significantly from the presence of a robust (i.e. non-sloppy) model parametrisation both at the parameter estimation and at the experimental design stage. Parameter estimates computed in the robust space are then transformed to the original, sloppy, parameter space applying algebraic transformations and returned as output to the user.

The performance of the presented algorithm was tested both in-silico and on a real system where an automated experimental platform has been employed for online kinetic model identification. The objective was the estimation of the kinetic parameters in a two-parameter model of benzoic acid esterification with ethanol catalysed by sulphuric acid in a flow reactor. In both cases, the reparametrisation algorithm iteratively reduced and eventually eliminated model sloppiness minimising the condition number of an originally ill-conditioned model identification problem. In the case study performed in-silico, a set of values for the kinetic parameters was assumed to simulate the experiments. Hence, it was possible to show that the ill-conditioned nature of the model was preventing a conventional model identification algorithm from retrieving the target value of the kinetic constants. The presented model identification algorithm implementing the proposed online reparametrisation method was instead capable of computing estimates that were statistically compatible with the assumed target values of the kinetic parameters.

In the real case, the target values for the kinetic constants were unknown and it was not possible to quantify directly the convergence of the parameter estimates to the target kinetic coefficients. However, it was possible to appreciate that, also in the experimental campaign performed on the real system, the model identification algorithm implementing the online reparametrisation routine iteratively reduced and eventually minimised the condition number. The minimisation of the condition number to unity and the concomitant elimination of model sloppiness resulted in an improved performance of optimisation algorithms employed in the course of the model identification process.

The proposed reparametrisation method was integrated as an optional step in an on-line model identification algorithm implemented in a Python script. It was shown that the computational performance of the algorithm was not affected significantly by the additional step of model reparametrisation. The modest computational cost associated to the reparametrisation step and the low memory requirement of the method make it suitable for implementation also on embedded devices. Future research activities will focus primarily on three aspects: 1) improving the efficiency of the proposed method by reducing the number of iterations required to minimise the condition number and eliminate model sloppiness; 2) validating the proposed approach on more complex model structures, e.g. kinetic models in-

675 involving a higher number of parameters and multiple measured model responses; 3) extending the proposed online reparametrisation algorithm including routines for applying nonlinear transformations to the parameter space.

Acknowledgements

This work was supported by the Hugh Walter Stern PhD Scholarship, University College London; the Department of Chemical Engineering, University College London; and the EP-
680 SRC Grant EP/J017868/1.

Symbols used

Latin symbols

A	pre-exponential factor
c_{ij}	correlation coefficient between θ_i and θ_j
C_i	concentration of species i
C_i^{IN}	concentration of species i at the inlet
C_i^{OUT}	concentration of species i at the outlet
d	scaling factor of parameter space (> 0)
E_a	activation energy
F	volumetric flowrate
k	kinetic constant
N	number of samples in the available dataset Y
N_f	number of functions in a given kinetic model
N_{MAX}	maximum number of samples collectable
N_u	number of independent inputs in a given kinetic model
N_x	number of state variables in a given kinetic model
N_y	number of output variables in a given kinetic model
N_θ	number of non-measurable parameters in a given model
R	ideal gas constant
t	time
T	temperature
U	vector space of model inputs
v	flow velocity along the axial coordinate of microchannel
$v_{\theta,ij}$	ij -th element of the covariance matrix \mathbf{V}_θ
Y	dataset available for model identification
z	axial coordinate of microchannel
$z_{\alpha/2}$	two-tailed score of standard normal distribution with significance α

Matrices and vectors

$\mathbf{1}_\theta$	column array whose entries are all equal to 1 [$N_\theta \times 1$]
---------------------	---

\mathbf{f}	column array of functions $[N_f \times 1]$
\mathbf{G}	linear transformation of parameter space $\Omega \rightarrow \Theta$ $[N_\theta \times N_\theta]$
\mathbf{G}_P	primary transformation of parameter space $\Omega \rightarrow \Theta$ $[N_\theta \times N_\theta]$
\mathbf{G}_S	secondary transformation of parameter space $\Omega \rightarrow \Theta$ $[N_\theta \times N_\theta]$
\mathbf{h}	column array of functions $[N_y \times 1]$
\mathbf{H}	observed Fisher information matrix $[N_\theta \times N_\theta]$
\mathbf{I}_θ	identity matrix $[N_\theta \times N_\theta]$
\mathbf{R}	matrix of rotation of parameter space $[N_\theta \times N_\theta]$
\mathbf{u}	column array of independent control variables (model inputs) $[N_u \times 1]$
\mathbf{U}	right normalised eigenbasis of \mathbf{H} $[N_\theta \times N_\theta]$
\mathbf{V}_θ	covariance of parameter estimates in Θ $[N_\theta \times N_\theta]$
\mathbf{V}_ω	covariance of parameter estimates in Ω $[N_\theta \times N_\theta]$
$\hat{\mathbf{V}}_\omega$	predicted covariance of parameter estimates in Ω $[N_\theta \times N_\theta]$
\mathbf{x}	column array of state variables $[N_x \times 1]$
\mathbf{y}	sample - column array of measured output variables $[N_y \times 1]$
\mathbf{y}_i	i -th sample in dataset Y $[N_y \times 1]$
$\hat{\mathbf{y}}$	column array of predicted output variables $[N_y \times 1]$
$\hat{\mathbf{y}}_i$	column array of predicted output variables for sample \mathbf{y}_i $[N_y \times 1]$
$\boldsymbol{\theta}$	column vector of parameters in parameter space Θ $[N_\theta \times 1]$
$\boldsymbol{\theta}^*$	column vector of target parameters in parameter space Θ $[N_\theta \times 1]$
$\hat{\boldsymbol{\theta}}$	maximum likelihood estimate for $\boldsymbol{\theta} \in \Theta$ $[N_\theta \times 1]$
$\boldsymbol{\Lambda}$	diagonal matrix whose ii -th element is λ_i $[N_\theta \times N_\theta]$
$\boldsymbol{\Sigma}$	covariance of measurement error for sample \mathbf{y} $[N_y \times N_y]$
$\boldsymbol{\varphi}$	experimental design vector
$\boldsymbol{\varphi}^*$	D-optimal experimental design vector
$\boldsymbol{\omega}$	column vector of parameters in parameter space Ω $[N_\theta \times 1]$
$\hat{\boldsymbol{\omega}}_P$	column vector of parameter estimates computed with $\mathbf{G} = \mathbf{G}_P$ $[N_\theta \times 1]$
$\hat{\boldsymbol{\omega}}_S$	column vector of parameter estimates computed with $\mathbf{G} = \mathbf{G}_S$ $[N_\theta \times 1]$

Greek symbols

α	statistical significance
θ_i	i -th model parameter
$\hat{\theta}_i$	estimate for the i -th model parameter
Θ	original vector space of model parameters
κ	condition number
λ_i	i -th eigenvalue of \mathbf{H}
ν_i	stoichiometric coefficient of the i -th species
Φ	log-likelihood function
χ_θ^2	χ^2 statistic of target parameters
χ_{ref}^2	95% value computed from a χ^2 distribution
χ_{sample}^2	sum of normalised squared residuals
Ω	transformed vector space of model parameters
∇	gradient operator in parameter space

Acronyms

ED	Experimental Design
HPLC	High-Performance Liquid Chromatograph
MBDoe	Model-Based Design of Experiments
RG	Regularisation
RP	Reparametrisation

References

- Agarwal, A. K., Brisk, M. L., 1985a. Sequential experimental design for precise parameter estimation. 1. Use of reparameterization. Industrial & Engineering Chemistry Process Design and Development 24 (1), 203–207.
- Agarwal, A. K., Brisk, M. L., 1985b. Sequential experimental design for precise parameter estimation. 2. Design criteria. Industrial & Engineering Chemistry Process Design and Development 24 (1), 207–210.
- Alberton, A. L., Schwaab, M., Schmal, M., Pinto, J. C., 2009. Experimental errors in kinetic tests and its influence on the precision of estimated parameters. Part I—Analysis of first-order reactions. Chemical Engineering Journal 155 (3), 816–823.
- Asprey, S. P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Computers & Chemical Engineering 24 (2), 1261–1267.
- Asprey, S. P., Naka, Y., 1999. Mathematical Problems in Fitting Kinetic Models - Some New Perspectives. Journal of Chemical Engineering of Japan 32 (3), 328–337.
- Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press.
- Bardow, A., 2008. Optimal experimental design of ill-posed problems: The METER approach. Computers & Chemical Engineering 32 (1), 115–124.
- Barz, T., Cárdenas, D. C. L., Arellano-Garcia, H., Wozny, G., 2013. Experimental evaluation of an approach to online redesign of experiments for parameter determination. AIChE Journal 59 (6), 1981–1995.
- Barz, T., López C., D. C., Cruz Bournazou, M. N., Körkel, S., Walter, S. F., 2016. Real-time adaptive input design for the determination of competitive adsorption isotherms in liquid chromatography. Computers & Chemical Engineering 94, 104–116.
- Benabbas, L., Asprey, S. P., Macchietto, S., 2005. Curvature-Based Methods for Designing Optimally Informative Experiments in Multiresponse Nonlinear Dynamic Situations. Industrial & Engineering Chemistry Research 44 (18), 7120–7131.
- Berger, R. J., Stitt, E. H., Marin, G. B., Kapteijn, F., Moulijn, J. A., 2001. Eurokin. Chemical Reaction Kinetics in Practice. CATTECH 5 (1), 36–60.

- 710 Bonvin, D., Georgakis, C., Pantelides, C. C., Barolo, M., Grover, M. A., Rodrigues, D.,
Schneider, R., Dochain, D., 2016. Linking Models and Experiments. *Industrial & Engi-
neering Chemistry Research* 55 (25), 6891–6903.
- Bournazou, M. N. C., Barz, T., Nickel, D. B., Cárdenas, D. C. L., Glauche, F., Knepper, A.,
715 Neubauer, P., 2016. Online optimal experimental re-design in robotic parallel fed-batch
cultivation facilities. *Biotechnology and Bioengineering* 114 (3), 610–619.
- Buzzi-Ferraris, G., Manenti, F., 2009. Kinetic models analysis. *Chemical Engineering Science*
64 (5), 1061–1074.
- Chakrabarty, A., Buzzard, G. T., Rundell, A. E., 2013. Model-based design of experiments
720 for cellular processes. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*
5 (2), 181–203.
- Chis, O. T., Banga, J. R., Balsa-Canto, E., 2014. Sloppy models can be identifiable.
arXiv:1403.1417 [q-bio].
- Dovi, V. G., Reverberi, A. P., Acevedo-Duarte, L., 1994. New procedure for optimal design
725 of sequential experiments in kinetic models. *Industrial & Engineering Chemistry Research*
33 (1), 62–68.
- Echtermeyer, A., Amar, Y., Zakrzewski, J., Lapkin, A., 2017. Self-optimisation and model-
based design of experiments for developing a C-H activation flow process. *Beilstein Journal
of Organic Chemistry* 13, 150–163.
- Elliott, C., Vijayakumar, V., Zink, W., Hansen, R., 2007. National Instruments LabVIEW: A
730 Programming Environment for Laboratory Automation and Measurement. *JALA: Journal
of the Association for Laboratory Automation* 12 (1), 17–24.
- Espie, D. M., Macchietto, S., 1988. Nonlinear transformations for parameter estimation.
Industrial & Engineering Chemistry Research 27 (11), 2175–2179.
- Fabry, D. C., Sugiono, E., Rueping, M., 2014. Self-Optimizing Reactor Systems: Algorithms,
735 On-line Analytics, Setups, and Strategies for Accelerating Continuous Flow Process Op-
timization. *Israel Journal of Chemistry* 54 (4), 341–350.
- Fedorov, V. V., 1972. *Theory Of Optimal Experiments*, 1972nd Edition. Academic Press.
- Franceschini, G., Macchietto, S., 2008a. Anti-Correlation Approach to Model-Based Exper-
iment Design: Application to a Biodiesel Production Process. *Industrial & Engineering*
740 *Chemistry Research* 47 (7), 2331–2348.
- Franceschini, G., Macchietto, S., 2008b. Model-based design of experiments for parameter
precision: State of the art. *Chemical Engineering Science* 63 (19), 4846–4872.
- Franceschini, G., Macchietto, S., 2008c. Novel anticorrelation criteria for design of experi-
ments: Algorithm and application. *AIChE Journal* 54 (12), 3221–3238.

- 745 Franceschini, G., Macchietto, S., 2008d. Novel anticorrelation criteria for model-based experiment design: Theory and formulations. *AIChE Journal* 54 (4), 1009–1024.
- Galvanin, F., Barolo, M., Bezzo, F., 2013. On the use of continuous glucose monitoring systems to design optimal clinical tests for the identification of type 1 diabetes models. *Computer Methods and Programs in Biomedicine* 109 (2), 157–170.
- 750 Galvanin, F., Macchietto, S., Bezzo, F., 2007. Model-Based Design of Parallel Experiments. *Industrial & Engineering Chemistry Research* 46 (3).
- Goodell, J. R., McMullen, J. P., Zaborenko, N., Maloney, J. R., Ho, C.-X., Jensen, K. F., Porco, J. A., Beeler, A. B., 2009. Development of an Automated Microfluidic Reaction Platform for Multidimensional Screening: Reaction Discovery Employing Bicyclo[3.2.1]octanoid Scaffolds. *The Journal of Organic Chemistry* 74 (16), 6169–6180.
- 755 Hansen, P. C., 2005. Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM.
- Higham, N. J., 1996. Accuracy and Stability of Numerical Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- 760 Holmes, N., Akien, G. R., Savage, R. J. D., Stanetty, C., Baxendale, I. R., Blacker, A. J., Taylor, B. A., Woodward, R. L., Meadows, R. E., Bourne, R. A., 2016. Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors. *Reaction Chemistry & Engineering* 1 (1), 96–100.
- Hosten, L. H., 1974. A sequential experimental design procedure for precise parameter estimation based upon the shape of the joint confidence region. *Chemical Engineering Science* 29 (11), 2247–2252.
- 765 Huber, P. J., 2004. Robust Statistics. John Wiley & Sons.
- Johansen, T. A., 1997. On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica* 33 (3), 441–446.
- 770 Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python.
- Levenspiel, O., 1998. Chemical Reaction Engineering, 3rd Edition. John Wiley & Sons, New York.
- Lopez C., D. C., Barz, T., Körkel, S., Wozny, G., 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers & Chemical Engineering* 77, 24–42.
- 775 MacKay, D. J. C., 1992. Bayesian methods for adaptive models. phd, California Institute of Technology.

- 780 Maheshwari, V., Rangaiah, G. P., Samavedham, L., 2013. Multiobjective Framework for Model-based Design of Experiments to Improve Parameter Precision and Minimize Parameter Correlation. *Industrial & Engineering Chemistry Research* 52 (24), 8289–8304.
- Malig, T. C., Koenig, J. D. B., Situ, H., Chehal, N. K., Hultin, P. G., Hein, J. E., 2017. Real-time HPLC-MS reaction progress monitoring using an automated analytical platform. *Reaction Chemistry & Engineering* 2 (3), 309–314.
- 785 McMullen, J. P., Jensen, K. F., 2010. An Automated Microfluidic System for Online Optimization in Chemical Synthesis. *Organic Process Research & Development* 14 (5), 1169–1176.
- McMullen, J. P., Jensen, K. F., 2011. Rapid Determination of Reaction Kinetics with an Automated Microfluidic System. *Organic Process Research & Development* 15 (2), 398–
790 407.
- Moore, J. S., Jensen, K. F., 2012. Automated Multitrajectory Method for Reaction Optimization in a Microfluidic System using Online IR Analysis. *Organic Process Research & Development* 16 (8), 1409–1415.
- Oliphant, T. E., 2015. *Guide to NumPy*, 2nd Edition. CreateSpace Independent Publishing
795 Platform, USA.
- Pipus, G., Plazl, I., Koloini, T., 2000. Esterification of benzoic acid in microwave tubular flow reactor. *Chemical Engineering Journal* 76 (3), 239–245.
- Prasad, V., Vlachos, D. G., 2008. Multiscale Model and Informatics-Based Optimal Design of Experiments: Application to the Catalytic Decomposition of Ammonia on Ruthenium.
800 *Industrial & Engineering Chemistry Research* 47 (17), 6555–6567.
- Pritchard, D. J., Bacon, D. W., 1978. Prospects for reducing correlations among parameter estimates in kinetic models. *Chemical Engineering Science* 33 (11), 1539–1543.
- Pukelsheim, F., 2006. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, New York, United States of America.
- 805 Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25 (15), 1923–1929.
- Rimensberger, T., Rippin, D. W. T., 1986. "Sequential experimental design for precise parameter estimation. 1. Use of reparameterization". *Comments. Industrial & Engineering
810 Chemistry Process Design and Development* 25 (4), 1042–1044.
- Rossi, D., Gargiulo, L., Valitov, G., Gavriilidis, A., Mazzei, L., 2017. Experimental characterization of axial dispersion in coiled flow inverters. *Chemical Engineering Research and Design* 120, 159–170.

- 815 Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science* 15 (4), 377–395.
- Schwaab, M., Lemos, L. P., Pinto, J. C., 2008. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 2: Problems involving multiple reparameterizations. *Chemical Engineering Science* 63 (11), 2895–2906.
- 820 Schwaab, M., Pinto, J. C., 2007. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant. *Chemical Engineering Science* 62 (10), 2750–2764.
- Silvey, S. D., 1975. *Statistical Inference*. CRC Press.
- 825 Stamati, I., Logist, F., Akkermans, S., Noriega Fernández, E., Van Impe, J., 2016. On the effect of sampling rate and experimental noise in the discrimination between microbial growth models in the suboptimal temperature range. *Computers & Chemical Engineering* 85, 84–93.
- Transtrum, M. K., Machta, B., Brown, K., Daniels, B. C., Myers, C. R., Sethna, J. P., 2015. Sloppiness and Emergent Theories in Physics, Biology, and Beyond. arXiv:1501.07668.
- 830 Transtrum, M. K., Machta, B. B., Sethna, J. P., 2010. Why are nonlinear fits so challenging? *Physical Review Letters* 104 (6).
- Versyck, K. J., Van Impe, J. F., 1997. On the unicity of optimal experimental design solutions for parameter estimation of microbial kinetics. In: 1997 European Control Conference (ECC). pp. 3509–3514.
- 835 Walpole, R. E., Myers, R. H., Myers, S. L., Ye, K. E., 2011. *Probability and Statistics for Engineers and Scientists*, 9th Edition. Pearson.
- Walsh, B., Hyde, J. R., Licence, P., Poliakov, M., 2005. The automation of continuous reactions in supercritical CO₂: the acid-catalysed etherification of short chain alcohols. *Green Chemistry* 7 (6), 456–463.
- 840 White, A., Tolman, M., Thames, H. D., Withers, H. R., Mason, K. A., Transtrum, M. K., 2016. The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems. *PLOS Computational Biology* 12 (12).
- Wilson, A. D., Schultz, J. A., Murphey, T. D., 2015. Trajectory Optimization for Well-Conditioned Parameter Estimation. *IEEE Transactions on Automation Science and Engineering* 12 (1), 28–36.
- 845 Zhelezov, O. I., 2017. N-dimensional Rotation Matrix Generation Algorithm. *American Journal of Computational and Applied Mathematics* 7 (2), 51–57.

Appendix A

Additional simulated cases

A total number of 20 experimental campaigns were simulated to further validate the results presented in the manuscript. This was done primarily to demonstrate that the performance achieved by the algorithm both in the RP and in the non-RP campaigns is insensitive to the choice of the dataset (i.e. it is insensitive to the choice of the random seed used to generate the experimental data in-silico).

The results obtained in the simulated campaigns are reported in Table A.1. Campaigns 1-10 were performed applying the online reparametrisation method (RP campaigns), while campaigns 11-20 were performed without online reparametrisation (non-RP campaigns). As one can see from Table A.1, the algorithm with online RP option active retrieved the target parameter value in all the campaigns, i.e. the final p -value of the target parameters is above 1.00% in campaigns 1-10. The condition number of the log-likelihood functions at the end of experimental campaigns 1-10 is 1.0, demonstrating that the application of the online RP led to the elimination of the model sloppiness. In the campaigns where the online RP is inactive, i.e. campaigns 11-20, the final p -value is 0.00%, demonstrating the failure of the algorithm in retrieving the target value of the parameters. The failure is associated to the high condition number of the log-likelihood function, which is around $10^3 - 10^4$ in campaigns 11-20.

Table A.1: Results obtained in 20 simulated experimental campaigns: experimental campaigns 1-10 were performed keeping the online reparametrisation option *active*; campaigns 11-20 were performed keeping the option for online reparametrisation *inactive*. The p -value of the target parameters $\theta^* = [15.27, 7.6]$ given the final parameter statistics is reported together with the condition number of the log-likelihood function at the end of the experimental campaigns.

Campaign number	Online reparametrisation	Final p -value of target parameters θ^*	Final condition number κ
1	Active	64.74%	$1.0 \cdot 10^0$
2	Active	98.91%	$1.0 \cdot 10^0$
3	Active	91.98%	$1.0 \cdot 10^0$
4	Active	20.59%	$1.0 \cdot 10^0$
5	Active	30.52%	$1.0 \cdot 10^0$
6	Active	67.93%	$1.0 \cdot 10^0$
7	Active	16.61%	$1.0 \cdot 10^0$
8	Active	92.17%	$1.0 \cdot 10^0$
9	Active	23.19%	$1.0 \cdot 10^0$
10	Active	71.59%	$1.0 \cdot 10^0$
11	Inactive	0.00%	$9.6 \cdot 10^3$
12	Inactive	0.00%	$9.4 \cdot 10^3$
13	Inactive	0.00%	$9.5 \cdot 10^3$
14	Inactive	0.00%	$9.3 \cdot 10^3$
15	Inactive	0.00%	$1.1 \cdot 10^4$
16	Inactive	0.00%	$9.5 \cdot 10^3$
17	Inactive	0.00%	$1.0 \cdot 10^4$
18	Inactive	0.00%	$9.2 \cdot 10^3$
19	Inactive	0.00%	$8.7 \cdot 10^3$
20	Inactive	0.00%	$9.3 \cdot 10^3$

Appendix B

Real case: additional information

Additional details are presented in this appendix regarding the non-RP campaign and the RP campaign performed on the experimental automated system. Information related to the campaign performed keeping the option for online model reparametrisation *inactive*, i.e. the non-RP campaign, is reported in Table B.1. Information on the campaign conducted keeping the option for online reparametrisation *active*, i.e. the RP campaign, is given in Table B.2. In Table B.1 and Table B.2 the following information is presented: 1) experimental conditions adopted to collect the samples, i.e. inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$, flowrate F and temperature T ; 2) sampled concentration of ethyl benzoate at the outlet $C_{\text{EB}}^{\text{OUT}}$; 3) parameter estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ returned by the model identification algorithm; 4) the pre-exponential factor and activation energy computed from the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ as $A = e^{\hat{\theta}_1}$ and $E_a = 10^4 \cdot \hat{\theta}_2$; 5) the sum of squared residuals χ_{sample}^2 and the reference value χ_{ref}^2 computed from a χ^2 distribution with degree of freedom equal to the number of samples minus the number of parameters and 95% of significance.

A sum of squared residuals χ_{sample}^2 larger than the reference value χ_{ref}^2 is interpreted as an index of inappropriate modelling assumptions (Silvey, 1975). As one can see from Table B.1, the χ_{sample}^2 after the collection of 9 samples in the non-RP campaign is 5.92. From Table B.2, it can be appreciated that the χ_{sample}^2 after the collection of 9 samples in the RP campaign is 1.83. Both in the non-RP and in the RP campaign the χ_{sample}^2 is smaller than the $\chi_{\text{ref}}^2 = 17.88$, thus demonstrating that the modelling assumptions (see Section 3.2) are not falsified by the experimental evidence.

As one can see from Table B.1, the experimental conditions designed by the algorithm for samples 5, 7 and 9 in the non-RP case were similar, i.e. inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, flowrate around $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and temperature $T = 413.0 \text{ K}$, i.e. the upper limit for the temperature. Samples 4, 6 and 8 were instead designed by the algorithm at conditions $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, flowrate $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and temperature in the range $T = 383.0 - 390.0 \text{ K}$. The designed samples in the non-RP case suggest the presence of two optimally informative sets of experimental conditions at maximum temperature $T = 413.0 \text{ K}$ and at temperature around $T = 385.0 \text{ K}$, given that the inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$ is set at the maximum and that flowrate F is set at the minimum.

An analogous situation can be observed in the RP case. As one can see from Table B.2, samples 4, 7 and 9 in the RP case were designed at conditions $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and $T = 413.0 \text{ K}$. Samples 5, 6 and 8 were instead designed at conditions $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and temperature around $T = 391.0 \text{ K}$.

Table B.1: Real case: experimental campaign with online RP option *inactive*. Experimental conditions, sampled concentrations, estimated kinetic parameters $\hat{\Theta}$ (and related Arrhenius constants) and information regarding the goodness-of-fit are reported for the 9 samples collected in the campaign.

Real Case - Online RP Inactive										
Sample	Experimental conditions φ			Sample	Estimates $\hat{\Theta}$		Arrhenius constants ¹		Goodness-of-fit ²	
number	C_{BA}^{IN} [mol L ⁻¹]	F [μ L min ⁻¹]	T [K]	C_{EB}^{OUT} [mol L ⁻¹]	$\hat{\theta}_1$	$\hat{\theta}_2$	A [s ⁻¹]	E_a [J mol ⁻¹ K ⁻¹]	χ_{sample}^2	χ_{ref}^2
1	1.50	20.00	413.0	0.370	-	-	-	-	-	-
2	1.00	10.00	393.0	0.161	-	-	-	-	-	-
3	1.25	15.00	403.0	0.240	16.16	7.94	$1.04 \cdot 10^7$	$7.94 \cdot 10^4$	$4.35 \cdot 10^{-4}$	3.84
4	1.55	7.50	383.0	0.175	16.44	8.03	$1.39 \cdot 10^7$	$8.03 \cdot 10^4$	$2.65 \cdot 10^{-2}$	5.99
5	1.55	7.58	413.0	0.848	17.15	8.26	$2.81 \cdot 10^7$	$8.26 \cdot 10^4$	1.04	7.81
6	1.55	7.50	390.2	0.284	16.80	8.14	$1.98 \cdot 10^7$	$8.14 \cdot 10^4$	1.31	9.49
7	1.55	7.50	413.0	0.876	17.23	8.28	$3.03 \cdot 10^7$	$8.28 \cdot 10^4$	3.56	11.07
8	1.55	7.50	388.5	0.254	17.15	8.26	$2.82 \cdot 10^7$	$8.26 \cdot 10^4$	3.59	12.59
9	1.55	7.50	413.0	0.887	17.42	8.34	$3.69 \cdot 10^7$	$8.34 \cdot 10^4$	5.92	14.07

¹ Pre-exponential factor and activation energy are computed from θ_1 and θ_2 as $A = e^{\theta_1}$ and $E_a = 10^4 \cdot \theta_2$

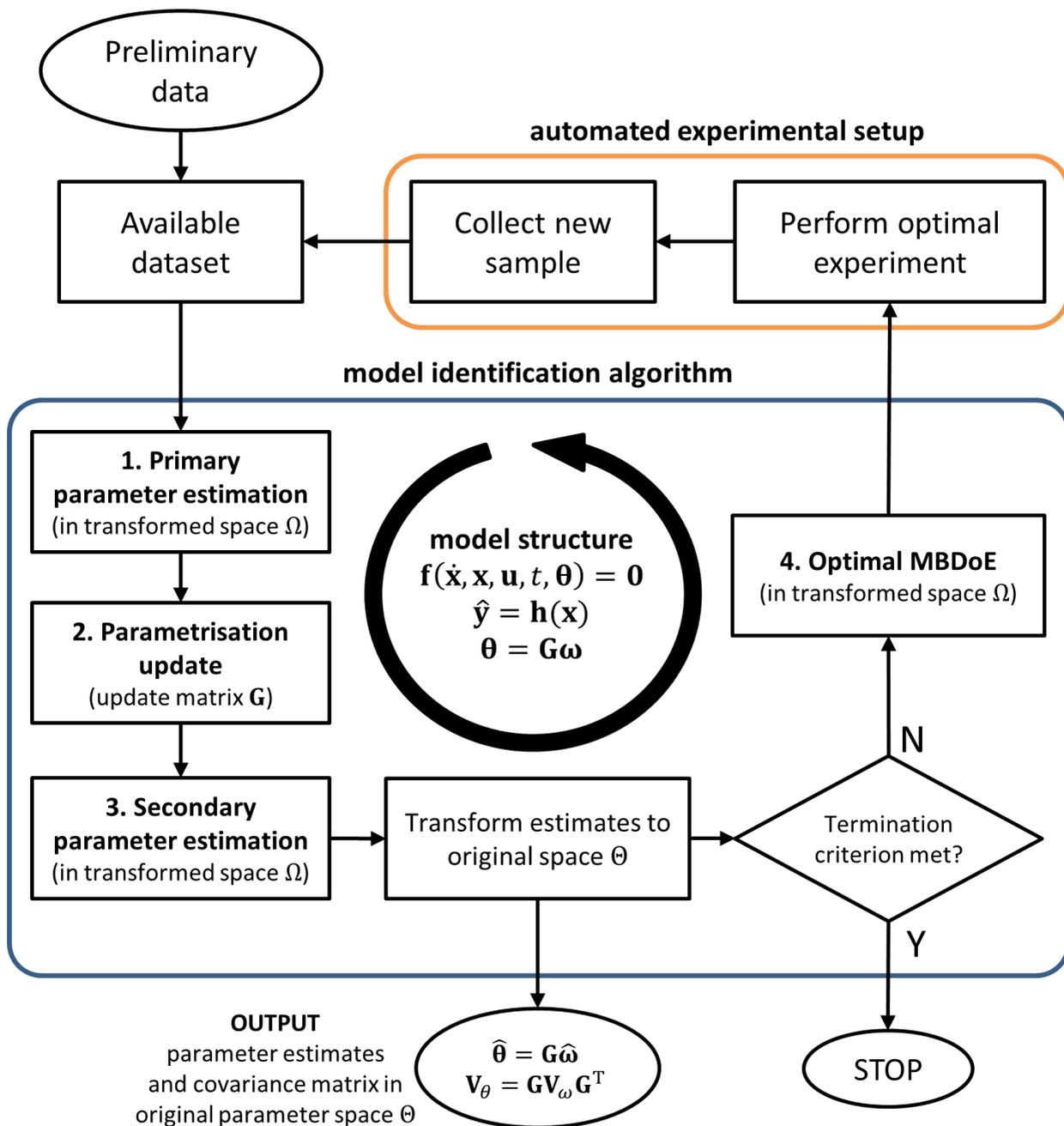
² A χ_{sample}^2 larger than χ_{ref}^2 is an index of inappropriate modelling assumptions

Table B.2: Real case: experimental campaign with online RP option *active*. Experimental conditions, sampled concentrations, estimated kinetic parameters $\hat{\Theta}$ (and related Arrhenius constants) and information regarding the goodness-of-fit are reported for the 9 samples collected in the campaign.

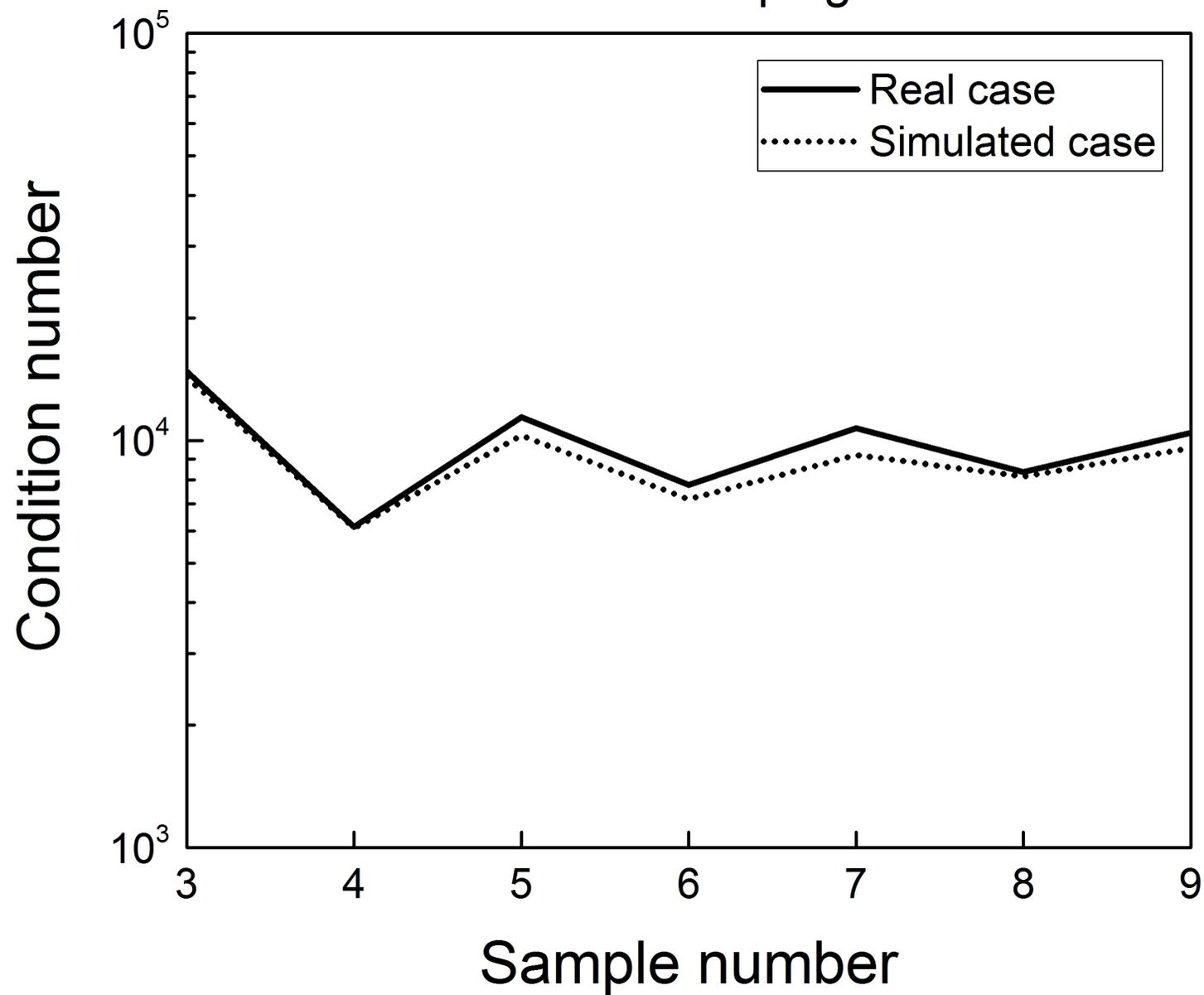
Real Case - Online RP Active										
Sample	Experimental conditions φ			Sample	Estimates $\hat{\Theta}$		Arrhenius constants ¹		Goodness-of-fit ²	
number	C_{BA}^{IN} [mol L ⁻¹]	F [μ L min ⁻¹]	T [K]	C_{EB}^{OUT} [mol L ⁻¹]	$\hat{\theta}_1$	$\hat{\theta}_2$	A [s ⁻¹]	E_a [J mol ⁻¹ K ⁻¹]	χ_{sample}^2	χ_{ref}^2
1	1.50	20.00	413.0	0.409	-	-	-	-	-	-
2	1.00	10.00	393.0	0.172	-	-	-	-	-	-
3	1.25	15.00	403.0	0.252	17.54	8.37	$4.13 \cdot 10^7$	$8.37 \cdot 10^4$	0.21	3.84
4	1.55	7.50	413.0	0.900	18.12	8.56	$7.39 \cdot 10^7$	$8.56 \cdot 10^4$	0.52	5.99
5	1.55	7.50	392.3	0.346	16.86	8.13	$2.10 \cdot 10^7$	$8.13 \cdot 10^4$	1.27	7.81
6	1.55	7.50	390.6	0.307	16.90	8.15	$2.18 \cdot 10^7$	$8.15 \cdot 10^4$	1.27	9.49
7	1.55	7.50	413.0	0.895	16.91	8.15	$2.20 \cdot 10^7$	$8.15 \cdot 10^4$	1.27	11.07
8	1.55	7.50	391.2	0.323	16.83	8.12	$2.04 \cdot 10^7$	$8.12 \cdot 10^4$	1.31	12.59
9	1.55	7.50	413.0	0.908	16.98	8.17	$2.36 \cdot 10^7$	$8.17 \cdot 10^4$	1.83	14.07

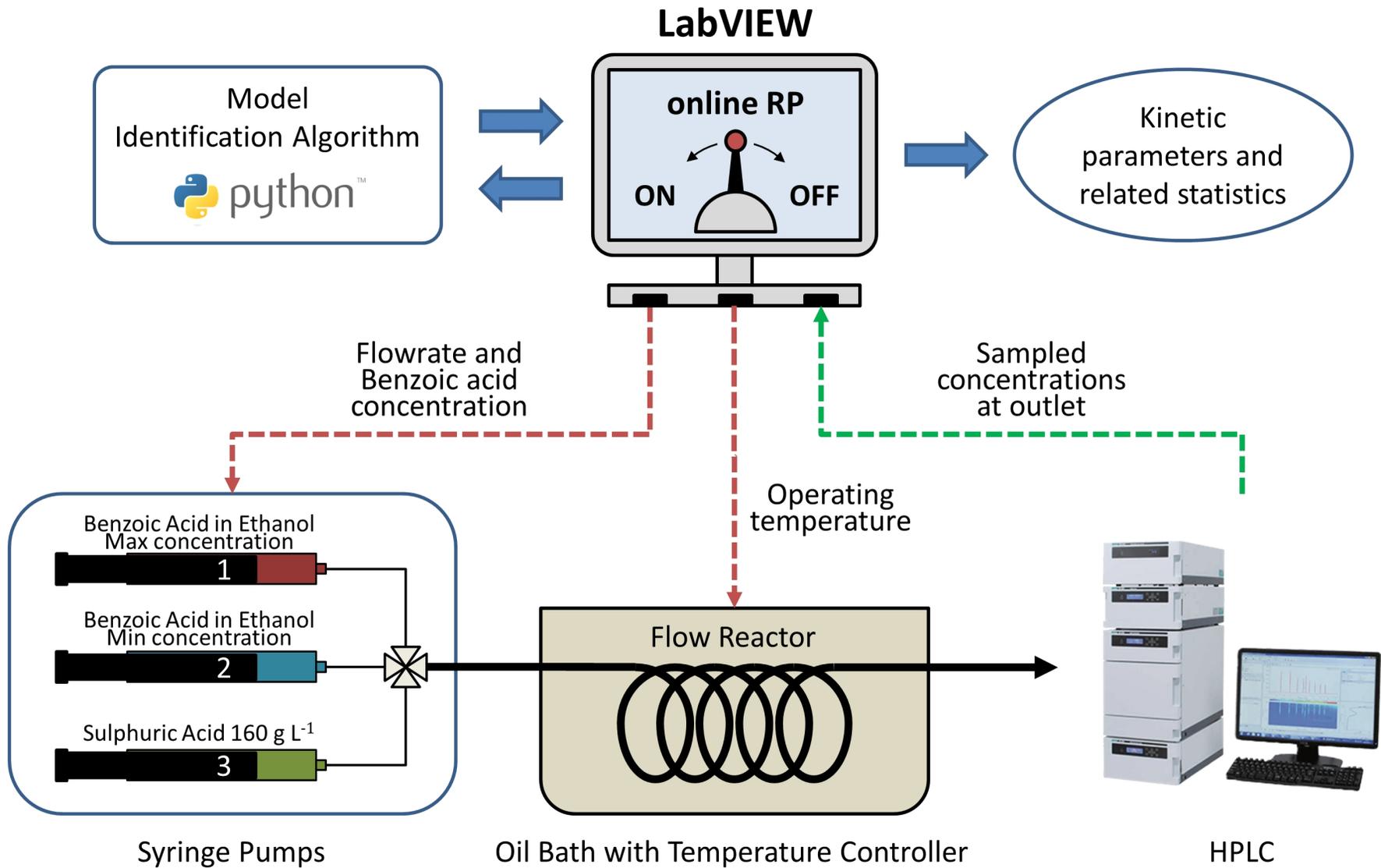
¹ Pre-exponential factor and activation energy are computed from θ_1 and θ_2 as $A = e^{\theta_1}$ and $E_a = 10^4 \cdot \theta_2$

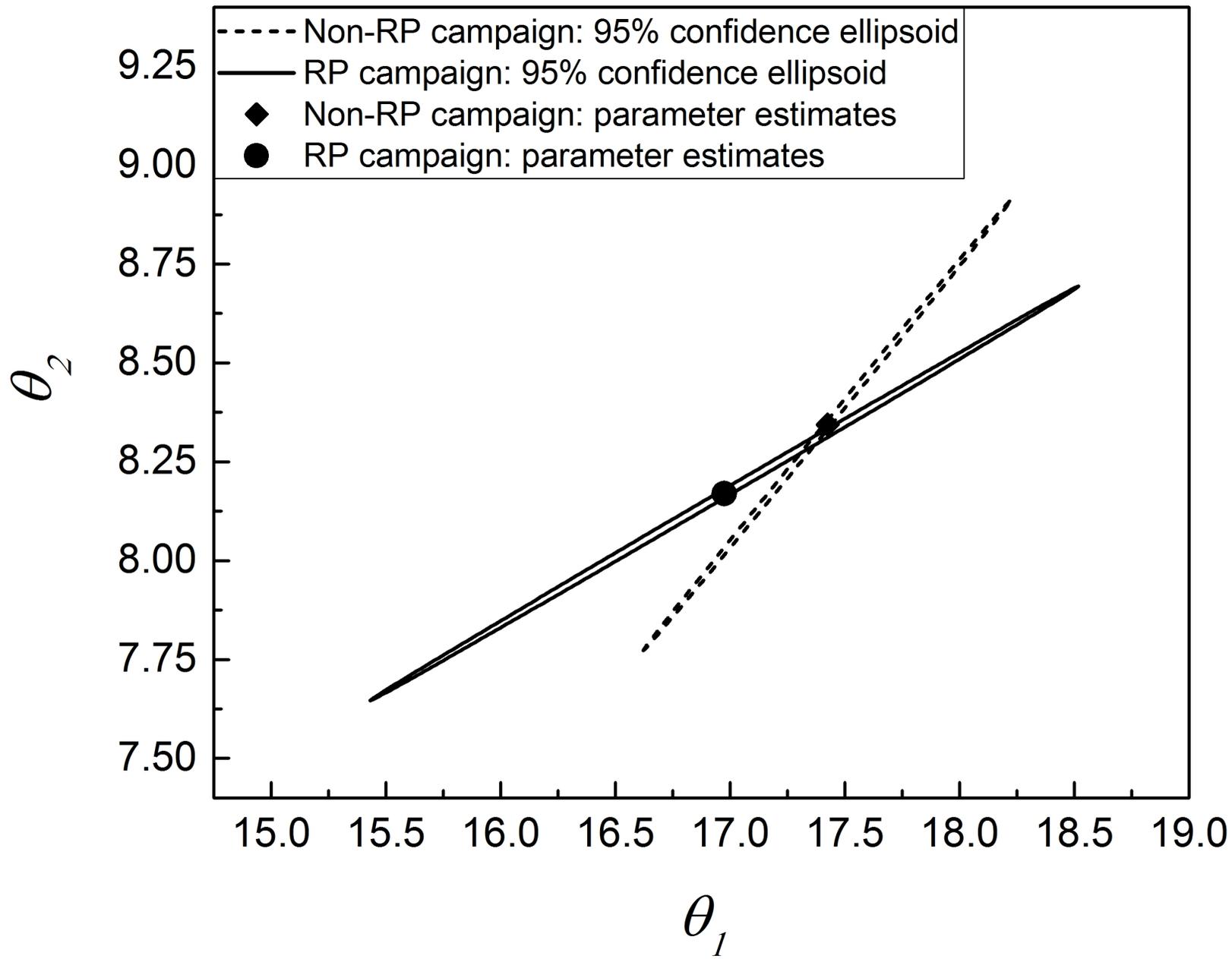
² A χ_{sample}^2 larger than χ_{ref}^2 is an index of inappropriate modelling assumptions

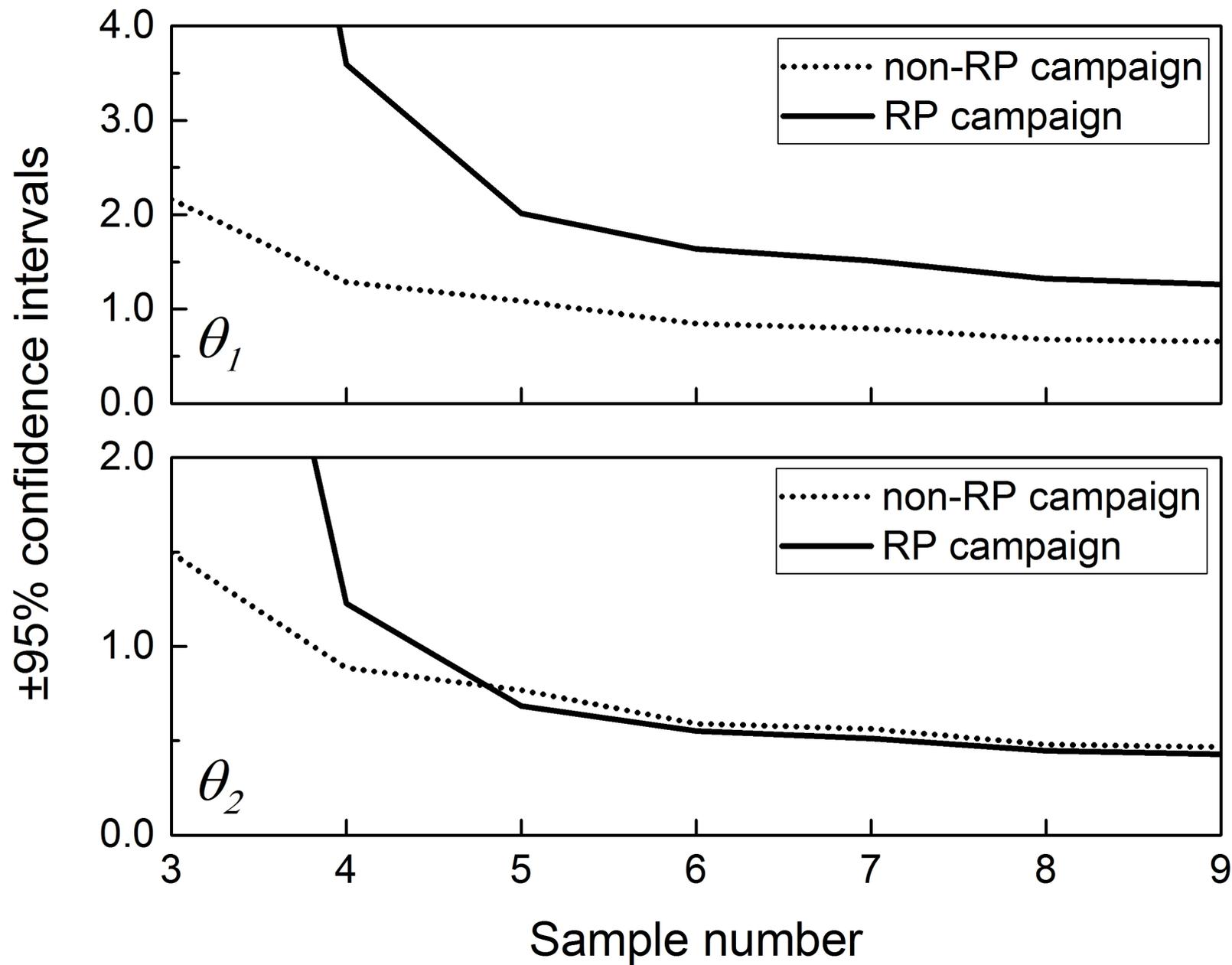


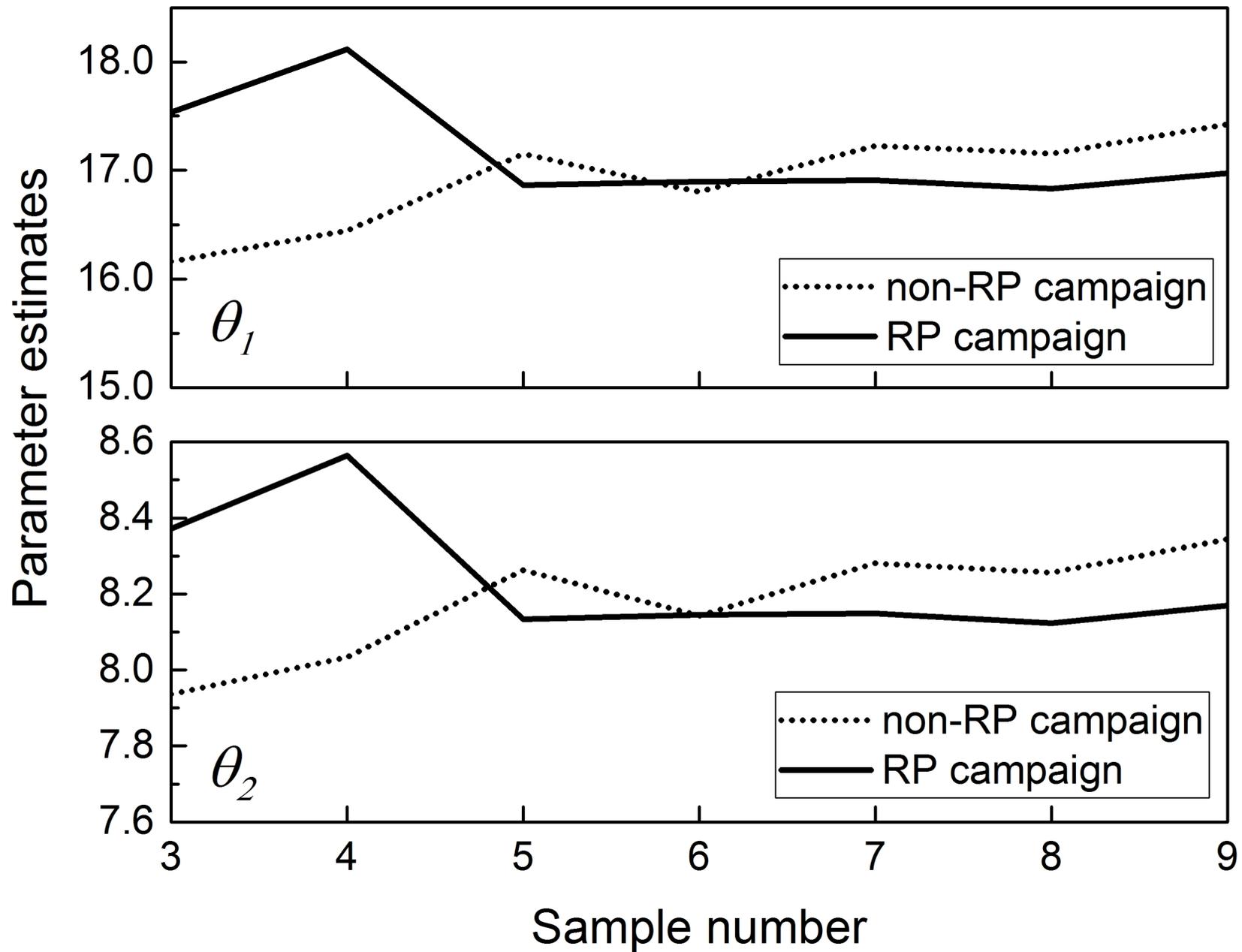
Non-RP campaigns



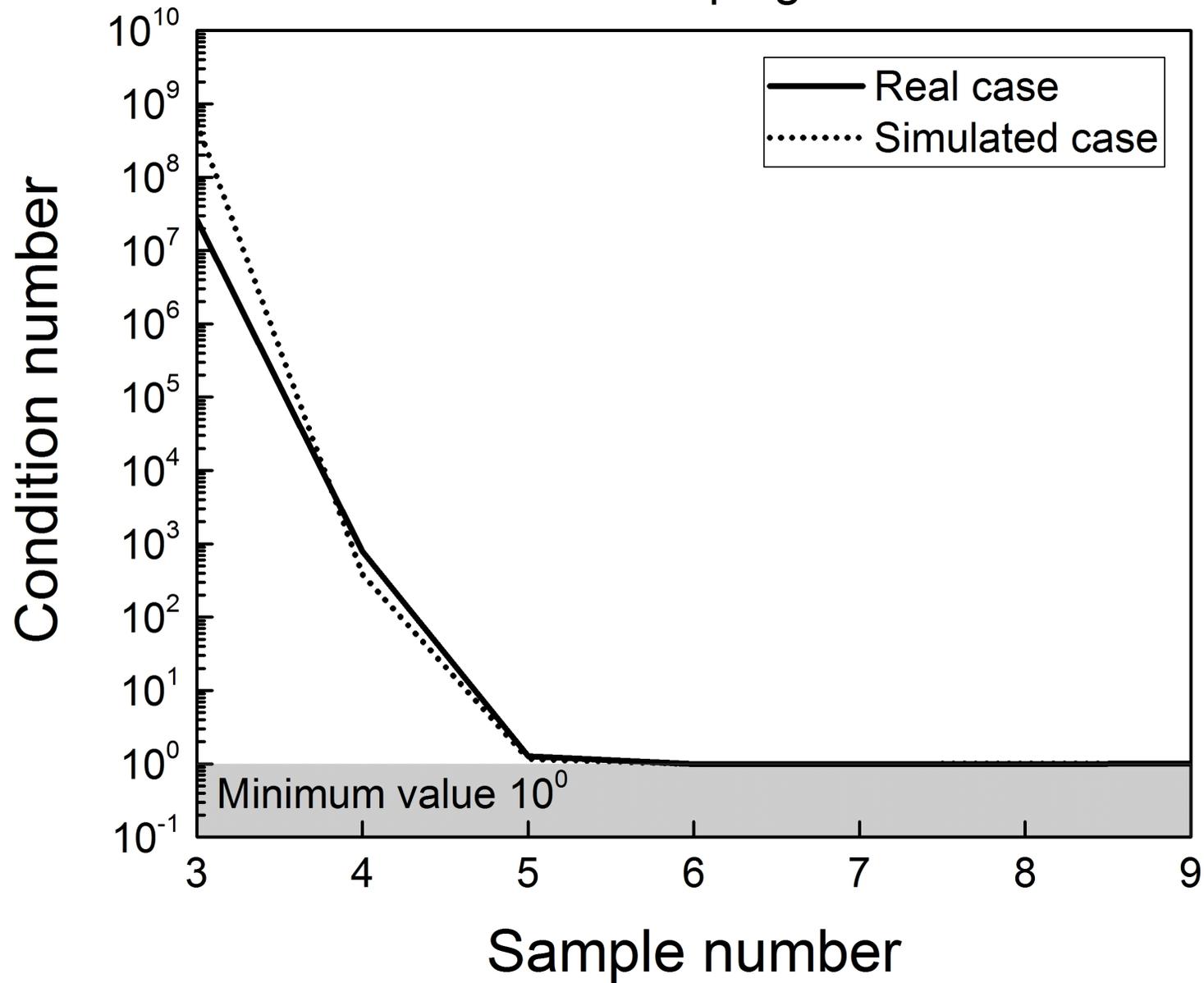


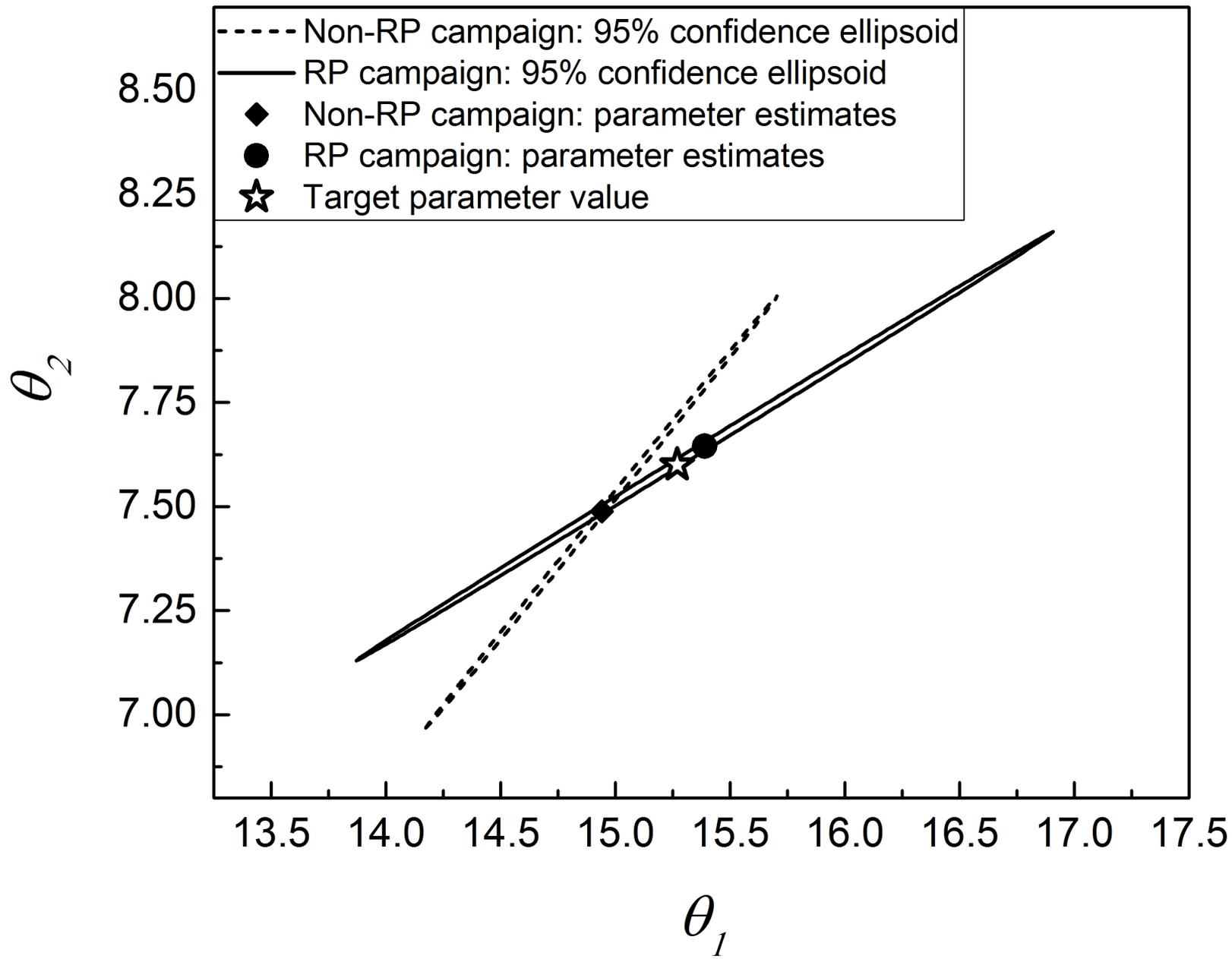


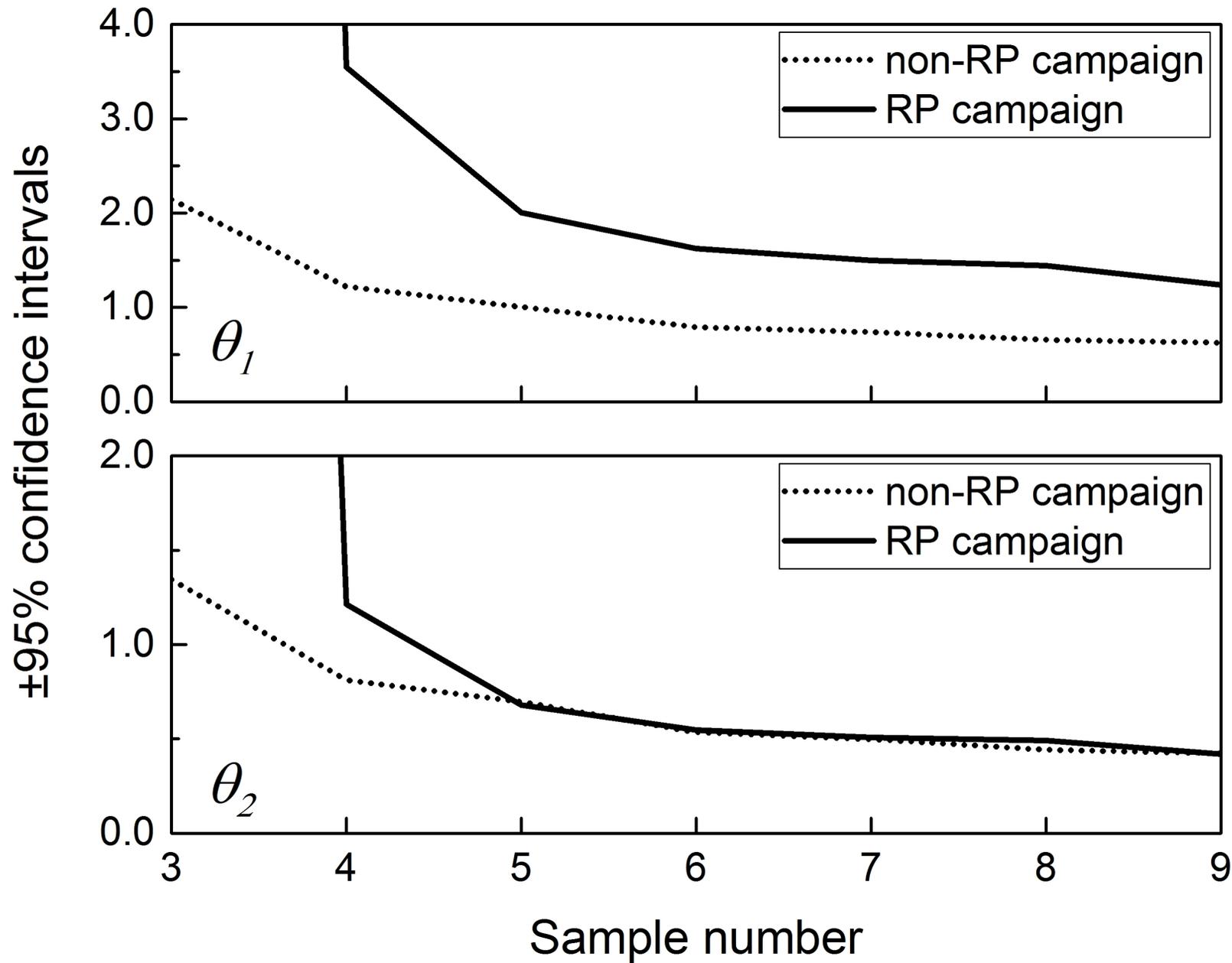


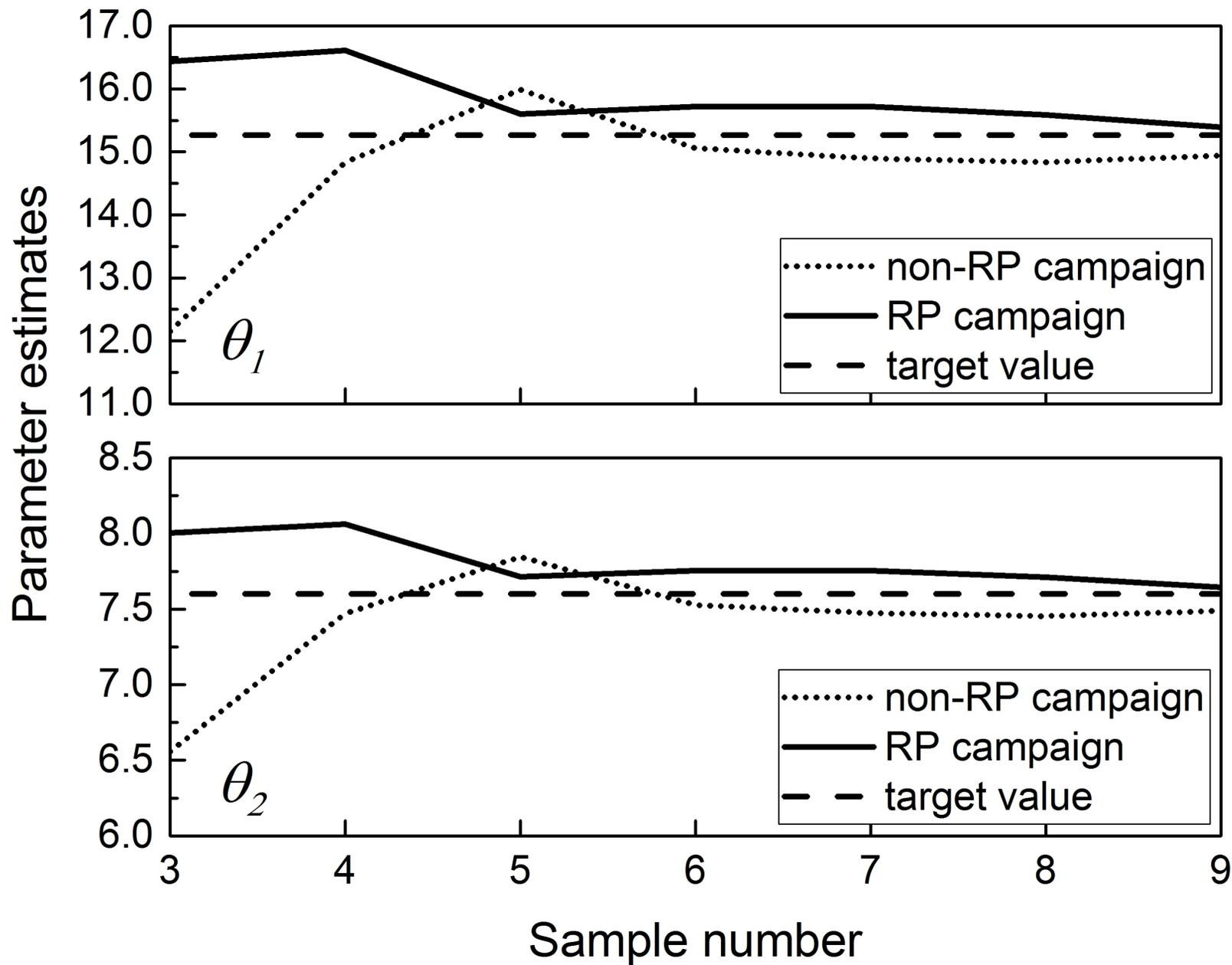


RP campaigns









An online reparametrisation approach for robust parameter estimation in automated model identification platforms

Marco Quaglio^a, Conor Waldron^a, Arun Pankajakshan^a, Enhong Cao^a,
5 Asterios Gavriilidis^a, Eric S. Fraga^a, and Federico Galvanin^{a,*}

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

*corresponding author e-mail: f.galvanin@ucl.ac.uk

Abstract

10 Automated model identification platforms were recently employed to identify parametric models online in the course of unmanned experimental campaigns. The algorithms controlling these platforms include two computational elements: *i*) a tool for parameter estimation; *ii*) a tool for model-based experimental design. Both tools require the solution of complex optimisation problems and their effective outcome relies
15 on their respective objective functions being well-conditioned. Ill-conditioned objective functions may arise when the model is characterised by a weak parametrisation, i.e. the model parameters are practically non-identifiable and/or extremely correlated. In this work, a robust reparametrisation technique is proposed and tested both in-silico and in an automated model identification platform. The benefit of reparametrisation
20 is demonstrated on a case study for the identification of a kinetic model of catalytic esterification of benzoic acid with ethanol in a flow microreactor.

keywords: online, identification, information, parametrization, design, experiment

1 Introduction

25 The kinetic modelling of chemical phenomena through the identification of an appropriate set of model equations is an important step in many research domains related to chemical engineering. Reliable kinetic models (i.e. models that accurately quantify the kinetic behaviour of the physical system) are regarded as key tools for supporting the design and intensification of chemical processes, performing non-empirical process optimisation and understanding which degrees of freedom in the physical system ultimately determine its ob-
30 servable behaviour (Berger et al., 2001). The identification of reliable models requires *i*) the determination of an opportune structure for the model equations and *ii*) the precise estimation of the model parameters. Both aspects typically require extensive amounts of time and resources for performing kinetic experiments. In the last decades, much effort

has been devoted by the scientific community to reducing the experimental burden required
35 to identify and validate kinetic models (Bonvin et al., 2016). Important steps towards the
reduction in the cost of kinetic studies are 1) the coupling of automated, small-scale flow
reactor technologies with online analysis equipment for the quick collection of experimental
data (Goodell et al., 2009) and 2) the employment of model-based design of experiments
(MBDoE) techniques for planning optimal experiments, minimise the cost, time and amount
40 of resources required for the experimentation (Asprey and Macchietto, 2000; Prasad and
Vlachos, 2008; Chakrabarty et al., 2013; Galvanin et al., 2013; Stamati et al., 2016).

Automated flow reactors have been employed in a wide variety of situations from process
monitoring (Malig et al., 2017) to screening of operating conditions (Walsh et al., 2005).
Automated flow reactors were also successfully coupled to algorithms for online sequential
45 design of experiments (McMullen and Jensen, 2010; Moore and Jensen, 2012; Fabry et al.,
2014; Holmes et al., 2016). After every experiment is terminated and new data are collected
by these platforms, algorithms construct black-box representations of the physical system
(e.g. response surfaces) for designing the following experiment with the aim of optimising
the reaction performance (e.g. the conversion or the yield). These self-optimising reactors
50 demonstrated the possibility for an automated platform of conducting experimental cam-
paigns with minimum human intervention. However, these platforms do not exploit the
collected data for the online development and identification of *physics-based* models. A ma-
jor consequence of this is that optimised reaction conditions identified through a black-box
approach in the lab-scale equipment are not necessarily transferable to the design, optimi-
55 sation and control of equipment at the industrial scale.

Only few works are available in the literature in which algorithms for online kinetic
model identification were coupled to automated reactor systems (McMullen and Jensen, 2011;
Bournazou et al., 2016; Echtermeyer et al., 2017). In these works, algorithms for parameter
estimation and optimal MBDoE were employed online to drive experimental campaigns
60 with the aim of selecting the best model among a set of given model structures (i.e. model
discrimination) (McMullen and Jensen, 2011) and/or improving the statistical quality of the
parameter estimates for a given model structure (McMullen and Jensen, 2011; Bournazou
et al., 2016; Echtermeyer et al., 2017). Automated model identification systems have the
potential of dramatically speeding up the modelling of kinetic phenomena and, consequently,
65 the discovery and the study of new chemical processes. However, the diffusion of these
promising systems in research laboratories is hampered by the high chance of numerical
failures whenever model identification algorithms are invoked.

The mathematical structure of kinetic models is frequently affected by problems of prac-
tical *identifiability*, i.e., the fitting quality of the data may be insensitive to a change in some
70 parameters and/or model parameters may be affected by extreme correlation. Whenever
the kinetic model exhibits this type of behaviour it is called *sloppy* (alternatively called ill-
conditioned model or poorly constrained model) (Chis et al., 2014) and its identification may
pose significant challenges even to state-of-the-art model identification algorithms (Asprey
and Naka, 1999; Transtrum et al., 2010, 2015; White et al., 2016). Parameter estimation
75 and optimal MBDoE problems are normally recast as optimisation problems and solved
numerically. In the presence of a sloppy parametrisation, the objective functions of the
aforementioned optimisation problems are ill-posed. The optimisation of ill-posed functions
may lead to significant numerical failures in the course of an unmanned experimental cam-

80 paign with the concomitant waste of experimental resources. Improving the robustness of automated model identification platforms towards model sloppiness is key to further promote their employment in the discovery and study of kinetic phenomena.

The main contribution of this manuscript is a computational strategy for online model reparametrisation (RP), i.e. a tool for transforming automatically the model parameter space in the course of the online model identification process. The presented tool is introduced to 85 enhance the robustness of unmanned platforms for model identification towards *numerical failures* derived by model sloppiness. Throughout this work, it is assumed that an opportune set of kinetic model equations is provided by the user to the model identification algorithm from the beginning of the unmanned experimental campaign. The benefit of the online RP is demonstrated experimentally on a case study where the objective is the identification of 90 a kinetic model of catalytic esterification of benzoic acid in a microreactor system.

2 Methods

2.1 Problem statement

An automated platform is available for performing experiments on a physical system of interest. An array \mathbf{y} of N_y physical quantities can be sampled by an online measurement 95 system. The kinetic behaviour of the physical system is described by a system of differential and algebraic equations as follows:

$$\begin{aligned} \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) &= \mathbf{0} \\ \hat{\mathbf{y}} &= \mathbf{h}(\mathbf{x}) \end{aligned} \tag{1}$$

In (1), \mathbf{f} is a $N_f \times 1$ array of model functions, \mathbf{x} is a $N_x \times 1$ array of state variables, $\dot{\mathbf{x}}$ is a $N_x \times 1$ array of time derivatives for the state variables¹, $\mathbf{u} \in U$ is a $N_u \times 1$ array of manipulable system inputs, t is time and $\boldsymbol{\theta} \in \Theta$ is a $N_\theta \times 1$ column array of model 100 parameters $\theta_1, \dots, \theta_{N_\theta}$. In (1), $\hat{\mathbf{y}}$ is a $N_y \times 1$ array of model predictions for the N_y measurable system states, expressed as a $N_y \times 1$ array of functions \mathbf{h} . It is assumed that the model (1) satisfies the requirements for structural identifiability, i.e., in principle, values of the model parameters $\boldsymbol{\theta}$ can be uniquely determined from the fitting of experimental data (Raue et al., 2009). The objective of the scientist is estimating the set of model parameters as precisely as 105 possible through an unmanned experimental campaign conducted on the automated platform given that the experimental budget allows for the collection of N_{MAX} samples of \mathbf{y} .

Whenever new data become available from the automated reactor system, the model identification algorithm is required to solve sequentially 1) a parameter estimation problem given the available dataset (Bard, 1974) and 2) a model-based design of experiments (MB- 110 DoE) problem to design the following experiment with the aim of minimising the predicted confidence region of parameter estimates (Franceschini and Macchietto, 2008b). The solution of both problems requires the employment of optimisation algorithms and their effectiveness

¹Only the derivatives of the states with respect to time are made explicit in the general model equations for simplicity of notation. However, in general, the model equations may be defined not only in the time domain, but also in the space domain and reactor model equations may involve functional relationships among partial derivatives of states with respect to time and space coordinates.

requires their respective objective function to be well-conditioned (Wilson et al., 2015; White et al., 2016). Ill-conditioned objective functions derive from the attempt of identifying models whose parametrisation is sloppy given the available dataset **and the level of noise present in the system** (Chis et al., 2014; White et al., 2016). Sloppiness arises when measured model responses are poorly sensitive to the change of some parameters and/or measurements do not carry sufficient information to bring parameter correlation below a critical threshold (typically considered as high as 95%). Whenever these circumstances occur, the eigenvalues of the covariance matrix of the parameter estimates span over a wide range of orders of magnitude, i.e. the condition number of the covariance matrix is very high.

Numerical failures may occur in the course of the model identification problem in the presence of a sloppy parametrisation. These may be classified as follows:

- *False convergence.* Ill-conditioned objective functions both in the parameter estimation and in the optimal MBDoE problem may cause numerical optimisation routines to fail in converging to the optimal solution (Higham, 1996).
- *Inaccuracy in the computation of gradients.* The calculation of the sensitivities (i.e. partial derivatives in the parameter space) using direct differential methods is frequently impractical. As a consequence, numerical differentiation routines are regularly employed in model building practice (Saltelli et al., 2000). The numerical computation of sensitivities requires a perturbation of the model parameter values. The computed sensitivities are *sensitive* to the choice of the perturbation. In the presence of a sloppy parametrisation, the applied perturbation may not be appropriate to accurately quantify the gradient in the parameter space (Higham, 1996). As a consequence, the Hessian and covariance matrix computed as functions of the parameter sensitivities may be inaccurate, affecting the model validation process and the design of following experiments (Pukelsheim, 2006).
- *Inaccuracy in the inversion of matrices.* In the presence of a sloppy parametrisation, the covariance matrix of the parameter estimates is ill-conditioned (White et al., 2016). The solution of an optimal MBDoE problem requires the inversion of an ill-conditioned covariance matrix if the parametrisation is sloppy (Franceschini and Macchietto, 2008b).

Different approaches have been proposed in the literature to address the identifiability problem of sloppy models (Dovi et al., 1994):

1. *Experimental-design-based (ED) methods.* These methods are based on the design of optimal experiments for *reshaping* the covariance matrix of the parameter estimates and improve the condition number. For more information on these approaches, the reader is referred to the relevant literature on design criteria for relaxing model sloppiness and reducing parameter correlation (Hosten, 1974; Pritchard and Bacon, 1978; Versyck and Van Impe, 1997; Galvanin et al., 2007; Franceschini and Macchietto, 2008a,d,c; Maheshwari et al., 2013; Chis et al., 2014; Wilson et al., 2015).

2. *Regularisation-based (RG) methods.* Regularisation involves the introduction of a bias in the parameter estimates with the aim of constraining their variance and, concomitantly, reducing the condition number associated to the parameter estimation problem (Barz et al., 2016). Popular regularisation techniques are *i*) Tikhonov regularization (Johansen, 1997; Hansen, 2005; Bardow, 2008) *ii*) truncated singular value decomposition (Hansen, 2005; Lopez C. et al., 2015) and *iii*) parameter subset selection (Barz et al., 2013; Lopez C. et al., 2015).
3. *Reparametrisation-based (RP) methods.* The aim of reparametrisation is transforming the original parameter space Θ into a robust parameter space Ω where both parameter estimation and MBDoE can be performed more effectively on well-conditioned objective functions (Agarwal and Brisk, 1985b,a). Although there is no theoretical advantage in the use of a reparametrised model (Rimensberger and Rippin, 1986; Dovi et al., 1994), the performance of model identification algorithms is sensitive to the type of parametrisation used (Espie and Macchietto, 1988). The effectiveness of RP-based methods has been recognised in many kinetic studies in the literature (Espie and Macchietto, 1988; Asprey and Naka, 1999; Benabbas et al., 2005; Schwaab and Pinto, 2007; Schwaab et al., 2008; Buzzi-Ferraris and Manenti, 2009).

These methods present strengths and weaknesses. ED-based methods are systematic. Optimal ED criteria to relax model sloppiness can be easily implemented into a computer program. However, even optimally designed experiments may not be sufficient to bring the condition number below critical levels. This weakness of ED-based methods is typically associated to either a too narrow range of explorable experimental conditions and/or an insufficient experimental budget to perform these optimal experiments. Furthermore, optimally designed experiments to reduce the condition number may not carry optimal amounts of information for the estimation of the model parameters. This limitation is typically overcome by designing experiments that represent a compromise between improving the parameter statistics and reducing the condition number (Franceschini and Macchietto, 2008c; Maheshwari et al., 2013).

An advantage of RG-based and RP-based methods is that they do not require the execution of experiments for improving the condition number and one can devote the entire experimental budget on improving the statistics of the parameter estimates. In RG-based approaches, the condition number is controlled through the introduction of prior information on the model parameter values. Systematic approaches, e.g. approaches based on Bayesian inference (MacKay, 1992), are available in the literature for supporting the selection of appropriate priors (Hansen, 2005). The introduction of prior information in the parameter estimation problem generally results in the computation of biased parameter estimates.

In contrast to RG-based approaches, RP-based methods do not involve the introduction of any bias in the model identification problem. Ad hoc strategies to reparametrise sloppy models were suggested for very specific kinetic model structures, e.g. Arrhenius-type reaction rates (Asprey and Naka, 1999; Schwaab and Pinto, 2007; Schwaab et al., 2008; Buzzi-Ferraris and Manenti, 2009). However, only few systematic approaches to the reparametrisation of sloppy models are available in the literature (Espie and Macchietto, 1988). An additional feature of RP-based methods is that whenever a model is reparametrised, the parametrisation

195 is fixed until the end of the experimental campaign. However, sloppiness is a consequence
of the combination of both the model parametrisation and the dataset available to identify
the model. There is no theoretical guarantee that the reparametrised model will not become
sloppy after the collection of new data (Wilson et al., 2015). The arising of sloppiness may be
200 i.e. by reparametrising the model every time new data are collected and included in the
parameter estimation problem. Nonetheless, online applications of RP-based methods seem
to be missing in the scientific literature.

In the following section, a RP-based framework for the identification of sloppy models
in automated model identification platforms is proposed. In the framework, a systematic
205 approach to model reparametrisation is introduced and applied online to maintain a small
condition number even when new data are collected by the automated system and included
in the parameter estimation problem.

2.2 Proposed methodology

The original set of equations (1) is initially extended including a linear system of equations
210 to transform the parameter space.

$$\begin{aligned}
\mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) &= \mathbf{0} \\
\hat{\mathbf{y}} &= \mathbf{h}(\mathbf{x}) \\
\boldsymbol{\theta} &= \mathbf{G}\boldsymbol{\omega}
\end{aligned} \tag{2}$$

In (2), $\boldsymbol{\omega} \in \Omega$ represents the $N_\theta \times 1$ array of model parameters in the transformed parameter space Ω , \mathbf{G} is a $N_\theta \times N_\theta$ matrix which transforms the parameter space Ω to the original model parameter space Θ . An online approach to model reparametrisation in automated
215 model identification platforms is now introduced with the aim of effectively estimating the
original parameter set $\boldsymbol{\theta} \in \Theta$. A block diagram showing the proposed procedure is given in
Figure 1. The procedure starts from the availability of preliminary experimental data and
the model structure (1). The parameter transformation \mathbf{G} is initially set equal to \mathbf{I}_θ , where
 \mathbf{I}_θ is the $N_\theta \times N_\theta$ identity matrix, i.e. the parameter spaces Θ and Ω are initially coincident.
220 The model identification algorithm is then called providing the available dataset as input.
The fundamental steps in the algorithm are now illustrated:

1. *A primary parameter estimation step.* At this stage, the set of transformed parameters $\boldsymbol{\omega}$ is estimated fitting the available dataset using a maximum likelihood approach (Bard, 1974). The Hessian of the likelihood function is then computed to characterise
225 the geometry of the parameter space and quantify its *sloppiness*.
2. *A parametrisation update step.* The Hessian matrix computed at the *primary parameter estimation* step is employed to compute and update the transformation matrix \mathbf{G} with the aim of minimising the condition number (i.e. eliminating the sloppiness) given the available dataset.

- 230 3. *A secondary parameter estimation step.* The model parameters $\boldsymbol{\omega} \in \Omega$ are estimated after the *parametrisation update* step and their statistical quality is quantified computing their covariance matrix \mathbf{V}_ω . Parameter estimates and related covariance computed in the transformed parameter space Ω are then transformed to the original parameter space Θ and returned as output.
- 235 4. *An optimal MBDoE for parameter precision step.* If parameter statistics in Θ are unsatisfactory and the experimental budget allows for additional samples to be collected, the experimental activity shall proceed. Optimal experimental conditions are identified at this stage through MBDoE techniques for parameter precision (Franceschini and Macchietto, 2008b) and transmitted to the automated platform for collecting the
- 240 next sample. Notice that in the proposed procedure the optimal MBDoE step occurs in the transformed parameter space Ω .

The illustrated steps constitute an iteration in the presented online framework. These are further detailed in the following subsections. **The computational burden associated with the application of the proposed methodology is comparable with standard parameter estimation algorithms based on parameter fitting.** The procedure shows how it is possible to achieve an effective estimation of parameters in a (potentially) sloppy parameter space Θ by invoking the parameter estimation and the MBDoE algorithms in a conveniently transformed, non-sloppy, parameter space Ω . The values of the estimates and the related covariance obtained in the robust space Ω are transformed to the original parameter space Θ by applying linear

250 transformations, which are computationally more robust operations than optimisations.

2.2.1 Primary parameter estimation

The available dataset Y is provided to the model identification algorithm (see Figure 1). The dataset Y consists N samples of \mathbf{y} , i.e. $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$. It is assumed that the measurements for \mathbf{y} are affected by Gaussian noise with zero mean and covariance $\boldsymbol{\Sigma}$. The transformation matrix \mathbf{G} is set equal to the *primary* transformation matrix \mathbf{G}_P . At the beginning of the model identification procedure \mathbf{G}_P is initialised as the identity matrix \mathbf{I}_θ . A primary estimation of the model parameters $\hat{\boldsymbol{\omega}}_P$ is performed as in (4) maximising the log-likelihood function (3).

255

$$\begin{aligned} \Phi(\boldsymbol{\omega}|Y)|_{\mathbf{G}=\mathbf{G}_P} = & -\frac{N}{2}[N_y \ln(2\pi) + \ln(\det(\boldsymbol{\Sigma}))] \\ & -\frac{1}{2} \sum_{i=1}^N [\mathbf{y}_i - \hat{\mathbf{y}}_i(\boldsymbol{\omega})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{y}_i - \hat{\mathbf{y}}_i(\boldsymbol{\omega})] |_{\mathbf{G}=\mathbf{G}_P} \end{aligned} \quad (3)$$

$$\hat{\boldsymbol{\omega}}_P = \arg \max_{\boldsymbol{\omega} \in \Omega} \Phi(\boldsymbol{\omega}|Y)|_{\mathbf{G}=\mathbf{G}_P} \quad (4)$$

In (3), the quantity $\hat{\mathbf{y}}_i$ represents the model prediction for the sample \mathbf{y}_i . The negative Hessian \mathbf{H} of the log-likelihood function is then computed to evaluate the geometrical properties of the log-likelihood profile in proximity of the maximum likelihood estimate as

260 in (5).

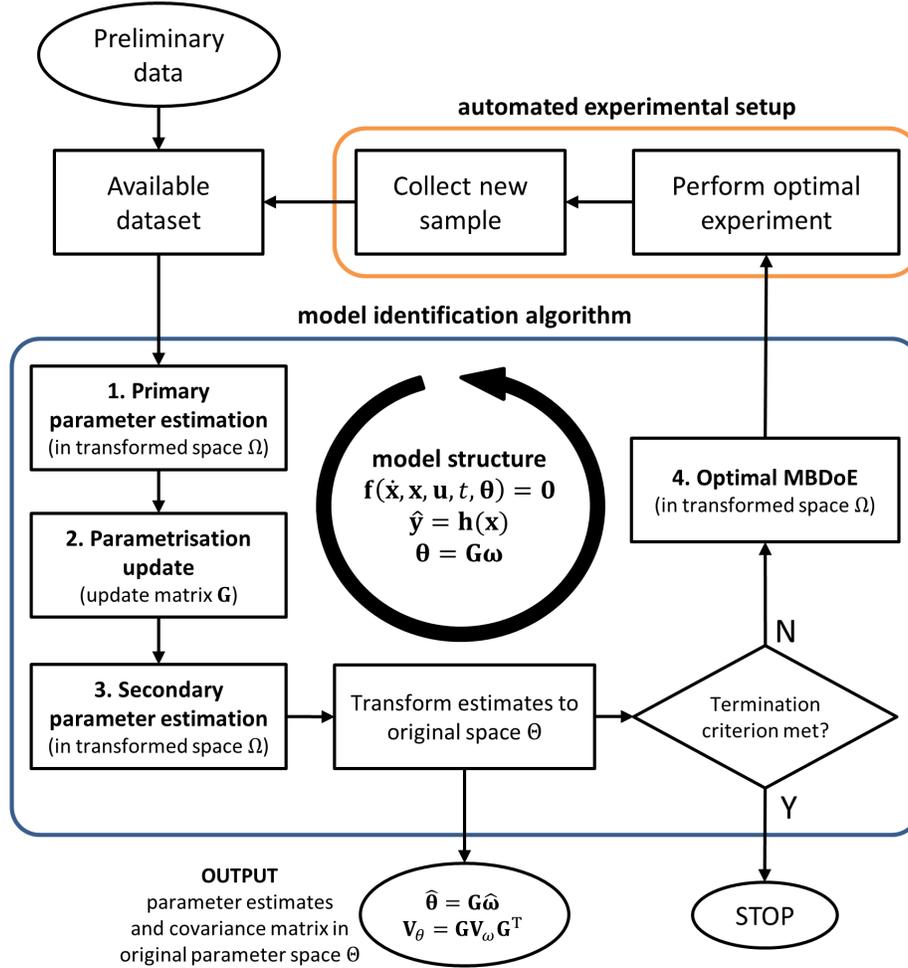


Figure 1: Proposed framework for the online identification of models in automated model identification platforms. Fundamental step in the procedure is the update of the parametrisation matrix \mathbf{G} after the collection and fitting of each sample. The online modification of the model parametrisation is performed to maintain a high computational performance at the parameter estimation and optimal MBDoe stages in the procedure.

$$\mathbf{H}(\hat{\boldsymbol{\omega}}_P)|_{\mathbf{G}=\mathbf{G}_P} = -\nabla\nabla^T\Phi(\hat{\boldsymbol{\omega}}_P|Y)|_{\mathbf{G}=\mathbf{G}_P} \quad (5)$$

In (5), the symbol ∇ defines the gradient operator in the parameter space Ω . Matrix \mathbf{H} is also known as the observed Fisher information matrix and its inverse quantifies the covariance matrix of the parameter estimates (Pukelsheim, 2006).

Notice that the model may be sloppy at the *primary parameter estimation* stage and the condition number may be very high, leading to numerical inaccuracy in the computation of the primary parameter estimate in (4) and in the computation of the Hessian in (5). Numerical results in Section 4 show that the performance of the online RP approach is not affected significantly by this aspect, but further analysis is required. Assessing the sensitivity of the proposed approach to numerical inaccuracies at the *primary parameter estimation* stage is going to be object of future research activities.

2.2.2 Parametrisation update

An eigendecomposition of the matrix (5) is performed at this stage with the aim of diagnosing the structure of the log-likelihood function in proximity of the maximum likelihood estimate and compute an opportune update to the transformation matrix \mathbf{G} . Let $\boldsymbol{\Lambda}$ be the diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \dots, \lambda_{N_\theta}$ of the observed Fisher information matrix (5). The eigenvalues of the observed Fisher information matrix represent the inverse eigenvalues of the covariance of the parameter estimates and the ratio between the maximum and the minimum eigenvalue represents the condition number κ .

$$\kappa = \frac{\max_i \lambda_i}{\min_i \lambda_i} \quad (6)$$

Let matrix \mathbf{U} be the matrix whose columns represent the right normalised eigenvectors of the observed Fisher information matrix (5). Matrix $\boldsymbol{\Lambda}$ and matrix \mathbf{U} quantify the sloppiness of the model in a more readable format. In fact, the eigenvalues and eigenvectors of the negative Hessian (5) respectively quantify the extent of the sloppiness and the directions in the parameter space which are associated to the sloppy behaviour of the model (Lopez C. et al., 2015). A family of *secondary* transformations \mathbf{G}_S can be constructed from \mathbf{G}_P , \mathbf{U} and $\boldsymbol{\Lambda}$ as in (7) with the aim of minimising the condition number of the problem (i.e. making $\kappa = 1.0$).

$$\mathbf{G}_S = d \mathbf{G}_P \mathbf{U} \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{R} \quad (7)$$

The family of transformations given in (7) is parametrised by the scalar $d > 0$ and by the matrix \mathbf{R} , which represent respectively a scaling factor and a rotation matrix in the parameter space. The condition number κ is not influenced by the choice of d and \mathbf{R} . However, the omission of d and \mathbf{R} from (7) (the omission is equivalent to setting $d = 1.0$ and $\mathbf{R} = \mathbf{I}_\theta$ in (7)) may result in a transformation to a new parameter space in which there is significant discrepancy in the orders of magnitude of the model parameters. Model identification algorithms are influenced by the relative scale of parameters, e.g. in the computation of the gradients and, consequently, in the computation of the covariance of parameter estimates (Saltelli et al., 2000). Working with parameters sharing the same order of magnitude is

therefore desirable to avoid discrepancies on how the model identification algorithm handles different directions of the parameter space. In this work, the scaling factor d and the matrix \mathbf{R} are computed to map the *primary* parameter estimate $\hat{\boldsymbol{\omega}}_P$ into the parameter vector whose entries are all equal to 100.0 (this value was chosen arbitrarily) (Zhelezov, 2017). More specifically, the rotation applied by \mathbf{R} and the scaling factor d are computed to satisfy the equality $\mathbf{G}_P \hat{\boldsymbol{\omega}}_P = 100.0 \cdot \mathbf{G}_S \mathbf{1}_\theta$ where the vector $\mathbf{1}_\theta$ is the $N_\theta \times 1$ array whose entries are all equal to 1.0.

The secondary transformation matrix \mathbf{G}_S , computed as in (7), is then used to *update* the primary transformation matrix \mathbf{G}_P that will be used at the following iteration in the procedure of Figure 1.

2.2.3 Secondary parameter estimation

The aim at the *secondary parameter estimation* stage is obtaining a more accurate estimate for the parameters in the transformed space Ω . This is done by repeating the estimation of the parameters $\boldsymbol{\omega}$ after the *parametrisation update* stage, i.e. after the transformation of the (possibly) sloppy parameter space in a more robust, non sloppy parameter space. The log-likelihood function of the model is optimised as in (8) with $\mathbf{G} = \mathbf{G}_S$ obtaining the *secondary* parameter estimate $\hat{\boldsymbol{\omega}}_S$.

$$\hat{\boldsymbol{\omega}}_S = \arg \max_{\boldsymbol{\omega} \in \Omega} \Phi(\boldsymbol{\omega} | Y) |_{\mathbf{G}=\mathbf{G}_S} \quad (8)$$

In principle, the *primary* and the *secondary* parameter estimates satisfy the equality $\mathbf{G}_P \hat{\boldsymbol{\omega}}_P = \mathbf{G}_S \hat{\boldsymbol{\omega}}_S$. However, numerical algorithms for parameter estimation are sensitive to the model parametrisation (Rimensberger and Rippin, 1986; Dovi et al., 1994). More specifically, the convergence rate of numerical optimisation routines to the maximum likelihood estimate is sensitive to the choice of the transformation matrix \mathbf{G} and the aforementioned equality may not be satisfied in practice (Higham, 1996). The covariance \mathbf{V}_ω is then computed for the *secondary* parameter estimates as the inverse of the observed Fisher information matrix (Bard, 1974).

$$\mathbf{V}_\omega = [\mathbf{H}(\hat{\boldsymbol{\omega}}_S) |_{\mathbf{G}=\mathbf{G}_S}]^{-1} \quad (9)$$

The parameter estimates $\hat{\boldsymbol{\theta}}$ and their associated covariance matrix \mathbf{V}_θ in the original parameter space Θ are then computed by applying the *secondary* transformation to the estimates $\hat{\boldsymbol{\omega}}_S$ and covariance \mathbf{V}_ω computed in the transformed space Ω .

$$\hat{\boldsymbol{\theta}} = \mathbf{G}_S \hat{\boldsymbol{\omega}}_S \quad (10)$$

$$\mathbf{V}_\theta = \mathbf{G}_S \mathbf{V}_\omega \mathbf{G}_S^T \quad (11)$$

In standard parameter estimation algorithms, the computation of the covariance \mathbf{V}_θ requires the inversion of the information matrix in the original parameter space Θ (Bard, 1974). However, in the presence of a sloppy parametrisation, the information matrix in Θ may be ill-conditioned. Notice that, in the proposed framework, the inversion of ill-conditioned matrices is avoided. In fact, matrix inversion is performed in a conveniently transformed

parameter space Ω , as in (9), where the information matrix is well-conditioned. The covariance in the original parameter space \mathbf{V}_θ is then computed as in (11) by applying algebraic transformations, which are numerically more robust operations than matrix inversions.

From the covariance \mathbf{V}_θ , it is possible to derive the confidence intervals for the estimates $\hat{\boldsymbol{\theta}} \in \Theta$ and the correlation coefficient c_{ij} between any estimated parameter pair $\hat{\theta}_i$ and $\hat{\theta}_j$ (Bard, 1974). Let $v_{\theta,ij}$ be the ij -th element of the covariance matrix \mathbf{V}_θ . The confidence interval with significance α for the i -th parameter estimate $\hat{\theta}_i$ can be computed as $\hat{\theta}_i \pm z_{\alpha/2} \sqrt{v_{\theta,ii}}$ where $z_{\alpha/2}$ represents a two-tailed value computed from a standard normal distribution with significance α . The correlation coefficient between any parameter pair $\hat{\theta}_i$ and $\hat{\theta}_j$ can be computed according to (12).

$$c_{ij} = \frac{v_{\theta,ij}}{\sqrt{v_{\theta,ii}v_{\theta,jj}}} \quad \forall i, j \quad (12)$$

The statistical quality of the parameter estimates $\hat{\boldsymbol{\theta}}$ in the original parameter space Θ can be checked through a statistical test (e.g. a t -test) for assessing parameter precision (Walpole et al., 2011).

2.2.4 Optimal MBDoE for parameter precision

If some parameter statistics are not satisfactory and the experimental budget allows for the collection of additional data then the experimental activity will continue with the collection of an additional sample from the automated experimental setup. The following sample will be collected with the aim of minimising the size of the confidence region of the parameter estimates $\hat{\boldsymbol{\theta}} \in \Theta$. Popular measures of the size of the confidence region are (Galvanin et al., 2007; Franceschini and Macchietto, 2008b) *i*) the determinant of the covariance matrix $\det(\mathbf{V}_\theta)$ (i.e. the D-criterion), which quantifies the volume of the confidence ellipsoid of the parameter estimates and *ii*) the trace of the covariance matrix $\text{Tr}(\mathbf{V}_\theta)$ (i.e. the A-criterion), which quantifies the volume of the hyperbox containing the confidence ellipsoid.

Optimal MBDoE problems for parameter precision may be ill-conditioned in the presence of a sloppy parametrisation (White et al., 2016). In fact, the solution of an optimal MBDoE problem requires the inversion of an ill-conditioned matrix if the parametrisation is sloppy. In this work it is proposed to solve the MBDoE problem in the robust parameter space Ω with the aim of minimising the size of the confidence region in the original parameter space Θ . In general, the optimal experimental conditions depend on the type of criterion adopted for the design and on the model parametrisation. In this study, the D-criterion is employed because it is invariant under transformations of the parameter space (Fedorov, 1972; Rimensberger and Rippin, 1986). In fact, the following equality holds:

$$\det(\mathbf{V}_\theta) = \det(\mathbf{G}_S)^2 \det(\mathbf{V}_\omega) \quad (13)$$

It is sufficient to notice that matrix \mathbf{G}_S is not modified at the *optimal MBDoE* stage of the procedure (see Figure 1), i.e. $\det(\mathbf{G}_S)$ represents a constant in the MBDoE problem. Therefore, minimising the determinant of the covariance $\det(\mathbf{V}_\omega)$ in the transformed parameter space Ω is equivalent to minimising the determinant of the covariance $\det(\mathbf{V}_\theta)$ in the original parameter space Θ .

The optimal MBDoe problem in the robust space Ω requires the computation of a prediction for the parameter covariance $\hat{\mathbf{V}}_\omega$ (i.e. the posterior covariance matrix) after the collection of the new sample.

$$\hat{\mathbf{V}}_\omega = [\mathbf{V}_\omega^{-1} + \nabla \hat{\mathbf{y}}(\hat{\boldsymbol{\omega}}_s) \boldsymbol{\Sigma}^{-1} \nabla \hat{\mathbf{y}}(\hat{\boldsymbol{\omega}}_s)^T |_{\mathbf{G}=\mathbf{G}_s}]^{-1} \quad (14)$$

In (14), the second addend in the bracket represents the expected Fisher information matrix of the sample to be designed, which is a function of the experimental design vector $\boldsymbol{\varphi}$. The inverse of the prior covariance matrix \mathbf{V}_ω is also included in (14) to quantify the preliminary information that is available from previously fitted samples. The prior covariance is updated at every iteration of the procedure in Figure 1, i.e. after the collection of each sample, according to (9). The D-optimal experimental conditions $\boldsymbol{\varphi}^*$ for the collection of the following sample are computed solving the following optimisation problem:

$$\boldsymbol{\varphi}^* = \arg \min_{\boldsymbol{\varphi}} \det(\hat{\mathbf{V}}_\omega) \quad (15)$$

The optimised conditions computed as in (15) are then transmitted to the automated experimental setup for the collection of the following sample (see Figure 1).

3 Case study

The proposed algorithm presented in Section 2.2 is tested on a case study. The objective is the identification in an automated platform of a kinetic model of benzoic acid esterification with ethanol (Pipus et al., 2000). The reaction is homogeneous and it is catalysed by sulphuric acid. A description of the automated model identification platform is given in Section 3.1. The modelling assumptions are presented in Section 3.2. The proposed online RP methodology is tested both in-silico (Section 4.1) and experimentally on an automated model identification platform (Section 4.2). For both the simulated and the real cases two experimental campaigns are performed:

- a campaign where the parametrisation matrix is not modified;
- a campaign where the parametrisation matrix is updated online.

The two campaigns are performed to assess the influence of the online RP on the model identification process. The methods adopted for the conduction of the experimental campaigns are detailed in Section 3.3.

3.1 Automated model identification platform

A simplified diagram for the online model identification platform is given in Figure 2. The esterification of benzoic acid with ethanol catalysed by sulphuric acid occurs in a flow microreactor. The microreactor is a 2 m long PEEK tube with a diameter of 250 μm . It is placed in a stirred oil bath whose temperature is controlled by a rope heater. The reactants and the catalyst are injected through the flow reactor by three syringe pumps. Syringe 1 and

400 syringe 2 are filled with two different mixtures of benzoic acid and ethanol. The feed concentration of benzoic acid in the reactor is manipulated by modifying the relative flowrates of syringe 1 and syringe 2. Syringe pump 3 is filled with a 160 g L^{-1} sulphuric acid solution. The flowrate of syringe 3 is kept at 10% of the overall flowrate to maintain a constant concentration of sulphuric acid at 16 g L^{-1} at the inlet of the flow reactor. The mixture
405 at the outlet of the reactor is analysed online by a Jasco HPLC using a 250 mm long, 4.6 mm internal diameter ODS hypersil column with a particle size of $5 \text{ }\mu\text{m}$ (Thermo Fisher Scientific). The HPLC method uses 1.25 mL min^{-1} of a 40% water and 60% acetonitrile mobile phase (percentages refer to volume fractions). The oven is maintained at 303 K and a UV detection at 274 nm is used to detect the composition of the outlet mixture. Samples
410 are diluted using an online auto-sampler device (Syrris Asia) applying a dilution factor of 250.

The experimental conditions which can be manipulated by the automated system are:

- the inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$ in the range 0.9 - 1.55 mol L^{-1} ;
- the flowrate F of the feed mixture to the reactor in the range 7.5 - $30.0 \text{ }\mu\text{L min}^{-1}$;
- 415 • the temperature of the oil bath T in the range 343.0 - 413.0 K.

These constitute independent directions of the explorable space of experimental conditions $\varphi = [C_{\text{BA}}^{\text{IN}}, F, T]$. The experimental setup is controlled through a LabVIEW interface (Elliott et al., 2007) implemented in a 32-bit Windows machine with Intel Core i7-3770 3.40 GHz processor and 4.0 GB of RAM. A script written in Python 2.7 implementing the
420 model identification algorithm presented in Section 2.2 is integrated with LabVIEW for the purposes of online parameter estimation and sample design. **The main Python packages employed in the script are NumPy 1.13 (Oliphant, 2015) for the manipulation of algebraic objects and SciPy 1.1 (Jones et al., 2001) for integrating the kinetic model equations and solving the optimisation problems associated with parameter estimation and MBDoE. Parameter estimation problems are solved using the *Nelder-Mead* method. MBDoE problems are solved employing the *SLSQP* solver.**

430 The *parametrisation update* stage of the algorithm (see Figure 1) was implemented in the Python script as an option that can be activated or deactivated from LabVIEW. This option was implemented to give more flexibility to the user in testing the model identification algorithm both in the presence and in the absence of the online RP method.

3.2 Modelling assumptions

The catalytic esterification of benzoic acid and ethanol is modelled as a single reaction system where benzoic acid (BA) and ethanol (Et) react to produce ethyl benzoate (EB) and water (W) (Pipus et al., 2000).



435 Available studies in the literature report that the reaction is reversible. However, if a large excess of ethanol in the reactor is maintained (as in this work), the reverse reaction can be

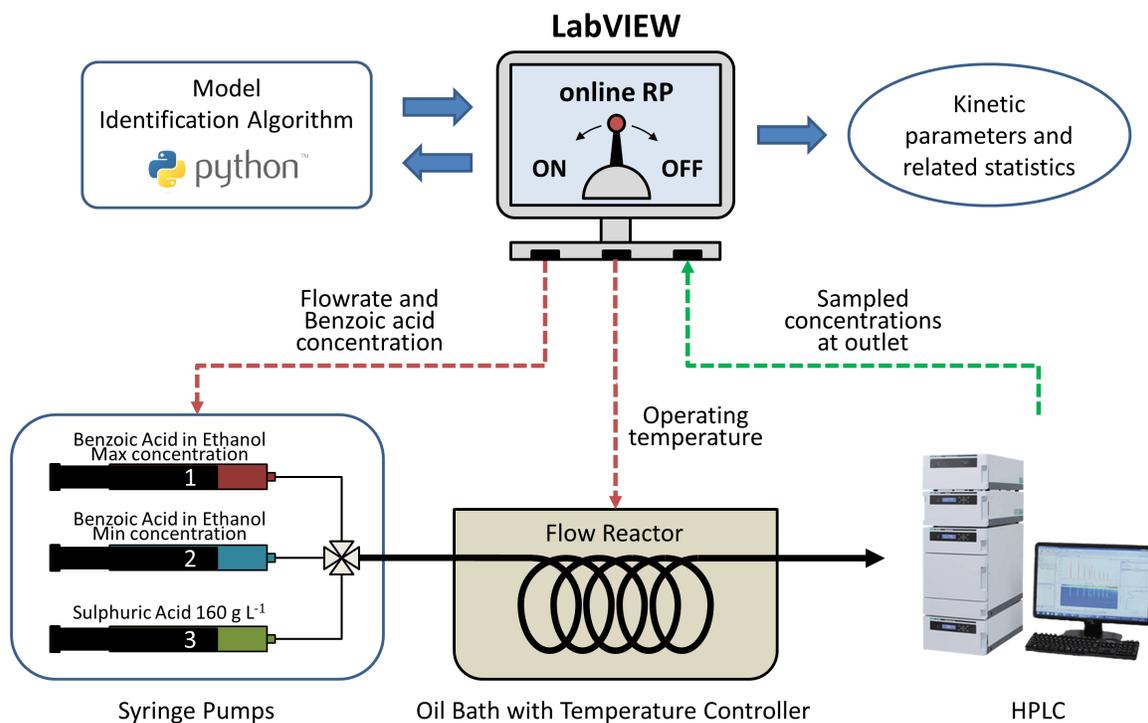


Figure 2: Simplified diagram representing the online model identification platform.

neglected (Pipus et al., 2000). The tubular reactor is modelled as an ideal plug flow reactor operated at isothermal conditions, i.e. thermal and mass transfer resistances are neglected. The validity of plug flow behaviour was checked by evaluating the vessel dispersion number (Levenspiel, 1998; Rossi et al., 2017). A maximum vessel dispersion number of $6.8 \cdot 10^{-4}$ was computed for the flowrate range considered in the study. The computed value is significantly smaller than $1.28 \cdot 10^{-2}$, i.e. the maximum vessel dispersion number recommended in the literature for the validity of the plug flow assumption (Levenspiel, 1998).

The reaction rate is assumed as first order with respect to benzoic acid. Following from the aforementioned assumptions, the steady-state kinetic behaviour of the system is modelled through the following set of ordinary differential equations (17):

$$v \frac{dC_i}{dz} = \nu_i k C_{BA}(z) \quad \forall i = \text{BA, Et, EB, W} \quad (17)$$

In (17), C_i is the concentration of the i -th component in the mixture expressed in molL^{-1} ; z represents the axial spatial coordinate of the tubular reactor expressed in m; v is the axial velocity of the liquid bulk expressed in m s^{-1} ; ν_i is the stoichiometric coefficient of the i -th component in the mixture; k is the rate constant expressed in s^{-1} .

An Arrhenius-type kinetic constant involving a set of two kinetic parameters $\theta = [\theta_1, \theta_2]$ is assumed with the following mathematical structure:

$$k = e^{\theta_1 - \frac{10^4 \theta_2}{RT}} \quad (18)$$

In (18), R is the ideal gas constant expressed in $\text{J mol}^{-1} \text{K}^{-1}$. As one can see from (18), the pre-exponential factor is included as exponent in the rate constant and the activation energy is multiplied by a scaling factor. The above structure for the kinetic rate constant was selected because it is generally recognised as robust within the literature on kinetic parameter estimation (Asprey and Naka, 1999; Buzzi-Ferraris and Manenti, 2009). In other words, parametrisation (18) generally leads to an improvement of the condition number with respect to the original form of the Arrhenius constant, i.e. $k = Ae^{-E_a/RT}$, parametrised by pre-exponential factor A and activation energy E_a .

3.3 Objective and methods

The objective of the study is the estimation of the kinetic parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$ with the smallest volume confidence region of $\hat{\boldsymbol{\theta}}$ by conducting an experimental campaign on the online model identification platform with an available budget of 9 samples. A sample is constituted by the single measurement of ethyl benzoate concentration at the outlet of the reactor, i.e. $\mathbf{y} = [C_{\text{EB}}^{\text{OUT}}]$ [mol L^{-1}]. The measurement error is modelled as Gaussian noise with covariance matrix $\boldsymbol{\Sigma} = [2.5 \cdot 10^{-5}]$, i.e. a standard deviation of 0.0165 molL^{-1} is assumed to model the Gaussian measurement noise for $C_{\text{EB}}^{\text{OUT}}$. The experimental conditions for the collection of samples 1, 2 and 3 are fixed to the values reported in Table 1. The following samples, i.e. samples from 4 to 9, are designed by the model identification algorithm by employing a D-optimal criterion, i.e. by solving an MBDoe problem in the form (15).

Two cases are proposed to test the model identification algorithm implemented in the online model identification platform:

1. *Simulated case: samples generated in-silico.* Samples are generated simulating the experiments with the kinetic model (17) setting the kinetic parameters equal to the value $\boldsymbol{\theta}^* = [15.27, 7.60]$ and adding Gaussian noise with covariance $\boldsymbol{\Sigma}$.
2. *Real case: samples collected from the experimental platform.* In this case, samples are collected from the experimental platform described in Section 3.1. An interval of 65 min is allowed between the collection of samples to let the system reach steady-state conditions.

For both the *Simulated* and the *Real* case, two experimental campaigns are performed: 1) a non-RP campaign in which the online reparametrisation is not activated; 2) a RP campaign in which the online reparametrisation is activated. This is done to provide a comparison of the performance of the model identification algorithm both in the presence and in the absence of the online RP method. In the *Simulated* case, the effect of the online RP is assessed comparing statistically the parameter estimates $\hat{\boldsymbol{\theta}}$ computed in the two campaigns with the target parameter value $\boldsymbol{\theta}^* = [15.27, 7.60]$. This is done by means of a χ^2 -test in the parameter space Θ . This involves testing the null hypothesis that the following statistic $\chi_{\hat{\boldsymbol{\theta}}}^2$ is distributed as a χ^2 distribution with degree of freedom $N_{\boldsymbol{\theta}} = 2$.

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \mathbf{V}_{\hat{\boldsymbol{\theta}}}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \chi_{\hat{\boldsymbol{\theta}}}^2 \sim \chi^2 \quad (19)$$

490 A small p -value associated to the statistic χ_{θ}^2 (e.g. smaller than 1.0%) is interpreted as an index of failure of the model identification algorithm in estimating the target parameter values. In the *Real* case, the target parameter value θ^* is unknown. Furthermore, a discrepancy in the parameter estimates between the RP and the non-RP campaigns is not only caused by numerical reasons, but also by problems of experimental repeatability caused by external
 495 disturbances (Alberton et al., 2009). The presence of disturbances can lead to changes in the parameters of the population from which experimental data are sampled and the concomitant inclusion of outliers in the dataset (Huber, 2004). It is recognised that, in the presence of such uncertainty sources, a statistical analysis to validate the models identified in the two campaigns would not be significant and it is therefore omitted.

500 Confidence intervals and correlation coefficient for the parameter estimates (see Section 2.2.3) are recorded in the course of the experimental campaigns and they are reported in the Results section. The condition number κ is also recorded in the course of the experimental campaigns and it is reported to demonstrate the performance of the online RP in improving and maintaining the well-posedness of the model identification problem.

Table 1: Experimental conditions φ adopted for the collection of samples 1 to 3 in the experimental campaigns: inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$; flowrate F ; temperature of the oil bath T .

Sample number	$C_{\text{BA}}^{\text{IN}}$ [mol L ⁻¹]	F [$\mu\text{L min}^{-1}$]	T [K]
1	1.50	20.0	413.0
2	1.00	10.0	393.0
3	1.25	15.0	403.0

505 4 Results

4.1 Simulated case: samples generated in-silico

The estimates for the kinetic parameters θ_1 and θ_2 for the non-RP campaign are reported in Table 2 together with information on their statistical quality. More specifically, the 95% confidence intervals and the correlation coefficient c_{12} between the kinetic parameters θ_1
 510 and θ_2 are reported. One can see from Table 2 that the correlation coefficient c_{12} remains above 99.96% in the course of the campaign. The parameter estimation and the MBDoE problems are solved in the original parameter space Θ where the condition number of the log-likelihood function remains above $6.1 \cdot 10^3$ throughout the whole experimental campaign. The χ^2 -test was conducted to compare statistically the computed parameter distribution
 515 with the target parameter value θ^* (see Section 3.3 for information on how the test statistic is computed). As one can see from Table 2, a p -value of 0.00% in the course of the non-RP campaign suggests that the parameter estimates computed by the algorithm are statistically inconsistent with the target parameter values.

Table 2: Simulated case: non-RP campaign. Parameter estimates are reported together with their respective 95% confidence intervals and correlation coefficient in the course of the experimental campaign. Parameter estimation and MBDoE problems are solved in the original parameter space Θ . The condition number of the log-likelihood function in Θ is reported in the table.

Simulated case - non-RP campaign					
Samples collected	Estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals		Correlation coefficient c_{12}	p -value of target parameters θ^*	Condition number κ in Θ
1	[- , -]		-	-	-
2	[- , -]		-	-	-
3	[12.15 \pm 2.14 , 6.56 \pm 1.35]		0.9998	0.00%	$1.4 \cdot 10^4$
4	[14.83 \pm 1.22 , 7.47 \pm 0.81]		0.9996	0.00%	$6.1 \cdot 10^3$
5	[15.99 \pm 1.01 , 7.85 \pm 0.70]		0.9998	0.00%	$1.0 \cdot 10^4$
6	[15.06 \pm 0.79 , 7.53 \pm 0.53]		0.9997	0.00%	$7.2 \cdot 10^3$
7	[14.90 \pm 0.74 , 7.47 \pm 0.50]		0.9997	0.00%	$9.2 \cdot 10^3$
8	[14.84 \pm 0.66 , 7.45 \pm 0.44]		0.9997	0.00%	$8.2 \cdot 10^3$
9	[14.94 \pm 0.63 , 7.49 \pm 0.42]		0.9998	0.00%	$9.6 \cdot 10^3$

Table 3: Simulated case: RP campaign. Parameter estimates in the course of the experimental campaign are reported together with their respective 95% confidence intervals and correlation coefficient. Parameter estimation and MBDoE problems are solved in the transformed parameter space Ω . The condition number of the log-likelihood function in Ω is reported in the table.

Simulated case - RP campaign					
Samples collected	Estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals		Correlation coefficient c_{12}	p -value of target parameters θ^*	Condition number κ in Ω
1	[- , -]		-	-	-
2	[- , -]		-	-	-
3	[16.44 \pm 64.52 , 8.01 \pm 25.05]		0.9999	0.00%	$5.5 \cdot 10^8$
4	[16.61 \pm 3.55 , 8.06 \pm 1.21]		0.9999	68.26%	$3.8 \cdot 10^2$
5	[15.60 \pm 2.01 , 7.72 \pm 0.68]		0.9998	86.41%	$1.2 \cdot 10^0$
6	[15.72 \pm 1.62 , 7.76 \pm 0.55]		0.9997	70.47%	$1.0 \cdot 10^0$
7	[15.72 \pm 1.50 , 7.76 \pm 0.51]		0.9998	69.84%	$1.0 \cdot 10^0$
8	[15.59 \pm 1.44 , 7.71 \pm 0.49]		0.9998	56.62%	$1.0 \cdot 10^0$
9	[15.39 \pm 1.24 , 7.64 \pm 0.42]		0.9998	64.74%	$1.0 \cdot 10^0$

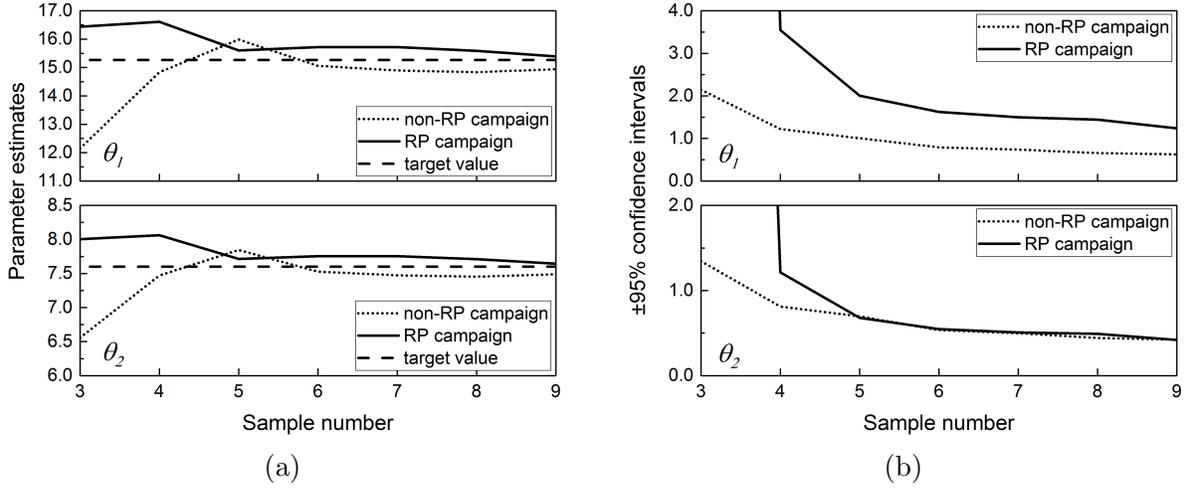


Figure 3: Simulated case: (a) parameter estimates and (b) $\pm 95\%$ confidence intervals throughout the non-RP campaign (dotted) and the RP campaign (solid). In subfigure (a), the target parameters are indicated by a dashed line.

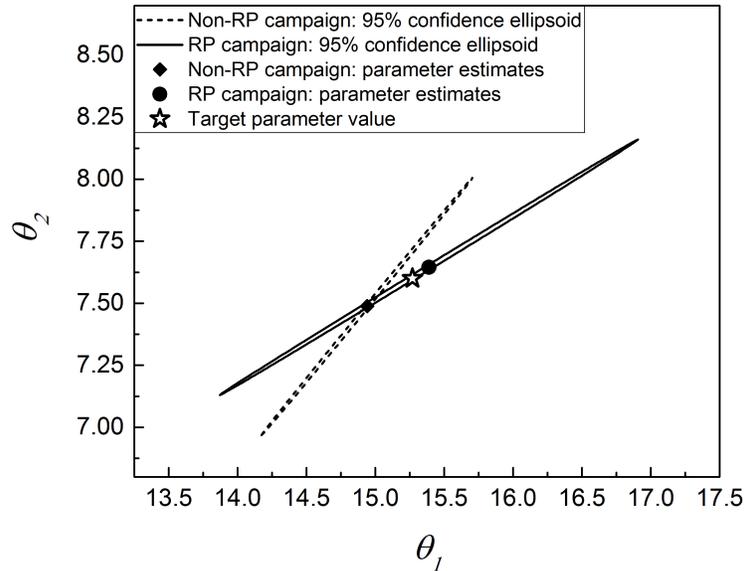


Figure 4: Simulated case: parameter estimates and related 95% confidence ellipsoids at the end of the non-RP campaign (dotted) and at the end of the RP campaign (solid). The target parameter value is highlighted in the graph by a star-shaped symbol.

Parameter estimates and related information on their statistical quality are given in
 520 Table 3 for the RP campaign. In the course of the RP campaign, the correlation coefficient
 c_{12} remains above 99.97%. In the RP campaign, the parameter estimation problem and the
 MBDoe problem are solved in the transformed parameter space Ω , where the transformation
 matrix \mathbf{G} is refined after the collection of each sample. The condition number of the log-
 likelihood function in Ω starts from a value of $5.5 \cdot 10^8$ at the first iteration of the model
 525 identification algorithm (i.e. after the collection of 3 samples) and it is reduced to 1.0 at
 the fourth iteration (i.e. after the collection of 6 samples). The benefit derived from the
 application of the online RP is validated by the χ^2 -test. The p -value of the target value θ^*
 given the computed covariance at the end of the model identification campaign is 64.74%.
 This confirms that the algorithm computed estimates that are statistically consistent with
 530 the target parameter value θ^* .

The parameter estimates and related 95% confidence intervals obtained in the non-RP
 campaign and in the RP campaign are compared graphically in Figure 3a and Figure 3b.
 In Figure 3a, one can see that both the methods present a similar convergence to the target
 parameter values, highlighted with dashed lines in the plot. In Figure 3b, one can see
 535 that the 95% confidence intervals for the parameters are significantly different between the
 non-RP and the RP campaign. In particular the confidence interval of parameter $\hat{\theta}_1$ is
 significantly larger in the RP case than in the non-RP case. The discrepancy is interpreted
 as a consequence of an inaccurate computation of the log-likelihood gradient in the non-RP
 case, which results in an underestimation in the variance of the estimate $\hat{\theta}_1$ (see Section 2.1
 540 for more details).

The final parameter estimates obtained in the non-RP and in the RP campaigns in
 the simulated case are compared graphically in Figure 4. In Figure 4 the final parameter
 estimates are plotted with their respective 95% confidence ellipsoids for the non-RP campaign
 (dotted) and for the RP campaign (solid). The target parameter value is highlighted in Figure
 545 4 by a star-shaped symbol. As one can see from Figure 4 the target value lies within the
 solid ellipsoid of the RP campaign, while it lies outside the dotted ellipsoid of the non-RP
 campaign. The graph shows that the non-RP campaign leads to the misleading conclusion
 that the target parameter values are not the parameters values of the physical system. The
 RP campaign led to a more reliable estimate of the kinetic parameter values.

550 Additional campaigns were performed in-silico to demonstrate that the performance of
 the model identification algorithm is insensitive to a change in the dataset, i.e. it is insensitive
 to a change in the random seed used to generate the data in-silico. The results obtained from
 20 simulated campaigns are reported in Appendix A. **Both in RP and non-RP campaigns,
 each algorithm iteration required only few seconds of CPU time.**

555 4.2 Real case: samples collected from the experimental platform

Two campaigns of experiments, i.e. a non-RP campaign and a RP campaign, were per-
 formed on the automated system. Experimental conditions investigated in the course of the
 campaign and the associated sampled concentrations are given in Appendix B. Parameter
 estimates $\hat{\theta}$ with associated confidence intervals and correlation coefficient are reported in
 560 Table 4 for the non-RP campaign and in Table 5 for the RP campaign. Numerical estimates
 in terms of pre-exponential factor and activation energy were also computed from $\hat{\theta}$. These

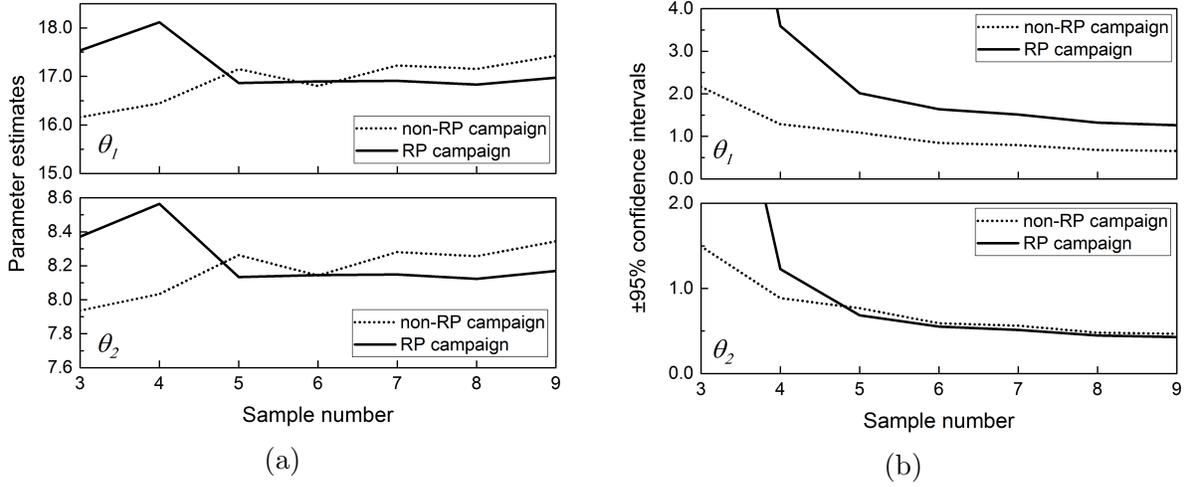


Figure 5: Real case: (a) parameter estimates and (b) $\pm 95\%$ confidence intervals throughout the non-RP campaign (dotted) and the RP campaign (solid).

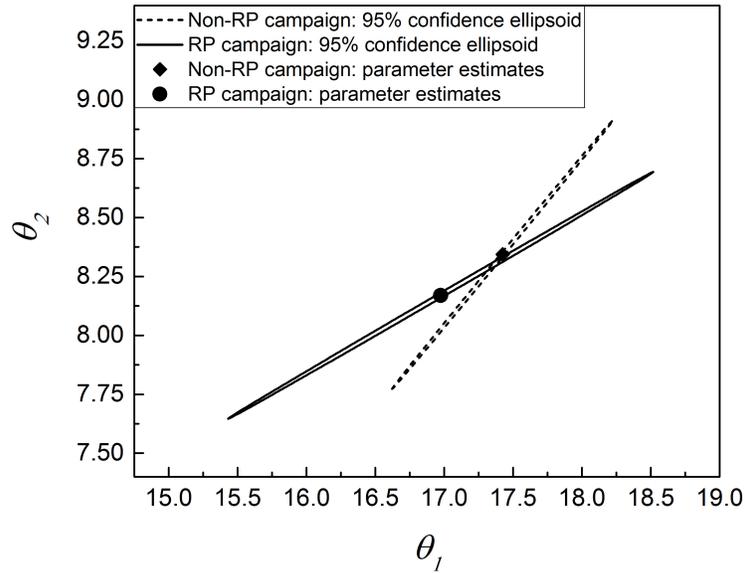


Figure 6: Real case: parameter estimates and related 95% confidence ellipsoids at the end of the non-RP campaign (dotted) and at the end of the RP campaign (solid).

Table 4: Real case: non-RP campaign. Parameter estimates in the course of the experimental campaign with 95% confidence intervals and correlation coefficient. Parameter estimation and MBDoE problems are solved in the original parameter space Θ . The condition number of the log-likelihood function in Θ is reported in the table.

Real case - non-RP campaign					
Samples collected	Estimates $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals			Correlation coefficient c_{12}	Condition number κ in Θ
1	-	,	-	-	-
2	-	,	-	-	-
3	16.16 \pm 2.16	,	7.94 \pm 1.49	0.9998	$1.5 \cdot 10^4$
4	16.44 \pm 1.29	,	8.03 \pm 0.89	0.9996	$6.1 \cdot 10^3$
5	17.15 \pm 1.09	,	8.26 \pm 0.77	0.9998	$1.1 \cdot 10^4$
6	16.80 \pm 0.85	,	8.14 \pm 0.59	0.9997	$7.8 \cdot 10^3$
7	17.23 \pm 0.79	,	8.28 \pm 0.56	0.9998	$1.1 \cdot 10^4$
8	17.15 \pm 0.68	,	8.26 \pm 0.48	0.9997	$8.4 \cdot 10^3$
9	17.42 \pm 0.66	,	8.34 \pm 0.47	0.9998	$1.0 \cdot 10^4$

Table 5: Real case: RP campaign. Parameter estimates in the course of the experimental campaign with 95% confidence intervals and correlation coefficient. Parameter estimation and MBDoE problems are solved in the transformed parameter space Ω . The condition number of the log-likelihood function in Ω is reported in the table.

Real case - RP campaign					
Samples collected	Estimates $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2]$ with 95% confidence intervals			Correlation coefficient c_{12}	Condition number κ in Ω
1	-	,	-	-	-
2	-	,	-	-	-
3	17.54 \pm 13.41	,	8.37 \pm 5.38	0.9999	$2.6 \cdot 10^7$
4	18.12 \pm 3.59	,	8.56 \pm 1.23	0.9999	$8.0 \cdot 10^2$
5	16.86 \pm 2.01	,	8.13 \pm 0.68	0.9998	$1.3 \cdot 10^0$
6	16.90 \pm 1.64	,	8.15 \pm 0.55	0.9997	$1.0 \cdot 10^0$
7	16.91 \pm 1.51	,	8.15 \pm 0.51	0.9998	$1.0 \cdot 10^0$
8	16.83 \pm 1.32	,	8.12 \pm 0.45	0.9997	$1.0 \cdot 10^0$
9	16.98 \pm 1.26	,	8.17 \pm 0.43	0.9998	$1.0 \cdot 10^0$

are reported in Appendix B.

In the course of the non-RP campaign (see Table 4), the parameter correlation c_{12} between $\hat{\theta}_1$ and $\hat{\theta}_2$ remains above 99.96%. In the non-RP campaign the parameter estimation and MBDoe problems are solved in the original parameter space Θ . The condition number of the log-likelihood function in Θ remains above $6.1 \cdot 10^3$ in the course of the non-RP campaign.

The correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is above 99.97% throughout the whole RP campaign (see Table 5). However, in the RP campaign, parameter estimation and MBDoe problems are solved in the transformed parameter space Ω . The condition number in Ω is reduced by the algorithm from an initial value of $2.6 \cdot 10^7$ to the minimum value 1.0 in four iterations (i.e. after the collection of 6 samples). The transformation matrix \mathbf{G} is then adjusted after the collection of each sample to maintain a condition number $\kappa = 1.0$ until the end of the experimental campaign.

The parameter estimates and related 95% confidence intervals obtained in the non-RP and in the RP campaigns are plotted in Figure 5a and Figure 5b. The 95% confidence ellipsoids associated to the final parameter estimates achieved in the non-RP campaign and in the RP campaign are plotted in Figure 6.

Notice that in this case it is not possible to quantify and compare the performance of the two campaigns in retrieving the target parameter value. The target kinetic parameters are in fact unknown in the real case. One can observe from Figure 5a that the estimates achieved in the RP campaign exhibit a convergent behaviour around the values $\boldsymbol{\theta} = [16.90, 8.15]$. Estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ in the non-RP campaign do not exhibit a convergent behaviour, but they tend to increase in the course of the non-RP campaign (see Figure 5a). It is not possible to assess whether the absence of convergence in the non-RP campaign is the consequence of an unknown systematic disturbance in the system. However, it is possible to appreciate that the application of the online RP method led to the minimisation of the condition number (see Table 5) with the concomitant improvement in the numerical performance of the optimisation algorithms. **Also in the real case, both in the RP and in the non-RP campaign, the CPU time required to complete each algorithm iteration was on the order of seconds.**

A goodness-of-fit test was also conducted to demonstrate that, both in the RP and in the non-RP campaign, the postulated first order single-reaction mechanism (see Section 3.2) provided an accurate representation of the chemical system. Nonetheless, it was recognised that an analysis on the goodness-of-fit was not significant for demonstrating the online RP method. It was chosen to report in Appendix B the numerical details regarding the analysis on the fitting quality.

4.3 Results discussion

Both in the simulated and in the real case, the 95% confidence intervals of the estimates after 9 collected samples differ significantly between the non-RP and the RP campaign (see Figure 3b and Figure 5b). In the simulated case, a χ^2 -test was conducted to compare the final statistics on the parameters computed in the RP campaign with the final statistics obtained in the non-RP campaign. It was shown that the confidence region of the parameter estimates computed in the RP campaign *contains* the target parameter value $\boldsymbol{\theta}^*$ while the ellipsoid computed in the non-RP campaign does *not contain* the target value $\boldsymbol{\theta}^*$ (see Figure 4). Hence, it was possible to demonstrate statistically that the campaign with online RP

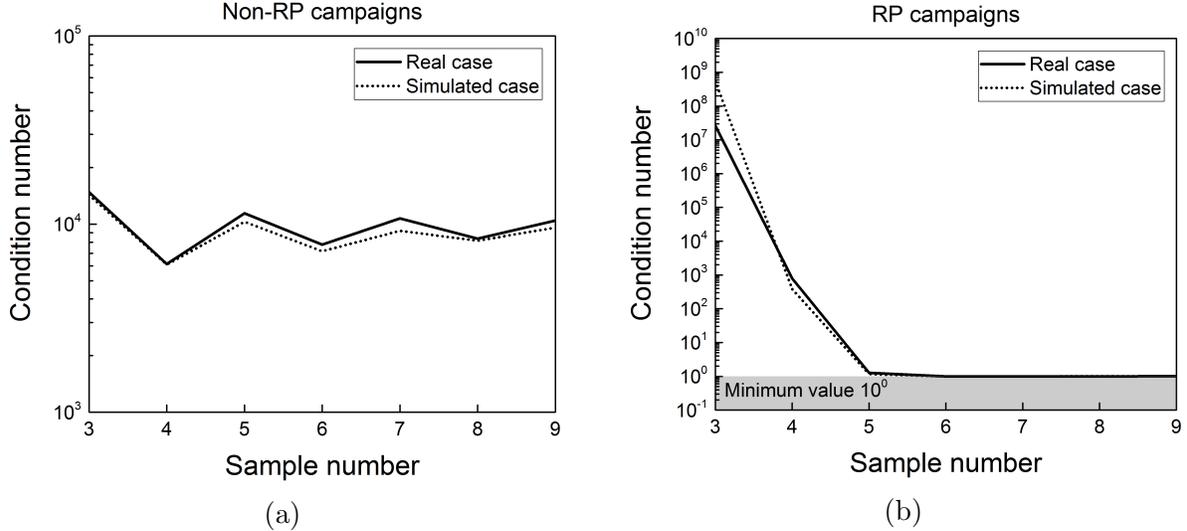


Figure 7: Condition number after each sample collected in the simulated case (dotted line) and in the real case (solid line): (a) non-RP campaigns; (b) RP campaigns.

605 led to a more accurate quantification of the uncertainty region associated to the computed parameter estimates.

Figure 7a and Figure 7b show the condition numbers in the course of the non-RP and RP campaigns respectively. In the non-RP campaigns (see Figure 7a), the condition number κ is around 10^4 and does not vary significantly in the course of the sample collection process. In the RP campaigns, both in the simulated and in the real case, the employment of the online RP method led to the minimisation of the condition number to $\kappa = 1.0$ in an initially ill-conditioned model identification problem (see Figure 7b). From Figure 7b, one can see that, both in the simulated and in the real case, the condition number is minimised to $\kappa = 1.0$ when sample 6 is collected, i.e. after 4 iterations in the model identification algorithm. This is explained by the fact that the update for the transformation matrix \mathbf{G} is evaluated as a function of the Hessian \mathbf{H} computed with the primary transformation matrix \mathbf{G}_P (see Section 2.2).

The condition number in the transformed space associated to \mathbf{G}_P may be very high at the first iteration of the algorithm. A high condition number at the *primary parameter estimation* step may lead to an inaccurate computation of the Hessian (i.e. an inaccurate quantification of the sloppiness) and, consequently, lead to the computation of an inappropriate update for \mathbf{G} . This does not appear to affect the performance of the online RP approach in the presented case study, but further analysis is required. It is object of future research activities to make the proposed algorithm insensitive towards numerical inaccuracies in the initial diagnosis of model sloppiness.

5 Conclusion

A model identification algorithm implementing a novel approach of online reparametrisation, i.e. an approach of online transformation of the model parameter space, is proposed

in this manuscript. The tool is designed specifically to deal with the problem of parameter estimation in the presence of *sloppy* model structures, i.e. models whose parameters are practically non-identifiable and/or highly correlated. The proposed approach to online reparametrisation is based on two fundamental steps: 1) a primary parameter estimation step, which is required to diagnose and quantify the sloppiness of the model parameter space; 2) a parametrisation update step in which the sloppy parameter space is transformed to a robust parameter space with the aim of reducing the sloppiness. Once the model parametrisation is updated, the parameter estimation is repeated solving an optimisation problem in the transformed, non-sloppy, parameter space. Additional samples are then designed by the algorithm employing techniques for optimal design of experiments with the aim of improving the statistical quality of parameter estimation. It is shown that optimisation algorithms benefit significantly from the presence of a robust (i.e. non-sloppy) model parametrisation both at the parameter estimation and at the experimental design stage. Parameter estimates computed in the robust space are then transformed to the original, sloppy, parameter space applying algebraic transformations and returned as output to the user.

The performance of the presented algorithm was tested both in-silico and on a real system where an automated experimental platform has been employed for online kinetic model identification. The objective was the estimation of the kinetic parameters in a two-parameter model of benzoic acid esterification with ethanol catalysed by sulphuric acid in a flow reactor. In both cases, the reparametrisation algorithm iteratively reduced and eventually eliminated model sloppiness minimising the condition number of an originally ill-conditioned model identification problem. In the case study performed in-silico, a set of values for the kinetic parameters was assumed to simulate the experiments. Hence, it was possible to show that the ill-conditioned nature of the model was preventing a conventional model identification algorithm from retrieving the target value of the kinetic constants. The presented model identification algorithm implementing the proposed online reparametrisation method was instead capable of computing estimates that were statistically compatible with the assumed target values of the kinetic parameters.

In the real case, the target values for the kinetic constants were unknown and it was not possible to quantify directly the convergence of the parameter estimates to the target kinetic coefficients. However, it was possible to appreciate that, also in the experimental campaign performed on the real system, the model identification algorithm implementing the online reparametrisation routine iteratively reduced and eventually minimised the condition number. The minimisation of the condition number to unity and the concomitant elimination of model sloppiness resulted in an improved performance of optimisation algorithms employed in the course of the model identification process.

The proposed reparametrisation method was integrated as an optional step in an online model identification algorithm implemented in a Python script. It was shown that the computational performance of the algorithm was not affected significantly by the additional step of model reparametrisation. The modest computational cost associated to the reparametrisation step and the low memory requirement of the method make it suitable for implementation also on embedded devices. Future research activities will focus primarily on three aspects: 1) improving the efficiency of the proposed method by reducing the number of iterations required to minimise the condition number and eliminate model sloppiness; 2) validating the proposed approach on more complex model structures, e.g. kinetic models in-

675 involving a higher number of parameters and multiple measured model responses; 3) extending the proposed online reparametrisation algorithm including routines for applying nonlinear transformations to the parameter space.

Acknowledgements

This work was supported by the Hugh Walter Stern PhD Scholarship, University College London; the Department of Chemical Engineering, University College London; and the EP-
680 SRC Grant EP/J017868/1.

Symbols used

Latin symbols

A	pre-exponential factor
c_{ij}	correlation coefficient between θ_i and θ_j
C_i	concentration of species i
C_i^{IN}	concentration of species i at the inlet
C_i^{OUT}	concentration of species i at the outlet
d	scaling factor of parameter space (> 0)
E_a	activation energy
F	volumetric flowrate
k	kinetic constant
N	number of samples in the available dataset Y
N_f	number of functions in a given kinetic model
N_{MAX}	maximum number of samples collectable
N_u	number of independent inputs in a given kinetic model
N_x	number of state variables in a given kinetic model
N_y	number of output variables in a given kinetic model
N_θ	number of non-measurable parameters in a given model
R	ideal gas constant
t	time
T	temperature
U	vector space of model inputs
v	flow velocity along the axial coordinate of microchannel
$v_{\theta,ij}$	ij -th element of the covariance matrix \mathbf{V}_θ
Y	dataset available for model identification
z	axial coordinate of microchannel
$z_{\alpha/2}$	two-tailed score of standard normal distribution with significance α

Matrices and vectors

$\mathbf{1}_\theta$	column array whose entries are all equal to 1 [$N_\theta \times 1$]
---------------------	---

\mathbf{f}	column array of functions $[N_f \times 1]$
\mathbf{G}	linear transformation of parameter space $\Omega \rightarrow \Theta$ $[N_\theta \times N_\theta]$
\mathbf{G}_P	primary transformation of parameter space $\Omega \rightarrow \Theta$ $[N_\theta \times N_\theta]$
\mathbf{G}_S	secondary transformation of parameter space $\Omega \rightarrow \Theta$ $[N_\theta \times N_\theta]$
\mathbf{h}	column array of functions $[N_y \times 1]$
\mathbf{H}	observed Fisher information matrix $[N_\theta \times N_\theta]$
\mathbf{I}_θ	identity matrix $[N_\theta \times N_\theta]$
\mathbf{R}	matrix of rotation of parameter space $[N_\theta \times N_\theta]$
\mathbf{u}	column array of independent control variables (model inputs) $[N_u \times 1]$
\mathbf{U}	right normalised eigenbasis of \mathbf{H} $[N_\theta \times N_\theta]$
\mathbf{V}_θ	covariance of parameter estimates in Θ $[N_\theta \times N_\theta]$
\mathbf{V}_ω	covariance of parameter estimates in Ω $[N_\theta \times N_\theta]$
$\hat{\mathbf{V}}_\omega$	predicted covariance of parameter estimates in Ω $[N_\theta \times N_\theta]$
\mathbf{x}	column array of state variables $[N_x \times 1]$
\mathbf{y}	sample - column array of measured output variables $[N_y \times 1]$
\mathbf{y}_i	i -th sample in dataset Y $[N_y \times 1]$
$\hat{\mathbf{y}}$	column array of predicted output variables $[N_y \times 1]$
$\hat{\mathbf{y}}_i$	column array of predicted output variables for sample \mathbf{y}_i $[N_y \times 1]$
$\boldsymbol{\theta}$	column vector of parameters in parameter space Θ $[N_\theta \times 1]$
$\boldsymbol{\theta}^*$	column vector of target parameters in parameter space Θ $[N_\theta \times 1]$
$\hat{\boldsymbol{\theta}}$	maximum likelihood estimate for $\boldsymbol{\theta} \in \Theta$ $[N_\theta \times 1]$
$\boldsymbol{\Lambda}$	diagonal matrix whose ii -th element is λ_i $[N_\theta \times N_\theta]$
$\boldsymbol{\Sigma}$	covariance of measurement error for sample \mathbf{y} $[N_y \times N_y]$
$\boldsymbol{\varphi}$	experimental design vector
$\boldsymbol{\varphi}^*$	D-optimal experimental design vector
$\boldsymbol{\omega}$	column vector of parameters in parameter space Ω $[N_\theta \times 1]$
$\hat{\boldsymbol{\omega}}_P$	column vector of parameter estimates computed with $\mathbf{G} = \mathbf{G}_P$ $[N_\theta \times 1]$
$\hat{\boldsymbol{\omega}}_S$	column vector of parameter estimates computed with $\mathbf{G} = \mathbf{G}_S$ $[N_\theta \times 1]$

Greek symbols

α	statistical significance
θ_i	i -th model parameter
$\hat{\theta}_i$	estimate for the i -th model parameter
Θ	original vector space of model parameters
κ	condition number
λ_i	i -th eigenvalue of \mathbf{H}
ν_i	stoichiometric coefficient of the i -th species
Φ	log-likelihood function
χ_θ^2	χ^2 statistic of target parameters
χ_{ref}^2	95% value computed from a χ^2 distribution
χ_{sample}^2	sum of normalised squared residuals
Ω	transformed vector space of model parameters
∇	gradient operator in parameter space

Acronyms

ED	Experimental Design
HPLC	High-Performance Liquid Chromatograph
MBDoe	Model-Based Design of Experiments
RG	Regularisation
RP	Reparametrisation

References

- Agarwal, A. K., Brisk, M. L., 1985a. Sequential experimental design for precise parameter estimation. 1. Use of reparameterization. Industrial & Engineering Chemistry Process Design and Development 24 (1), 203–207.
- Agarwal, A. K., Brisk, M. L., 1985b. Sequential experimental design for precise parameter estimation. 2. Design criteria. Industrial & Engineering Chemistry Process Design and Development 24 (1), 207–210.
- Alberton, A. L., Schwaab, M., Schmal, M., Pinto, J. C., 2009. Experimental errors in kinetic tests and its influence on the precision of estimated parameters. Part I—Analysis of first-order reactions. Chemical Engineering Journal 155 (3), 816–823.
- Asprey, S. P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Computers & Chemical Engineering 24 (2), 1261–1267.
- Asprey, S. P., Naka, Y., 1999. Mathematical Problems in Fitting Kinetic Models - Some New Perspectives. Journal of Chemical Engineering of Japan 32 (3), 328–337.
- Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press.
- Bardow, A., 2008. Optimal experimental design of ill-posed problems: The METER approach. Computers & Chemical Engineering 32 (1), 115–124.
- Barz, T., Cárdenas, D. C. L., Arellano-Garcia, H., Wozny, G., 2013. Experimental evaluation of an approach to online redesign of experiments for parameter determination. AIChE Journal 59 (6), 1981–1995.
- Barz, T., López C., D. C., Cruz Bournazou, M. N., Körkel, S., Walter, S. F., 2016. Real-time adaptive input design for the determination of competitive adsorption isotherms in liquid chromatography. Computers & Chemical Engineering 94, 104–116.
- Benabbas, L., Asprey, S. P., Macchietto, S., 2005. Curvature-Based Methods for Designing Optimally Informative Experiments in Multiresponse Nonlinear Dynamic Situations. Industrial & Engineering Chemistry Research 44 (18), 7120–7131.
- Berger, R. J., Stitt, E. H., Marin, G. B., Kapteijn, F., Moulijn, J. A., 2001. Eurokin. Chemical Reaction Kinetics in Practice. CATTECH 5 (1), 36–60.

- 710 Bonvin, D., Georgakis, C., Pantelides, C. C., Barolo, M., Grover, M. A., Rodrigues, D.,
Schneider, R., Dochain, D., 2016. Linking Models and Experiments. *Industrial & Engi-
neering Chemistry Research* 55 (25), 6891–6903.
- Bournazou, M. N. C., Barz, T., Nickel, D. B., Cárdenas, D. C. L., Glauche, F., Knepper, A.,
715 Neubauer, P., 2016. Online optimal experimental re-design in robotic parallel fed-batch
cultivation facilities. *Biotechnology and Bioengineering* 114 (3), 610–619.
- Buzzi-Ferraris, G., Manenti, F., 2009. Kinetic models analysis. *Chemical Engineering Science*
64 (5), 1061–1074.
- Chakrabarty, A., Buzzard, G. T., Rundell, A. E., 2013. Model-based design of experiments
720 for cellular processes. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*
5 (2), 181–203.
- Chis, O. T., Banga, J. R., Balsa-Canto, E., 2014. Sloppy models can be identifiable.
arXiv:1403.1417 [q-bio].
- Dovi, V. G., Reverberi, A. P., Acevedo-Duarte, L., 1994. New procedure for optimal design
725 of sequential experiments in kinetic models. *Industrial & Engineering Chemistry Research*
33 (1), 62–68.
- Echtermeyer, A., Amar, Y., Zakrzewski, J., Lapkin, A., 2017. Self-optimisation and model-
based design of experiments for developing a C-H activation flow process. *Beilstein Journal
of Organic Chemistry* 13, 150–163.
- Elliott, C., Vijayakumar, V., Zink, W., Hansen, R., 2007. National Instruments LabVIEW: A
730 Programming Environment for Laboratory Automation and Measurement. *JALA: Journal
of the Association for Laboratory Automation* 12 (1), 17–24.
- Espie, D. M., Macchietto, S., 1988. Nonlinear transformations for parameter estimation.
Industrial & Engineering Chemistry Research 27 (11), 2175–2179.
- Fabry, D. C., Sugiono, E., Rueping, M., 2014. Self-Optimizing Reactor Systems: Algorithms,
735 On-line Analytics, Setups, and Strategies for Accelerating Continuous Flow Process Op-
timization. *Israel Journal of Chemistry* 54 (4), 341–350.
- Fedorov, V. V., 1972. *Theory Of Optimal Experiments*, 1972nd Edition. Academic Press.
- Franceschini, G., Macchietto, S., 2008a. Anti-Correlation Approach to Model-Based Exper-
iment Design: Application to a Biodiesel Production Process. *Industrial & Engineering*
740 *Chemistry Research* 47 (7), 2331–2348.
- Franceschini, G., Macchietto, S., 2008b. Model-based design of experiments for parameter
precision: State of the art. *Chemical Engineering Science* 63 (19), 4846–4872.
- Franceschini, G., Macchietto, S., 2008c. Novel anticorrelation criteria for design of experi-
ments: Algorithm and application. *AIChE Journal* 54 (12), 3221–3238.

- 745 Franceschini, G., Macchietto, S., 2008d. Novel anticorrelation criteria for model-based experiment design: Theory and formulations. *AIChE Journal* 54 (4), 1009–1024.
- Galvanin, F., Barolo, M., Bezzo, F., 2013. On the use of continuous glucose monitoring systems to design optimal clinical tests for the identification of type 1 diabetes models. *Computer Methods and Programs in Biomedicine* 109 (2), 157–170.
- 750 Galvanin, F., Macchietto, S., Bezzo, F., 2007. Model-Based Design of Parallel Experiments. *Industrial & Engineering Chemistry Research* 46 (3).
- Goodell, J. R., McMullen, J. P., Zaborenko, N., Maloney, J. R., Ho, C.-X., Jensen, K. F., Porco, J. A., Beeler, A. B., 2009. Development of an Automated Microfluidic Reaction Platform for Multidimensional Screening: Reaction Discovery Employing Bicyclo[3.2.1]octanoid Scaffolds. *The Journal of Organic Chemistry* 74 (16), 6169–6180.
- 755 Hansen, P. C., 2005. Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM.
- Higham, N. J., 1996. Accuracy and Stability of Numerical Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- 760 Holmes, N., Akien, G. R., Savage, R. J. D., Stanetty, C., Baxendale, I. R., Blacker, A. J., Taylor, B. A., Woodward, R. L., Meadows, R. E., Bourne, R. A., 2016. Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors. *Reaction Chemistry & Engineering* 1 (1), 96–100.
- Hosten, L. H., 1974. A sequential experimental design procedure for precise parameter estimation based upon the shape of the joint confidence region. *Chemical Engineering Science* 29 (11), 2247–2252.
- 765 Huber, P. J., 2004. Robust Statistics. John Wiley & Sons.
- Johansen, T. A., 1997. On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica* 33 (3), 441–446.
- 770 Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python.
- Levenspiel, O., 1998. Chemical Reaction Engineering, 3rd Edition. John Wiley & Sons, New York.
- Lopez C., D. C., Barz, T., Körkel, S., Wozny, G., 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers & Chemical Engineering* 77, 24–42.
- 775 MacKay, D. J. C., 1992. Bayesian methods for adaptive models. phd, California Institute of Technology.

- 780 Maheshwari, V., Rangaiah, G. P., Samavedham, L., 2013. Multiobjective Framework for Model-based Design of Experiments to Improve Parameter Precision and Minimize Parameter Correlation. *Industrial & Engineering Chemistry Research* 52 (24), 8289–8304.
- Malig, T. C., Koenig, J. D. B., Situ, H., Chehal, N. K., Hultin, P. G., Hein, J. E., 2017. Real-time HPLC-MS reaction progress monitoring using an automated analytical platform. *Reaction Chemistry & Engineering* 2 (3), 309–314.
- 785 McMullen, J. P., Jensen, K. F., 2010. An Automated Microfluidic System for Online Optimization in Chemical Synthesis. *Organic Process Research & Development* 14 (5), 1169–1176.
- McMullen, J. P., Jensen, K. F., 2011. Rapid Determination of Reaction Kinetics with an Automated Microfluidic System. *Organic Process Research & Development* 15 (2), 398–
790 407.
- Moore, J. S., Jensen, K. F., 2012. Automated Multitrajectory Method for Reaction Optimization in a Microfluidic System using Online IR Analysis. *Organic Process Research & Development* 16 (8), 1409–1415.
- Oliphant, T. E., 2015. *Guide to NumPy*, 2nd Edition. CreateSpace Independent Publishing
795 Platform, USA.
- Pipus, G., Plazl, I., Koloini, T., 2000. Esterification of benzoic acid in microwave tubular flow reactor. *Chemical Engineering Journal* 76 (3), 239–245.
- Prasad, V., Vlachos, D. G., 2008. Multiscale Model and Informatics-Based Optimal Design of Experiments: Application to the Catalytic Decomposition of Ammonia on Ruthenium.
800 *Industrial & Engineering Chemistry Research* 47 (17), 6555–6567.
- Pritchard, D. J., Bacon, D. W., 1978. Prospects for reducing correlations among parameter estimates in kinetic models. *Chemical Engineering Science* 33 (11), 1539–1543.
- Pukelsheim, F., 2006. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, New York, United States of America.
- 805 Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25 (15), 1923–1929.
- Rimensberger, T., Rippin, D. W. T., 1986. "Sequential experimental design for precise parameter estimation. 1. Use of reparameterization". *Comments. Industrial & Engineering
810 Chemistry Process Design and Development* 25 (4), 1042–1044.
- Rossi, D., Gargiulo, L., Valitov, G., Gavriilidis, A., Mazzei, L., 2017. Experimental characterization of axial dispersion in coiled flow inverters. *Chemical Engineering Research and Design* 120, 159–170.

- 815 Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science* 15 (4), 377–395.
- Schwaab, M., Lemos, L. P., Pinto, J. C., 2008. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 2: Problems involving multiple reparameterizations. *Chemical Engineering Science* 63 (11), 2895–2906.
- 820 Schwaab, M., Pinto, J. C., 2007. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant. *Chemical Engineering Science* 62 (10), 2750–2764.
- Silvey, S. D., 1975. *Statistical Inference*. CRC Press.
- 825 Stamati, I., Logist, F., Akkermans, S., Noriega Fernández, E., Van Impe, J., 2016. On the effect of sampling rate and experimental noise in the discrimination between microbial growth models in the suboptimal temperature range. *Computers & Chemical Engineering* 85, 84–93.
- Transtrum, M. K., Machta, B., Brown, K., Daniels, B. C., Myers, C. R., Sethna, J. P., 2015. Sloppiness and Emergent Theories in Physics, Biology, and Beyond. arXiv:1501.07668.
- 830 Transtrum, M. K., Machta, B. B., Sethna, J. P., 2010. Why are nonlinear fits so challenging? *Physical Review Letters* 104 (6).
- Versyck, K. J., Van Impe, J. F., 1997. On the unicity of optimal experimental design solutions for parameter estimation of microbial kinetics. In: 1997 European Control Conference (ECC). pp. 3509–3514.
- 835 Walpole, R. E., Myers, R. H., Myers, S. L., Ye, K. E., 2011. *Probability and Statistics for Engineers and Scientists*, 9th Edition. Pearson.
- Walsh, B., Hyde, J. R., Licence, P., Poliakov, M., 2005. The automation of continuous reactions in supercritical CO₂: the acid-catalysed etherification of short chain alcohols. *Green Chemistry* 7 (6), 456–463.
- 840 White, A., Tolman, M., Thames, H. D., Withers, H. R., Mason, K. A., Transtrum, M. K., 2016. The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems. *PLOS Computational Biology* 12 (12).
- Wilson, A. D., Schultz, J. A., Murphey, T. D., 2015. Trajectory Optimization for Well-Conditioned Parameter Estimation. *IEEE Transactions on Automation Science and Engineering* 12 (1), 28–36.
- 845 Zhelezov, O. I., 2017. N-dimensional Rotation Matrix Generation Algorithm. *American Journal of Computational and Applied Mathematics* 7 (2), 51–57.

Appendix A

Additional simulated cases

A total number of 20 experimental campaigns were simulated to further validate the results presented in the manuscript. This was done primarily to demonstrate that the performance achieved by the algorithm both in the RP and in the non-RP campaigns is insensitive to the choice of the dataset (i.e. it is insensitive to the choice of the random seed used to generate the experimental data in-silico).

The results obtained in the simulated campaigns are reported in Table A.1. Campaigns 1-10 were performed applying the online reparametrisation method (RP campaigns), while campaigns 11-20 were performed without online reparametrisation (non-RP campaigns). As one can see from Table A.1, the algorithm with online RP option active retrieved the target parameter value in all the campaigns, i.e. the final p -value of the target parameters is above 1.00% in campaigns 1-10. The condition number of the log-likelihood functions at the end of experimental campaigns 1-10 is 1.0, demonstrating that the application of the online RP led to the elimination of the model sloppiness. In the campaigns where the online RP is inactive, i.e. campaigns 11-20, the final p -value is 0.00%, demonstrating the failure of the algorithm in retrieving the target value of the parameters. The failure is associated to the high condition number of the log-likelihood function, which is around $10^3 - 10^4$ in campaigns 11-20.

Table A.1: Results obtained in 20 simulated experimental campaigns: experimental campaigns 1-10 were performed keeping the online reparametrisation option *active*; campaigns 11-20 were performed keeping the option for online reparametrisation *inactive*. The p -value of the target parameters $\theta^* = [15.27, 7.6]$ given the final parameter statistics is reported together with the condition number of the log-likelihood function at the end of the experimental campaigns.

Campaign number	Online reparametrisation	Final p -value of target parameters θ^*	Final condition number κ
1	Active	64.74%	$1.0 \cdot 10^0$
2	Active	98.91%	$1.0 \cdot 10^0$
3	Active	91.98%	$1.0 \cdot 10^0$
4	Active	20.59%	$1.0 \cdot 10^0$
5	Active	30.52%	$1.0 \cdot 10^0$
6	Active	67.93%	$1.0 \cdot 10^0$
7	Active	16.61%	$1.0 \cdot 10^0$
8	Active	92.17%	$1.0 \cdot 10^0$
9	Active	23.19%	$1.0 \cdot 10^0$
10	Active	71.59%	$1.0 \cdot 10^0$
11	Inactive	0.00%	$9.6 \cdot 10^3$
12	Inactive	0.00%	$9.4 \cdot 10^3$
13	Inactive	0.00%	$9.5 \cdot 10^3$
14	Inactive	0.00%	$9.3 \cdot 10^3$
15	Inactive	0.00%	$1.1 \cdot 10^4$
16	Inactive	0.00%	$9.5 \cdot 10^3$
17	Inactive	0.00%	$1.0 \cdot 10^4$
18	Inactive	0.00%	$9.2 \cdot 10^3$
19	Inactive	0.00%	$8.7 \cdot 10^3$
20	Inactive	0.00%	$9.3 \cdot 10^3$

Appendix B

Real case: additional information

Additional details are presented in this appendix regarding the non-RP campaign and the RP campaign performed on the experimental automated system. Information related to the campaign performed keeping the option for online model reparametrisation *inactive*, i.e. the non-RP campaign, is reported in Table B.1. Information on the campaign conducted keeping the option for online reparametrisation *active*, i.e. the RP campaign, is given in Table B.2. In Table B.1 and Table B.2 the following information is presented: 1) experimental conditions adopted to collect the samples, i.e. inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$, flowrate F and temperature T ; 2) sampled concentration of ethyl benzoate at the outlet $C_{\text{EB}}^{\text{OUT}}$; 3) parameter estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$ returned by the model identification algorithm; 4) the pre-exponential factor and activation energy computed from the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ as $A = e^{\hat{\theta}_1}$ and $E_a = 10^4 \cdot \hat{\theta}_2$; 5) the sum of squared residuals χ_{sample}^2 and the reference value χ_{ref}^2 computed from a χ^2 distribution with degree of freedom equal to the number of samples minus the number of parameters and 95% of significance.

A sum of squared residuals χ_{sample}^2 larger than the reference value χ_{ref}^2 is interpreted as an index of inappropriate modelling assumptions (Silvey, 1975). As one can see from Table B.1, the χ_{sample}^2 after the collection of 9 samples in the non-RP campaign is 5.92. From Table B.2, it can be appreciated that the χ_{sample}^2 after the collection of 9 samples in the RP campaign is 1.83. Both in the non-RP and in the RP campaign the χ_{sample}^2 is smaller than the $\chi_{\text{ref}}^2 = 17.88$, thus demonstrating that the modelling assumptions (see Section 3.2) are not falsified by the experimental evidence.

As one can see from Table B.1, the experimental conditions designed by the algorithm for samples 5, 7 and 9 in the non-RP case were similar, i.e. inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, flowrate around $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and temperature $T = 413.0 \text{ K}$, i.e. the upper limit for the temperature. Samples 4, 6 and 8 were instead designed by the algorithm at conditions $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, flowrate $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and temperature in the range $T = 383.0 - 390.0 \text{ K}$. The designed samples in the non-RP case suggest the presence of two optimally informative sets of experimental conditions at maximum temperature $T = 413.0 \text{ K}$ and at temperature around $T = 385.0 \text{ K}$, given that the inlet concentration of benzoic acid $C_{\text{BA}}^{\text{IN}}$ is set at the maximum and that flowrate F is set at the minimum.

An analogous situation can be observed in the RP case. As one can see from Table B.2, samples 4, 7 and 9 in the RP case were designed at conditions $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and $T = 413.0 \text{ K}$. Samples 5, 6 and 8 were instead designed at conditions $C_{\text{BA}}^{\text{IN}} = 1.55 \text{ mol L}^{-1}$, $F = 7.5 \text{ }\mu\text{L min}^{-1}$ and temperature around $T = 391.0 \text{ K}$.

Table B.1: Real case: experimental campaign with online RP option *inactive*. Experimental conditions, sampled concentrations, estimated kinetic parameters $\hat{\Theta}$ (and related Arrhenius constants) and information regarding the goodness-of-fit are reported for the 9 samples collected in the campaign.

Real Case - Online RP Inactive										
Sample	Experimental conditions φ			Sample	Estimates $\hat{\Theta}$		Arrhenius constants ¹		Goodness-of-fit ²	
number	C_{BA}^{IN} [mol L ⁻¹]	F [μ L min ⁻¹]	T [K]	C_{EB}^{OUT} [mol L ⁻¹]	$\hat{\theta}_1$	$\hat{\theta}_2$	A [s ⁻¹]	E_a [J mol ⁻¹ K ⁻¹]	χ_{sample}^2	χ_{ref}^2
1	1.50	20.00	413.0	0.370	-	-	-	-	-	-
2	1.00	10.00	393.0	0.161	-	-	-	-	-	-
3	1.25	15.00	403.0	0.240	16.16	7.94	$1.04 \cdot 10^7$	$7.94 \cdot 10^4$	$4.35 \cdot 10^{-4}$	3.84
4	1.55	7.50	383.0	0.175	16.44	8.03	$1.39 \cdot 10^7$	$8.03 \cdot 10^4$	$2.65 \cdot 10^{-2}$	5.99
5	1.55	7.58	413.0	0.848	17.15	8.26	$2.81 \cdot 10^7$	$8.26 \cdot 10^4$	1.04	7.81
6	1.55	7.50	390.2	0.284	16.80	8.14	$1.98 \cdot 10^7$	$8.14 \cdot 10^4$	1.31	9.49
7	1.55	7.50	413.0	0.876	17.23	8.28	$3.03 \cdot 10^7$	$8.28 \cdot 10^4$	3.56	11.07
8	1.55	7.50	388.5	0.254	17.15	8.26	$2.82 \cdot 10^7$	$8.26 \cdot 10^4$	3.59	12.59
9	1.55	7.50	413.0	0.887	17.42	8.34	$3.69 \cdot 10^7$	$8.34 \cdot 10^4$	5.92	14.07

¹ Pre-exponential factor and activation energy are computed from θ_1 and θ_2 as $A = e^{\theta_1}$ and $E_a = 10^4 \cdot \theta_2$

² A χ_{sample}^2 larger than χ_{ref}^2 is an index of inappropriate modelling assumptions

Table B.2: Real case: experimental campaign with online RP option *active*. Experimental conditions, sampled concentrations, estimated kinetic parameters $\hat{\Theta}$ (and related Arrhenius constants) and information regarding the goodness-of-fit are reported for the 9 samples collected in the campaign.

Real Case - Online RP Active										
Sample	Experimental conditions φ			Sample	Estimates $\hat{\Theta}$		Arrhenius constants ¹		Goodness-of-fit ²	
number	C_{BA}^{IN} [mol L ⁻¹]	F [μ L min ⁻¹]	T [K]	C_{EB}^{OUT} [mol L ⁻¹]	$\hat{\theta}_1$	$\hat{\theta}_2$	A [s ⁻¹]	E_a [J mol ⁻¹ K ⁻¹]	χ_{sample}^2	χ_{ref}^2
1	1.50	20.00	413.0	0.409	-	-	-	-	-	-
2	1.00	10.00	393.0	0.172	-	-	-	-	-	-
3	1.25	15.00	403.0	0.252	17.54	8.37	$4.13 \cdot 10^7$	$8.37 \cdot 10^4$	0.21	3.84
4	1.55	7.50	413.0	0.900	18.12	8.56	$7.39 \cdot 10^7$	$8.56 \cdot 10^4$	0.52	5.99
5	1.55	7.50	392.3	0.346	16.86	8.13	$2.10 \cdot 10^7$	$8.13 \cdot 10^4$	1.27	7.81
6	1.55	7.50	390.6	0.307	16.90	8.15	$2.18 \cdot 10^7$	$8.15 \cdot 10^4$	1.27	9.49
7	1.55	7.50	413.0	0.895	16.91	8.15	$2.20 \cdot 10^7$	$8.15 \cdot 10^4$	1.27	11.07
8	1.55	7.50	391.2	0.323	16.83	8.12	$2.04 \cdot 10^7$	$8.12 \cdot 10^4$	1.31	12.59
9	1.55	7.50	413.0	0.908	16.98	8.17	$2.36 \cdot 10^7$	$8.17 \cdot 10^4$	1.83	14.07

¹ Pre-exponential factor and activation energy are computed from θ_1 and θ_2 as $A = e^{\theta_1}$ and $E_a = 10^4 \cdot \theta_2$

² A χ_{sample}^2 larger than χ_{ref}^2 is an index of inappropriate modelling assumptions