Second-order statistical regression and conditioning of replicate transient kinetic data [★]

Raf Roelant [a], Denis Constales [b], Roger Van Keer [b],
Guy B. Marin [a,*]

[a]*Laboratory for Chemical Technology, Department of Chemical Engineering, Ghent University, Krijgslaan 281 (S5), B-9000 Gent, Belgium*

[b]*Department of Mathematical Analysis, Ghent University, Galglaan 2, B-9000 Gent, Belgium*

Abstract

A new method has been developed to estimate physico-chemical parameters from transient kinetic data: second-order statistical regression (SOSR). It allows to account for heteroskedasticity and nonwhiteness of the noise in the time series measured. SOSR makes use of replicates to estimate the second-order statistics, i.e. the autocovariance pattern of the noise. A sample principal noise component analysis of the experimental time series allows nonlinear least-squares (NLSQ) regression of the latter. The method has been validated by regression of artificially generated experimental data and the results have been compared with those obtained with direct NLSQ regression. The SOSR has also been applied to the irreversible adsorption of oxygen on a reduced vanadia/silica catalyst and the interaction of propane with a copper/ceria catalyst, as studied with a Temporal Analysis of Products (TAP) setup. In general, compared with those obtained with direct NLSQ regression, the parameter estimates and their confidence intervals are more accurate.

Key words: Transient response, Parameter identification, Replicate experiments, Kinetics, Reaction engineering, Temporal Analysis of Products

1 Introduction

Chemical reaction kinetics are a reflection of the reaction mechanism. Therefore, kinetic data are used with advantage to derive features of this reaction mechanism. In this regard, it was shown by several authors that more information can be extracted from transient experiments than from stationary ones, both for non-catalytic (Hulburt & Kim, 1966) and catalytic reaction systems (Tamaru, 1964; Boudart, 1968; Biloen, 1983; Zamostny & Belohlav, 2002). Furusawa et al. (1976), Renken (1993), Efstathiou & Verykios (1997) and Bennett (2000) have reviewed the transient experimental techniques available. Yet another review by Mills & Lerou (1993) was written from the perspective of the industrial catalyst designer. The role of transient experiments becomes crucial if reactions are studied which take place in a transient regime in practice. The industrially most relevant example is fluid catalytic cracking, where the catalyst deactivates on a time scale comparable to that of the cracking itself. This imposes the use of riser reactors. Next to this category of non-steady state operation of catalysts, the existence of oscillations in the feed composition is an obvious reason for transient regimes, an important example being catalytic automotive emission converters, where exhaust gases are oscillatorily fed from a combustion engine.

Nonlinear least-squares (NLSQ) regression of kinetic data is the key to the statistically sound discrimination between rival kinetic models, and hence between the corresponding reaction mechanisms (Froment & Hosten, 1981). NLSQ regression with a certain physico-chemical reaction model yields estimates of the parameters occurring in this model and valuable statistical judgements about their significance in the form of confidence intervals. However, the unbiasedness of the estimates, the validity of the confidence intervals and that of the model discrimination algorithms depend on whether the experimental data obey certain statistical prerequisites which are discussed in section 2 of this paper. All too frequently, these conditions are tacitly assumed fulfilled. However, the assumption that the experimental errors are homoskedastic and uncorrelated are not always valid. The autocorrelation of experimental errors when recording time series is well-known and is often referred to as 'coloured noise'. The necessary conditions unverified, there is a real danger that experimentalists draw unfounded conclusions. They may, for example, conclude that there is significant experimental support for certain mechanistic details of a reaction while in there is not. This paper presents a second-order statistical regression as a way to prevent such mistakes if replicate experimental data are available.

2 Theoretical background

2.1 Experimental time series

The experimentalist controls the conditions at which experiments are performed. For kinetic experiments, these are typically the (inlet) temperature, the inlet concentrations of the reactants and, for reactions involving gases, the (inlet) pressure. As a rule, during transient experiments, at least one of these is forced to undergo a well-defined variation in time. In the steady-state isotopic transient kinetic analysis (SSITKA) for example, an isotopic reactant concentration step is applied to the reactor. In the temperature programmed desorption (TPD), the desorption of products from a catalyst surface is registered in response to a reaction temperature ramp. From now on, the experimental conditions and those parameters characterizing their forced change will be referred to as the experimental conditions in a broad sense. Let them be collected in a column vector $\boldsymbol{\xi}$. In a way dependent on $\boldsymbol{\xi}$, n_v physical variables, V_1, V_2, ..., V_{n_v}, evolve as the experiment progresses. They are recorded by measuring devices, resulting in *time series* of values measured at equally spaced time points t_i: $t_{i+1} - t_i = \Delta t$, for all $i \in \{1, 2, \ldots, n_t\}$. Δt is called the *sampling period* or *sampling interval*. Assume n_e multiresponse experiments are carried out at conditions $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, ... and $\boldsymbol{\xi}_{n_e}$. At each condition $\boldsymbol{\xi}_e$, $e \in \{1, 2, \ldots, n_e\}$, this results in one time series for each physical variable V_v, $v \in \{1, 2, \ldots, n_v\}$:

$$
\mathbf{y}_{e,v} = \begin{pmatrix} y_{e,v}(t_1) \\ y_{e,v}(t_2) \\ \vdots \\ y_{e,v}(t_{n_t}) \end{pmatrix} \in \mathbb{R}^{n_t \times 1}, \tag{1}
$$

represented here as a column vector. Think of these time series as being stacked in a single composite column vector $\mathbf{y} \in \mathbb{R}^{n \times 1}$, with

$$
n = n_e n_v n_t. \tag{2}
$$

Throughout this paper, typical experimental data from a Temporal Analysis of Products (TAP) setup are used for illustration purposes (Gleaves et al., 1988, 1997; Yablonsky et al., 2003). The latter contains a tubular reactor (length typically 3 cm, diameter 5 mm) loaded with a fixed bed of solid catalyst and/or inert particles. Vacuum (10^{-4} to 10^{-5} Pa) is maintained to ensure that gas transport occurs via well-defined Knudsen diffusion. The reactor is isothermal in time and space, with a temperature to be set by the user. At the beginning of each experiment, a reactant pulse having a typical width between 250 and 500 μs is fed at one side of the reactor. A mass spectrometer monitors the flow of products and unconverted reactants leaving the reactor at the other side during typically 1 s. It records a fixed mass of

the spectrum so that unless use is made of a multitrack system (Nijhuis et al., 1997), to observe multiple components, the experiment has to be repeated as least as many times as there are components to be monitored. For TAP, the variables V_1, V_2, ..., V_{n_v} are thus the mass spectrometer measurements, in volt, at the selected spectrum masses. The pressure varies in time and space in a way fully determined by the reactant pulse size. Hence, unless it is unknown, the latter is an experimental condition but the former is not. The reaction temperature is another experimental condition. As an example, Figure 1 shows a typical experimental TAP pulse response.

2.2 Model-calculated time series

The TAP reactor is modelled by a system of partial differential equations with initial and boundary conditions (Gleaves et al., 1988; Constales et al., 2004, 2006). Usually, numerical techniques are applied to integrate the model equations, although an analytical integration is possible in some cases. In general, time series equivalent to those experimentally recorded can be calculated from reactor simulations. As a rule, use has to be made of calibration data of the measurement equipment in a final stage. For TAP for example, the mass spectrum of all products has to be known to be able to convert the product flow pulse responses in moles per second, calculated from the reactor model, to the mass spectrometer signals measured in volt.

The time series calculated from the model depend on some physico-chemical parameters b_1, b_2, ..., b_p (mostly kinetic and transport parameters) appearing in the reactor model. These are collected in a vector $\mathbf{b} \in \mathbb{R}^{p \times 1}$. The model-calculated time series can therefore be represented as

$$\mathbf{f}_v(\boldsymbol{\xi}_e, \mathbf{b}) = \begin{pmatrix} f_v(t_1, \boldsymbol{\xi}_e, \mathbf{b}) \\ f_v(t_2, \boldsymbol{\xi}_e, \mathbf{b}) \\ \vdots \\ f_v(t_{n_t}, \boldsymbol{\xi}_e, \mathbf{b}) \end{pmatrix}, \tag{3}$$

with $e \in \{1, 2, \ldots, n_e\}$ and $v \in \{1, 2, \ldots, n_v\}$. Furthermore, let $\mathbf{f}(\mathbf{b}) \in \mathbb{R}^{n \times 1}$ be the model-calculated analogue of $\mathbf{y}$: a composite of all vectors $\mathbf{f}_v(\boldsymbol{\xi}_e, \mathbf{b})$. In $\mathbf{f}(\mathbf{b})$, $\mathbf{f}$ is the *model function* (Graybill & Iyer, 1994).

2.3 Least-squares regression of time series

Regression analysis is a statistical tool in the hands of the empiricist to investigate the causal relation between certain well-known *independent variables*, possibly controlled by the empiricist, on the one hand, and some observed *dependent*

variables on the other. Regression analysis is useful in those cases where a causal model is known (or assumed valid) except for some parameters occurring in it, such as pre-exponential factors and activation energies in kinetic models. By regression, these parameters are estimated as those that make the model-calculated dependent variables coincide as near as possible with the empirical ones.

By far, the most widely applied regression technique is NLSQ regression. Some recent text books on the matter are (Seber & Lee, 2003; Seber & Wild, 2003) (mathematical) and (Graybill & Iyer, 1994; Draper & Smith, 1998) (applied). If the model-calculated dependent variables are linear functions of the unknown parameters, their estimates can be found in a direct linear-algebraic way. If not, they are to be found iteratively, starting from some initial estimates, which is computationally more expensive. The Levenberg-Marquardt iterative method is commonly used (Levenberg, 1944; Marquardt, 1963).

Hougen & Watson (1947) introduced the least-squares regression approach as a means to estimate kinetic parameters from stationary kinetic experiments. They rearranged the concentration and rate variables in Langmuir-Hinshelwood-Hougen-Watson type rate equations to render them linear in the parameters to be estimated. This allowed them to apply the computationally inexpensive linear least-squares regression theory. With the advent of the digital computer, direct nonlinear regression of the reaction rates was preferred over this method (Kittrell et al., 1965). Cutlip et al. (1972) first applied nonlinear least-squares regression to transient kinetic data.

In transient kinetic experiments, the independent variables are the experimental conditions $\boldsymbol{\xi}_{e,v}$ and the time t. The dependent variables are the experimental recordings $\mathbf{y}_{e,v}$. The physico-chemical parameters $\mathbf{b}$ are to be estimated by nonlinear least-squares regression. Figure 2 gives a schematic overview. The number of parameters that can be estimated is limited to the dimension of the data set $\mathbf{y}$:

$$p \leq n. \tag{4}$$

Strictly speaking, least-squares regression should only be applied under the following six conditions.

Condition 1 *The independent variables contain no experimental errors.*

Condition 2 *The model is* adequate.

Condition 3 *The mean experimental error is zero.*

Condition 4 *All errors are normally distributed.*

Condition 5 *The errors are homoskedastic, which means that their standard deviation is finite and constant.*

Condition 6 *The errors are uncorrelated.*

Condition 1 requires $\boldsymbol{\xi}$ and t to be known without (significant) error. In kinetic studies, the experimental conditions are usually temperature and pressure. These are typically known with very good accuracy and precision. The same is true for the time t. The adequacy of the model referred to in condition 2 means that it applies with some real, unknown parameters $\boldsymbol{\beta}$, aside from additive random experimental errors $\boldsymbol{\varepsilon}$ (Draper & Smith, 1998):

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon}. \tag{5}$$

In fulfilment of the other conditions, $\boldsymbol{\varepsilon}$ has to be a random error vector with multi-dimensional normal distribution with mean zero ($\mathbf{0}_n$) and variance matrix $\mathbf{I}_n \sigma^2$:

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}_n, \mathbf{I}_n \sigma^2\right). \tag{6}$$

If conditions 1 to 6 are fulfilled, the parameters $\hat{\boldsymbol{\beta}}$ minimizing of the residual sum of squares $S(\mathbf{b})$ are the maximum-likelihood estimates of the real parameters $\boldsymbol{\beta}$. (Joint) confidence regions of the parameter estimates $\hat{\boldsymbol{\beta}}$ can be estimated, which allow the experimentalist to judge the significance of the regression and of the individual parameter estimates. Appendix A gives a concise overview of the theory.

Generally, at best a simplifying model can be conceived of the experimental system, failing to grasp all its subtleties. Condition 2 is then fulfilled at most in approximation, and practically translates to the requirement that the bias between experimental and model-calculated time series, so-called *lack of fit* (Graybill & Iyer, 1994; Draper & Smith, 1998), is small compared with the experimental error. In the context of time series the latter consists of *noise*: similarly to $\mathbf{y}$ and $\mathbf{f}(\mathbf{b})$, $\boldsymbol{\varepsilon}$ in (5) is a composite vector containing the noise $\boldsymbol{\varepsilon}_{e,v}$ in all time series $\mathbf{y}_{e,v}$. As is always the case in regression analysis, condition 2 is assumed valid. Condition 3 expresses a basic experimental requirement and is also assumed fulfilled here on that account.

Assumption 1 *The conditions 1, 2 and 3 are fulfilled.*

The fulfilment of the conditions 4, 5 and 6 should not be taken for granted. For TAP pulse responses, at least, they are generally not met. This was demonstrated previously by Roelant et al. (2007) and Schuurman (2007). Condition 5 requires the level of noise to be constant in time, its variance σ^2 unvarying. Moreover, in absolute terms, the noise must be equally important in all experimental time series, σ^2 being independent of the experimental conditions $\boldsymbol{\xi}_e$ and the variable V_v monitored. Fulfilment of condition 6 implies different noise time series to have no cross-correlation. In this paper, the latter is assumed to be valid as a relaxed form of condition 6.

Assumption 2 *Different experimental time series have no cross-correlation.*

6

In particular, assumption 2 involves that the n_v time series measured during each experiment have no cross-correlation. This is the case for most TAP systems, where a single 'experiment' requires a series of n_v independently repeated experiments at the same experimental conditions (see section 2.1). Additionally to assumption 2, condition 6 requires all noise to be *white*, i.e. free of time autocorrelation. This is not assumed to be the case in this paper: the population variance matrix

$$\mathbf{V}(\boldsymbol{\varepsilon}_{e,v}) = E\left[\boldsymbol{\varepsilon}_{e,v}\boldsymbol{\varepsilon}_{e,v}^T\right] \tag{7}$$

need not be diagonal.

This paper will show how the conditions 4, 5 and 6 can be fulfilled. Section 3 attends to the case where $\mathbf{V}(\boldsymbol{\varepsilon}_{e,v})$ is known, a situation rarely met in practice. By contrast, in section 4, $\mathbf{V}(\boldsymbol{\varepsilon}_{e,v})$ is not assumed known.

3 Conditioning data with known noise population variance matrix for least-squares regression: population principal component analysis and rescaling

By its nature, the population variance matrix $\mathbf{V}(\boldsymbol{\varepsilon}_{e,v})$ of the noise $\boldsymbol{\varepsilon}_{e,v}$ in an experimental time series $\mathbf{y}_{e,v}$ is a symmetric, positive definite matrix. Thus, there exists an eigendecomposition

$$\mathbf{V}(\boldsymbol{\varepsilon}_{e,v}) = \mathbf{U}_{e,v}\boldsymbol{\Lambda}_{e,v}\mathbf{U}_{e,v}^T, \tag{8}$$

where $\boldsymbol{\Lambda}_{e,v} \in \mathbb{R}^{n_t \times n_t}$ is a diagonal matrix holding the positive eigenvalues of $\mathbf{V}(\boldsymbol{\varepsilon}_{e,v})$ in its diagonal, ranked from high to low, i.e.

$$\boldsymbol{\Lambda}_{e,v} = \begin{pmatrix} \sigma_{e,v,1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{e,v,2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{e,v,n_t}^2 \end{pmatrix}, \tag{9}$$

with

$$\sigma_{e,v,1} \geqslant \sigma_{e,v,2} \geqslant \ldots \geqslant \sigma_{e,v,n_t} > 0 \tag{10}$$

and $\mathbf{U}_{e,v}$ is a $n_t \times n_t$ matrix having the associated, orthonormal eigenvectors as columns:

$$\mathbf{U}_{e,v} = \begin{pmatrix} | & | & & | \\ \mathbf{u}_{e,v,1} & \mathbf{u}_{e,v,2} & \cdots & \mathbf{u}_{e,v,n_t} \\ | & | & & | \end{pmatrix}, \tag{11}$$

with

$$\mathbf{u}_{e,v,k}^T\mathbf{u}_{e,v,l} = \begin{cases} 1, \text{ if } k = l \\ 0, \text{ else.} \end{cases} \tag{12}$$

7

In matrix form, the orthonormality (12) of the eigenvectors translates to

$$\mathbf{U}_{e,v}^T \mathbf{U}_{e,v} = \mathbf{I}_{n_t}. \tag{13}$$

Consider the linear transformation of time series $\mathbf{y}_{e,v}$ by $\mathbf{U}_{e,v}^T$:

$$\mathbf{y}'_{e,v} = \mathbf{U}_{e,v}^T \mathbf{y}_{e,v}. \tag{14}$$

The random part $\boldsymbol{\varepsilon}_{e,v}$ of $\mathbf{y}_{e,v}$ is hereby transformed to the random part $\boldsymbol{\varepsilon}'_{e,v}$ of $\mathbf{y}'_{e,v}$:

$$\boldsymbol{\varepsilon}'_{e,v} = \mathbf{U}_{e,v}^T \boldsymbol{\varepsilon}_{e,v}. \tag{15}$$

The variance matrix of $\boldsymbol{\varepsilon}'_{e,v}$ is found as

$$
\begin{aligned}
\mathbf{V}(\boldsymbol{\varepsilon}'_{e,v}) &= E\left[\boldsymbol{\varepsilon}'_{e,v} {\boldsymbol{\varepsilon}'_{e,v}}^T \right] \\
&\overset{(15)}{=} E\left[\mathbf{U}_{e,v}^T \boldsymbol{\varepsilon}_{e,v} \boldsymbol{\varepsilon}_{e,v}^T \mathbf{U}_{e,v} \right] \\
&\overset{(7)}{=} \mathbf{U}_{e,v}^T \mathbf{V}(\boldsymbol{\varepsilon}_{e,v}) \mathbf{U}_{e,v} \\
&\overset{(8)}{=} \mathbf{U}_{e,v}^T \mathbf{U}_{e,v} \boldsymbol{\Lambda}_{e,v} \mathbf{U}_{e,v}^T \mathbf{U}_{e,v} \\
&\overset{(13)}{=} \boldsymbol{\Lambda}_{e,v}.
\end{aligned}
\tag{16}
$$

$$\tag{17}$$

$\boldsymbol{\varepsilon}'_{e,v}$ is the result of expressing $\boldsymbol{\varepsilon}_{e,v}$ in a coordinate system constituted by the orthonormal eigenvectors of $\mathbf{V}(\boldsymbol{\varepsilon}'_{e,v})$. In other words, for each k, the kth component of $\boldsymbol{\varepsilon}'_{e,v}$ is the euclidean length of the orthogonal projection of the random noise vector $\boldsymbol{\varepsilon}_{e,v}$ on $\mathbf{u}_{e,v,k}$. According to (17), (9) and (10), these components are mutually uncorrelated and have variances decreasing as a function of k. Apparently, the largest variability of the noise is parallel to the first eigenvector $\mathbf{u}_{e,v,1}$, the second largest variability parallel to the second eigenvector $\mathbf{u}_{e,v,2}$, etc. For this reason, the eigenvectors, and particularly the first few of them, are called the *population principal components* (pPCs) of the noise. The linear transformation (15) is called the *population principal component analysis* (pPCA) or the *discrete Karhunen-Loève transformation* of the noise (Therrien, 1992; Massart et al., 1997; Vandeginste et al., 1998; Jolliffe, 2004). In fulfilment of condition 6 for least-squares regression, transformation (14) projects the time series on the pPCs of their noise. Simple rescaling of the components of the resulting vector $\mathbf{y}'_{e,v}$ by the reciprocal of their standard deviation suffices to also render their errors homoskedastic, in fulfilment of condition 5. Define

$$
\begin{aligned}
\mathbf{y}''_{e,v} &= \boldsymbol{\Lambda}_{e,v}^{-1/2} \mathbf{y}'_{e,v} \\
&= \boldsymbol{\Lambda}_{e,v}^{-1/2} \mathbf{U}_{e,v}^T \mathbf{y}_{e,v}
\end{aligned}
\tag{18}
$$

with

$$\Lambda_{e,v}^{-1/2} = \begin{pmatrix} 1/\sigma_{e,v,1} & 0 & \cdots & 0 \\ 0 & 1/\sigma_{e,v,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_{e,v,n_t} \end{pmatrix}. \tag{19}$$

Then, taking into account that $\Lambda_{e,v}^{-1/2}$ is symmetric, the variance matrix of the random part

$$\boldsymbol{\varepsilon}_{e,v}'' = \Lambda_{e,v}^{-1/2}\boldsymbol{\varepsilon}_{e,v}' \tag{20}$$

of $\mathbf{y}_{e,v}''$ becomes

$$\begin{aligned} \mathbf{V}(\boldsymbol{\varepsilon}_{e,v}'') &= E\left[\boldsymbol{\varepsilon}_{e,v}''\boldsymbol{\varepsilon}_{e,v}''^T\right] \\ &\overset{(20)}{=} E\left[\Lambda_{e,v}^{-1/2}\boldsymbol{\varepsilon}_{e,v}'\boldsymbol{\varepsilon}_{e,v}'^T\Lambda_{e,v}^{-1/2}\right] \\ &\overset{(16)}{=} \Lambda_{e,v}^{-1/2}\mathbf{V}(\boldsymbol{\varepsilon}_{e,v}')\Lambda_{e,v}^{-1/2} \\ &\overset{(17)}{=} \Lambda_{e,v}^{-1/2}\Lambda_{e,v}\Lambda_{e,v}^{-1/2} \\ &= \mathbf{I}_{n_t}. \end{aligned} \tag{21}$$

Assumption 2 implies that there is no cross correlation between the different transformed noise vectors $\boldsymbol{\varepsilon}_{e,v}''$. Hence, it follows from (21) that the variance matrix of their composite vector $\boldsymbol{\varepsilon}''$ is

$$\mathbf{V}(\boldsymbol{\varepsilon}'') = \mathbf{I}_n. \tag{22}$$

This fulfils conditions 5 and 6 for regression of $\mathbf{y}''$ with $\mathbf{f}''(\mathbf{b})$, the composite vectors of $\mathbf{y}_{e,v}''$ and

$$\mathbf{f}_v''(\boldsymbol{\xi}_e, \mathbf{b}) = \Lambda_{e,v}^{-1/2}\mathbf{U}_{e,v}^T\mathbf{f}_v(\boldsymbol{\xi}_e, \mathbf{b}), \tag{23}$$

respectively. The fulfilment of the conditions 1, 2 and 3 (see assumption 1) holds. However, condition 4 generally remains unfulfilled. Moreover, observe that the matrices $\Lambda_{e,v}$ and $\mathbf{U}_{e,v}$ were calculated from the noise's population variance matrix $\mathbf{V}(\boldsymbol{\varepsilon}_{e,v})$, while this matrix is usually not known. These problems can be solved making use of replicate experiments, as will be explained in the sections 4.1 and 4.2.

4 Conditioning replicate experimental data with unknown noise variance matrix for least-squares regression: second-order statistical regression

4.1 Sample principal component analysis and rescaling

It is often possible to replicate transient experiments at low additional cost. The replicate time series recorded must be stochastically homogeneous and mutually

independent. Therefore, for catalytic experiments, the experimentalist has to make sure that any changes in the catalyst's surface between the replicate experiments remain negligible. The availability of replicate experiments offers three advantages.

(1) Replicate time series can be averaged to obtain time series with an increased signal-to-noise ratio. Regression of such averages thus yields more significant information.
(2) The noise of the average of, say, minimum ten time series can be considered normally distributed in good approximation, whatever the distribution of the original noise, because of the central limit theorem. This fulfils condition 4 for least-squares regression of the average time series.
(3) From replicates, second-order statistical information (the variance matrix) can be derived about the noise. This information can be used to transform linearly the time series to fulfil the conditions 5 and 6. An approach similar to the one presented in section 3 can be followed.

Advantage 2 involves an assumption.

Assumption 3 *A sufficient amount of replicate time series is available to consider the noise of the average time series normally distributed in good approximation.*

Assume there are $r_e \geq 2$ replicates of each experimental time series $\mathbf{y}_{e,v}$: $\mathbf{y}_{e,v}^{(1)}$, $\mathbf{y}_{e,v}^{(2)}$, $\ldots$, $\mathbf{y}_{e,v}^{(r_e)}$. Not all experiments have to be replicated an equal amount of times, which is why a subscript e is provided in r_e. Define $\bar{\mathbf{y}}_{e,v}$ as the average time series and the error matrix $\mathbf{E}_{e,v} \in \mathbb{R}^{n_t \times r_e}$ as the block matrix

$$
\mathbf{E}_{e,v} = \begin{pmatrix} \vline & \vline & & \vline \\ \mathbf{y}_{e,v}^{(1)} - \bar{\mathbf{y}}_{e,v} & \mathbf{y}_{e,v}^{(2)} - \bar{\mathbf{y}}_{e,v} & \cdots & \mathbf{y}_{e,v}^{(r_e)} - \bar{\mathbf{y}}_{e,v} \\ \vline & \vline & & \vline \end{pmatrix} \tag{24}
$$

Then the population variance matrix $\mathbf{V}(\boldsymbol{\varepsilon}_{e,v})$ of the noise $\boldsymbol{\varepsilon}_{e,v}$ in the time series $\mathbf{y}_{e,v}$ can be estimated from the replicates as the sample variance matrix

$$
\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v}) = \frac{1}{r_e - 1} \mathbf{E}_{e,v} \mathbf{E}_{e,v}^T. \tag{25}
$$

Normally, the following assumption applies.

Assumption 4 *The number of replicate experiments does not exceed the number of samples in one time series:*
$$
r_e \leq n_t. \tag{26}
$$

In the case of TAP for example, r_e would typically be 20 while n_t would be 1000.

By construction and assumption 4, $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v}) \in \mathbb{R}^{n_t \times n_t}$ is a symmetric, positive-semi-definite matrix of rank $r_e - 1$. Thus, there exists an eigendecomposition

$$\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v}) = \hat{\mathbf{U}}_{e,v}\hat{\boldsymbol{\Lambda}}_{e,v}\hat{\mathbf{U}}_{e,v}^T, \tag{27}$$

where $\hat{\boldsymbol{\Lambda}}_{e,v} \in \mathbb{R}^{(r_e-1) \times (r_e-1)}$ is a diagonal matrix with the $r_e - 1$ nonzero eigenvalues of $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v})$ as diagonal elements, ranked from high to low, i.e.

$$\hat{\boldsymbol{\Lambda}}_{e,v} = \begin{pmatrix} s_{e,v,1}^2 & 0 & \cdots & 0 \\ 0 & s_{e,v,2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{e,v,r_e-1}^2 \end{pmatrix}, \tag{28}$$

with

$$s_{e,v,1} \geqslant s_{e,v,2} \geqslant \ldots \geqslant s_{e,v,r_e-1} > 0 \tag{29}$$

and $\hat{\mathbf{U}}_{e,v}$ is a $n_t \times (r_e - 1)$ matrix with the associated eigenvectors as columns:

$$\hat{\mathbf{U}}_{e,v} = \begin{pmatrix} | & | & & | \\ \hat{\mathbf{u}}_{e,v,1} & \hat{\mathbf{u}}_{e,v,2} & \cdots & \hat{\mathbf{u}}_{e,v,r_e-1} \\ | & | & & | \end{pmatrix}. \tag{30}$$

The eigenvectors are orthonormal:

$$\hat{\mathbf{u}}_{e,v,k}^T\hat{\mathbf{u}}_{e,v,l} = \begin{cases} 1, \text{ if } k = l \\ 0, \text{ else.} \end{cases} \tag{31}$$

In matrix form:

$$\hat{\mathbf{U}}_{e,v}^T\hat{\mathbf{U}}_{e,v} = \mathbf{I}_{r_e-1}. \tag{32}$$

Analogously to (14), say

$$\mathbf{y}_{e,v}' = \hat{\mathbf{U}}_{e,v}^T\mathbf{y}_{e,v}. \tag{33}$$

$\mathbf{y}_{e,v}'$ is $\mathbf{y}_{e,v}$ orthogonally projected on the space spanned by the eigenvectors of $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v})$ corresponding to nonzero eigenvectors, expressed in the orthonormal coordinate system they constitute. Observe that the transformed replicates of $\mathbf{y}_{e,v}$ can be considered replicates themselves of $\mathbf{y}_{e,v}'$. For each $k \in \{1, 2, \ldots, r_e\}$, define

$$\mathbf{y}_{e,v}'^{(k)} = \hat{\mathbf{U}}_{e,v}^T\mathbf{y}_{e,v}^{(k)}. \tag{34}$$

Moreover the average of the transforms is equal to the transformed average, $\overline{\mathbf{y}}'_{e,v}$. Corresponding to (24), define the error matrix

$$\mathbf{E}'_{e,v} = \begin{pmatrix} | & | & & | \\ \mathbf{y}'^{(1)}_{e,v} - \overline{\mathbf{y}}'_{e,v} & \mathbf{y}'^{(2)}_{e,v} - \overline{\mathbf{y}}'_{e,v} & \cdots & \mathbf{y}'^{(r_e)}_{e,v} - \overline{\mathbf{y}}'_{e,v} \\ | & | & & | \end{pmatrix} \tag{35}$$

$$= \hat{\mathbf{U}}^T_{e,v} \mathbf{E}_{e,v}. \tag{36}$$

The variance matrix $\mathbf{V}(\boldsymbol{\varepsilon}'_{e,v})$ is estimated as

$$
\begin{aligned}
\hat{\mathbf{V}}(\boldsymbol{\varepsilon}'_{e,v}) &= \frac{1}{r_e - 1} \mathbf{E}'_{e,v} {\mathbf{E}'_{e,v}}^T \\
&\stackrel{(36)}{=} \frac{1}{r_e - 1} \hat{\mathbf{U}}^T_{e,v} \mathbf{E}_{e,v} \mathbf{E}^T_{e,v} \hat{\mathbf{U}}_{e,v} \\
&\stackrel{(25)}{=} \hat{\mathbf{U}}^T_{e,v} \hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v}) \hat{\mathbf{U}}_{e,v} \\
&\stackrel{(27)}{=} \hat{\mathbf{U}}^T_{e,v} \hat{\mathbf{U}}_{e,v} \hat{\boldsymbol{\Lambda}}_{e,v} \hat{\mathbf{U}}^T_{e,v} \hat{\mathbf{U}}_{e,v} \\
&\stackrel{(32)}{=} \hat{\boldsymbol{\Lambda}}_{e,v}.
\end{aligned} \tag{37}
$$

As this matrix is diagonal, the transformation is expected to fulfil condition 6 for least-squares regression. The eigenvectors $\hat{\mathbf{u}}_{e,v,1}$, $\hat{\mathbf{u}}_{e,v,2}$, …, $\hat{\mathbf{u}}_{e,v,r_e-1}$ are called *sample principal components* (sPCs) of the noise. Transformation (36) is analogous to the pPCA (15) and is referred to as sample principal component analysis (sPCA). As before, rescaling enables to meet condition 5. Define

$$
\begin{aligned}
\mathbf{y}''_{e,v} &= \hat{\boldsymbol{\Lambda}}^{-1/2}_{e,v} \mathbf{y}'_{e,v} \\
&= \hat{\boldsymbol{\Lambda}}^{-1/2}_{e,v} \hat{\mathbf{U}}^T_{e,v} \mathbf{y}_{e,v},
\end{aligned} \tag{38}
$$

where $\hat{\boldsymbol{\Lambda}}^{-1/2}_{e,v}$ is defined analogously to $\boldsymbol{\Lambda}^{-1/2}_{e,v}$ in (19). Then the sample variance matrix of the error $\boldsymbol{\varepsilon}''_{e,v}$ of $\mathbf{y}''_{e,v}$ is the unit matrix:

$$\hat{\mathbf{V}}\left(\boldsymbol{\varepsilon}''_{e,v}\right) = \mathbf{I}_{r_e-1}. \tag{39}$$

The sample variance matrix $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}''_{e,v})$ is an estimate of the unknown population variance matrix $\mathbf{V}(\boldsymbol{\varepsilon}''_{e,v})$ based on the replicate data available. While transformation (38) was constructed to equate the first with the unit matrix $\mathbf{I}_{r_e-1}$, the latter is close but probably not equal to $\mathbf{I}_{r_e-1}$:

$$\mathbf{V}(\boldsymbol{\varepsilon}''_{e,v}) \approx \mathbf{I}_{r_e-1}. \tag{40}$$

Correspondingly, the components of $\boldsymbol{\varepsilon}''_{e,v}$ are neither fully uncorrelated nor perfectly homoskedastic. However, the approximate validity of (40) improves as more replicates become available.

As a consequence of (39), the sample variance matrix of the error vector $\bar{\boldsymbol{\varepsilon}}_{e,v}''$ of the transformed average time series, $\bar{\mathbf{y}}_{e,v}'' = \hat{\boldsymbol{\Lambda}}_{e,v}^{-1/2}\hat{\mathbf{U}}_{e,v}^T\bar{\mathbf{y}}_{e,v}$, is calculated as

$$\hat{\mathbf{V}}\left(\bar{\boldsymbol{\varepsilon}}_{e,v}''\right) = \frac{1}{r_e - 1}\mathbf{I}_{r_e - 1}. \tag{41}$$

A subsequent rescaling

$$\begin{aligned}\bar{\mathbf{y}}_{e,v}''' &= \sqrt{r_e - 1}\,\bar{\mathbf{y}}_{e,v}'' \\ &= \sqrt{r_e - 1}\,\hat{\boldsymbol{\Lambda}}_{e,v}^{-1/2}\hat{\mathbf{U}}_{e,v}^T\bar{\mathbf{y}}_{e,v}\end{aligned} \tag{42}$$

enables to obtain

$$\hat{\mathbf{V}}\left(\bar{\boldsymbol{\varepsilon}}_{e,v}'''\right) = \mathbf{I}_{r_e - 1}, \tag{43}$$

where $\bar{\boldsymbol{\varepsilon}}_{e,v}'''$ is the random part of $\bar{\mathbf{y}}_{e,v}'''$. All transformed average time series $\bar{\mathbf{y}}_{e,v}''' \in \mathbb{R}^{(r_e - 1)\times 1}$ can then be stacked in a single vector $\bar{\mathbf{y}}''' \in \mathbb{R}^{n'''\times 1}$, with n''' the total number of replicate experimental times series,

$$n''' = n_v \sum_{e=1}^{n_e}(r_e - 1). \tag{44}$$

It follows from assumption 2 and equation (43) that the sample variance matrix of the random part $\bar{\boldsymbol{\varepsilon}}'''$ of $\bar{\mathbf{y}}'''$ is calculated as

$$\hat{\mathbf{V}}(\bar{\boldsymbol{\varepsilon}}''') = \mathbf{I}_{n'''}. \tag{45}$$

$\bar{\mathbf{y}}'''$ can now be regressed with its model-calculated analogue $\mathbf{f}'''(\mathbf{b})$, a composite vector of transformed model-calculated time series $\mathbf{f}_v'''(\boldsymbol{\xi}_e, \mathbf{b})$:

$$\mathbf{f}_v'''(\boldsymbol{\xi}_e, \mathbf{b}) = \sqrt{r_e - 1}\,\hat{\boldsymbol{\Lambda}}_{e,v}^{-1/2}\hat{\mathbf{U}}_{e,v}^T\mathbf{f}_v(\boldsymbol{\xi}_e, \mathbf{b}). \tag{46}$$

Conditions 5 and 6 are expected to be fulfilled by (45). Moreover, as the transformation presented is linear, the fulfilment of the conditions 2, 3 and 4, assured by the assumptions 1 and 3, holds. In summary, $\bar{\mathbf{y}}'''$ is provisionally found suited for regression with $\mathbf{f}'''(\mathbf{b})$. Obviously, this regression is only possible if the number of parameters is not larger than the dimensionality of these vectors. This is assumed in this paper.

Assumption 5 *The number of parameters does not exceed the total number of replicate experimental time series:*

$$p \le n'''. \tag{47}$$

Taking into account equations (2), (44) and (26), comparison of (47) with (4) teaches that by application of the sPCA (33), the number of parameters that can be estimated was reduced. In reality, rarely more than, say, ten parameters have

to be estimated. Keeping in mind assumption 3, the limitation is unlikely to raise problems in practice. The dimensionality reduction would decrease the computational load of the regression, were it not that the most time-consuming part is the evaluation of the model function $\mathbf{f}$, which generally requires the numerical integration of one or more systems of partial differential equations. On the other hand, because $\mathbf{y}'''$ and $\mathbf{f}'''(\mathbf{b})$ have less components than their originals, the transformation is noninvertible. Hence, a part of the information is lost for regression. This can lead to an unnecessary increase of the width of the parameter confidence intervals. Preferably, $\mathbf{y}$ is well aligned with the vector space $\mathscr{S}$ spanned by all sPCs. In that case, the orthonormal coordinate system they constitute is able to describe $\mathbf{y}$ well, $\mathbf{y}$'s orthogonal projection on $\mathscr{S}$ not differing so much from its original. The angle θ between $\mathbf{y}$ and the space $\mathscr{S}$ can be calculated from

$$\cos \theta = \frac{\|\bar{\mathbf{y}}'\|}{\|\bar{\mathbf{y}}\|}, \tag{48}$$

where $\bar{\mathbf{y}}'$ is the composite of all vectors $\bar{\mathbf{y}}'_{e,v}$. θ is a measure for the distortion of the experimental set by switching to a smaller coordinate system at sPCA. In other words, θ is a rough measure of the fraction of information lost. In the most favourable case, evidently, $\theta \approx 0°$.

Figure 3 visualizes the first four sPCs calculated from a set of eight hundred replicate TAP pulse responses, one of which was shown in Figure 1. Figure 4 shows the sample standard deviations along the first few sPCs. The data used stem from argon pulsed over quartz at room temperature. Hence, the experiment was non-reactive, so that the reactor bed was not affected by the gas pulses. This made it possible to collect a number of replicates as large as eight hundred (see remark on catalytic experiments at the beginning of this section). Figures 3 and 4 also show sPCs and sample standard deviations calculated from a reduced set of 20 replicates, a realistic number. The angle θ calculated for the full set of 800 replicates was found to be $1.3°$. This would perhaps be acceptable, but for the reduced set of 20 replicates, $\theta = 53°$. This is by all means unacceptable. Section 4.2 will show how preconditioning the data enables reduction of the information loss.

4.2 Data preconditioning

On the basis of replicates, section 4.1 presented the sPCA as a way to decorrelate the experimental data. This was at the expense of a dimensionality reduction. The coordinate system constituted by the sPCs $\hat{\mathbf{u}}_{e,v,1}$, $\hat{\mathbf{u}}_{e,v,2}$, $\ldots$, $\hat{\mathbf{u}}_{e,v,r_e-1}$ was found unable to describe the time series well, resulting in a serious loss of experimental information at sPCA. Compare the sPCs calculated for a set of replicate TAP pulse responses, shown in Figure 3, with one of these responses depicted in Figure 1. While the pulse response shows a relatively slow variation, the sPCs show very fast random fluctuations. This is probably a general feature. It is no surprise that the

coordinate system constituted by the sPCs is unable to give a good representation of the (average) pulse response. A better result would be obtained if the sPCs were smoother. This can be achieved by transforming the original time series so that their noise, although still random, becomes smoother. To this end, the experimental time series will be submitted to an extra conditioning transformation before least-squares regression.

Observe that *any* linear transformation preserves the fulfilment of the conditions 2, 3 and 4 (assumptions 1 and 3). It is therefore allowed to transform linearly the time series, before performing a sPCA and rescaling in fulfilment of conditions 5 and 6. Indeed, say $\mathbf{T}$ is a nonsingular $n_t \times n_t$ matrix. Then $\mathbf{T}\mathbf{y}_{e,v}$ contains all the information of $\mathbf{y}_{e,v}$. Instead of regressing the $\overline{\mathbf{y}}_{e,v}'''$, defined by (42), with $\mathbf{f}_v'''(\boldsymbol{\xi}_e, \mathbf{b})$ from (46), vectors

$$\overline{\mathbf{y}}_{e,v,\mathbf{T}}''' = \sqrt{r_e - 1}\,\hat{\boldsymbol{\Lambda}}_{e,v,\mathbf{T}}^{-1/2}\hat{\mathbf{U}}_{e,v,\mathbf{T}}^T \mathbf{T}\overline{\mathbf{y}}_{e,v} \tag{49}$$

can just as well be regressed with

$$\mathbf{f}_{v,\mathbf{T}}'''(\boldsymbol{\xi}_e, \mathbf{b}) = \sqrt{r_e - 1}\,\hat{\boldsymbol{\Lambda}}_{e,v,\mathbf{T}}^{-1/2}\hat{\mathbf{U}}_{e,v,\mathbf{T}}^T \mathbf{T}\mathbf{f}_v(\boldsymbol{\xi}_e, \mathbf{b}). \tag{50}$$

In the latter two expressions, a subscript $\mathbf{T}$ has been added to the matrices $\hat{\mathbf{U}}_{e,v}$ and $\hat{\boldsymbol{\Lambda}}_{e,v}$ to indicate that an eigendecomposition is performed of $\hat{\mathbf{V}}(\mathbf{T}\boldsymbol{\varepsilon}_{e,v})$ instead of $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v})$, so that $\hat{\mathbf{U}}_{e,v,\mathbf{T}}$ and $\hat{\boldsymbol{\Lambda}}_{e,v,\mathbf{T}}$ indeed depend on $\mathbf{T}$. Many matrices $\mathbf{T}$ have the desired smoothing effect on the noise. As an example, the lower triangular matrix,

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \Delta t, \tag{51}$$

involves the replacement of each sample $y(t_i)$ in the time series by the Riemann sum $\sum_{j=1}^{i} y(t_j)\Delta t$ of the preceding samples and itself.

Figure 5 shows the average $\overline{\mathbf{y}}$ of the eight hundred Ar pulse responses used in section 4.1, and its Riemann cumulative $\mathbf{T}\overline{\mathbf{y}}$. Still for the same set of experimental data, Figure 6, analogous to Figure 3, visualizes the first four sPCs of the transformed pulse responses. Figure 7, analogous to Figure 4, shows the sample standard deviations along the first few sPCs. As desired, the sPCs are smoother. This was explained intuitively at the beginning of this section, but there is also a mathematical explanation. The linear transformation $\mathbf{T}$ renders the unknown population variance matrix near-singular in such a way that the population standard deviations along the latter $n_t - (r_e - 1)$ pPCs, $\sigma_{r_e}, \sigma_{r_e+1}, \ldots, \sigma_{n_t}$, become negligible. As a consequence, the sPCs and the corresponding standard deviations can be considered nearly unbiased estimates of the pPCs and the corresponding standard deviations. The pPCs are generally smooth and hence so are their estimates, the sPCs, apart

15

from their random part. That the sPCs and the corresponding standard deviations indeed evolve to the smooth pPCs and the corresponding standard deviations is apparent in Figure 6 and Figure 7. As intended, the relative smoothness of the sPCs causes a better alignment of the space $\mathscr{S}$ they span and the experimental data $\mathbf{y}$. This is illustrated by the angle θ which is $0.0022°$ for the complete data set of eight hundred replicates and still $0.47°$ for the reduced set of twenty replicates.

The regression of $\bar{\mathbf{y}}_{\mathbf{T}}'''$, composite vector of the $\bar{\mathbf{y}}_{e,v,\mathbf{T}}'''$ with $\mathbf{f}_{\mathbf{T}}'''(\mathbf{b})$, composite of the $\mathbf{f}_{v,\mathbf{T}}'''(\boldsymbol{\xi}_e, \mathbf{b})$ with a suitable $\mathbf{T}$ will be called *second-order statistical regression* (SOSR). Figure 8 gives a summarizing overview.

Figure 7 illustrates that the sample standard deviation along the successive sPCs drops to zero very quickly. This means that the first few of them already account for most of the variability of the noise. In some cases, a physical meaning can be attributed to one or more of the principal components. For example, comparison of Figure 6 (a) with Figure 5 shows that the first sPC, in particular, assumes the shape of the Riemann-cumulated pulse response itself. Accordingly, the first sPC shown in Figure 3 (a), assumes the shape of the pulse response itself. These observations show that the most important variability in the pulse responses is in their size and not so much in their shape. This is rooted in the experimental hardware, the size of the TAP inlet pulse being poorly reproducible. The pulse size is proportional to the Riemann sum of the average pulse response, which is

$$S_R = \Delta t \cdot \sum_{j=1}^{n_t} (\bar{\mathbf{y}})_j = (\mathbf{T}\bar{\mathbf{y}})_{n_t}, \tag{52}$$

where $(\mathbf{T}\bar{\mathbf{y}})_{n_t}$ is the n_tth component of vector $\mathbf{T}\bar{\mathbf{y}}$ with $\mathbf{T}$ chosen as in (51). S_R's standard deviation relates to the standard deviation $s_{\mathbf{T},1}$ corresponding to the first sPC as

$$s_{S_R} = s_{\mathbf{T},1}(\hat{\mathbf{u}}_{\mathbf{T},1})_{n_t}. \tag{53}$$

The relative error of S_R and therefore also of the pulse size is quantified here as

$$\frac{s_{S_R}}{S_R} \approx 1.13\% \tag{54}$$

using (52) and (53) for the full set of eight hundred replicates. This is a typical number (Gleaves et al., 1997; Roelant et al., 2007).

4.3   *Practical implementation of the SOSR*

Figure 2 provided a schematic overview of how physico-chemical parameters can be estimated by NLSQ regression of experimental time series. It contains two computational operations: the simulation and the NLSQ regression. For TAP, the simulation and regression were carried out numerically as outlined by van der Linde et

16

al. (1997), making use of the Livermore solver for ordinary differential equations (LSODE) from ODEPACK (Hindmarsh, 1983) for the simulation and ODRPACK (in ordinary least-squares mode) for the regression (Boggs et al., 1987).

Comparison of Figure 8 with Figure 2 shows that the difference between SOSR and NLSQ regression is the application of a three-part conditioning transformation before the regression stage: preconditioning, sPCA and rescaling. Here, the preconditioning transformation used was Riemann cumulation. However, other smoothing operations would work as well. The subsequent sPCA and rescaling are based on a once-only calculation of the nonzero eigenvalues and corresponding -vectors of the sample variance matrices $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v})$, calculated from replicate data as in (25). Explicit calculation of $\hat{\mathbf{V}}(\boldsymbol{\varepsilon}_{e,v})$ is not necessary, as the eigenvalues and -vectors can be found directly from the computationally efficient *singular value decomposition* of the error matrices $\mathbf{E}_{e,v}$. The reader is referred to (Therrien, 1992), (Vandeginste et al., 1998) or (Jolliffe, 2004) for details. Here, the DGESVD routine from LAPACK was used. The main program was written in Fortran 95.

5   Validation of the second-order statistical regression

5.1   *Numerical experiment*

The theory presented above was tested with a numerical experiment. A realistic TAP experiment was simulated with the model of Constales et al. (2004). An unspecified inert species A was thought pulsed in a 3 cm long TAP reactor filled with an inert packing at a constant, unspecified temperature, $n_e = 1$. As only species A is to be monitored: $n_v = 1$. The only model parameter is the effective Knudsen diffusivity $\mathscr{D}_e$ of A through the reactor tube. A simulation was carried out with a realistic value of $2 \cdot 10^{-3}\,\mathrm{m}^2/\mathrm{s}$ for $\mathscr{D}_e$. The A pulse response calculated from this simulation, was hereupon subjected four hundred times to artificial TAP noise with some typical characteristics, see (Roelant et al., 2007):

(1) A normally distributed variability of the pulse size with relative standard deviation of 5 %.
(2) Oscillatory noise with a frequency of 50 Hz and a time-varying amplitude of 2 % of the calculated response signal strength. For each pulse response, the initial phase was chosen between 0 and 2π in a uniformly random way.
(3) Ornstein-Uhlenbeck noise, also called Gauss-Markov noise, with a correlation time of 0.789 ms. The standard deviation was taken partly constant (0.5 % of the response maximum), and partly proportional (5 %) to the signal strength.

The set of four hundred pulse responses was divided in twenty groups of twenty. Each of these were used to estimate $\mathscr{D}_e$ by regression. This was done twice each

time: once by NLSQ regression of the average pulse response and once by SOSR. Conditions 1, 2 and 3 presented in section 2.3 are fulfilled by construction. Hence, the artificial TAP pulse responses are consistent with assumption 1. The responses were also constructed to be consistent with assumption 2.

Figure 9 shows the 95 % confidence intervals obtained for $\mathscr{D}_e$. In the case of the NLSQ regression, the true value lies within the confidence interval only once. In the case of the new regression, this number increases to fifteen. With a confidence limit of 95 %, the true value would be expected to lie in the confidence interval in nineteen out of twenty cases. That a similar number is not attained even with the new regression approach, can be understood because a number of replicates as small as twenty gives rise to imperfect second-order statistical estimates. The sPCs inherit to a certain extent the randomness of the data set and are thus not able to fulfil conditions 5 and 6 perfectly (see sections 4.1 and 4.2). Note that the confidence intervals aside, the parameter estimates themselves are more accurate with the new regression approach than with the classical NLSQ regression.

Figure 10 shows the ranked parameter estimates in a normal probability graph. The fraction of estimates that fall below values indicated on the x-axis are presented on the y-axis. The y-axis has a normal probability scaling. If the estimates are normally distributed around the true value, a straight line through the point $(2 \cdot 10^{-3}\,\mathrm{m}^2/\mathrm{s}, 50\,\%)$ is expected. This is more or less the case for both regression approaches, which is in conformity with property 1 presented in appendix A.

5.2   Regression of experimental TAP data

5.2.1   Irreversible adsorption of oxygen on a vanadia based catalyst

A TAP experiment was performed to study the model V150 catalyst (Poelman et al., 2007): $V_2O_5|SiO_2$–ZrO_2. Oxygen was pulsed over a three-zone TAP reactor at 773 K. The diameter of the reactor tube was 5 mm. A 4.0 mm long catalytic bed was sandwiched between two inert zones filled with quartz beads. The one at the inlet was 4.4 mm long, the one at the outlet 20.0 mm. The interparticle porosity of all three zones was 0.42. The total catalyst mass was 189.2 mg. Twenty replicate oxygen pulse responses were collected after reduction of the catalyst surface. The calibration coefficient was estimated at $9.9 \cdot 10^5$ V·s/mol.

The interaction between oxygen and the catalyst was modelled as simple irreversible adsorption. The kinetics of this adsorption are of first order in oxygen, with a specific adsorption rate r_a in mol/kg.s, expressed as

$$r_a = k_a C_{0_2}, \tag{55}$$

where k_a is the specific adsorption rate coefficient in m^3/kg.s and C_{0_2} the concen-

tration of oxygen in the interparticle pores, in mol/m^3.

The oxygen pulse responses were regressed with the TAP model of Constales et al. (2004), once by NLSQ regression of the average pulse response and once by SOSR. Aside from the kinetic parameter k_a, 2 physical parameters were estimated: the average oxygen pulse size N_{O_2} and the Knudsen diffusion coefficient $\mathscr{D}_{e,O_2}$ through the reactor bed. Knudsen diffusion is indeed expected to proceed at an equal speed in all three zones as both the catalytic and inert particles were nonporous and had the same size (diameters of 250 to 425 μm). An additional parameter estimated was the average pulse response's baseline position U_0: the voltage corresponding to a zero reactor outflow.

Differences between the pulse responses calculated from direct NLSQ regression and SOSR are hardly observable. Therefore, only the latter is shown in Figure 11. However, there is quite some difference in the corresponding parameter estimates, which are displayed in Table 1. SOSR applies a transformation to fulfil the conditions for the subsequent NLSQ regression. Therefore, the parameter estimates from SOSR are more reliable than those obtained from direct NLSQ regression. The same is true a fortiori for the statistics accompanying the parameter estimates: F and t statistics for global and individual parameter regression significance, where the latter determine the width of the parameter confidence intervals. Indeed, no meaning should be attached to these statistics if they are supplied by the direct NLSQ regression. Table 1 shows that both for NLSQ and SOSR, the F value is larger than the 95% quantile, indicating that the regression is significant. However, the difference is much more pronounced in the case of NLSQ regression. Analogously, the confidence intervals are narrower. This way, the significance of regression is clearly misrepresented.

For the sake of completeness, Table 2 displays the estimates of the binary correlation coefficients $\rho\left(\hat{\beta}_i, \hat{\beta}_j\right)$ between the parameter estimates resulting from SOSR:

$$\rho\left(\hat{\beta}_i, \hat{\beta}_j\right) = \mathrm{cov}\left(\hat{\beta}_i, \hat{\beta}_j\right) \Big/ \sqrt{\mathrm{var}\left(\hat{\beta}_i\right) \cdot \mathrm{var}\left(\hat{\beta}_j\right)}. \tag{56}$$

They follow directly from the estimated variance matrix $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\beta}}\right)$, see (70) in appendix A. The estimated binary correlation coefficients all are between -0.9 and 0.9. Such favourable figures would certainly be unattainable with steady-state kinetic data. k_a is actually the only parameter of interest. Its estimate is most correlated with the Knudsen diffusivity $\mathscr{D}_{e,O_2}$, which indicates that if the latter would be known in advance, a narrower confidence interval would be obtained for k_a. At the same time however, it is clear that the value assumed for $\mathscr{D}_{e,O_2}$ should be accurate, since it strongly affects the estimate for k_a. In general, an accurate independent determination of the physical parameters $\mathscr{D}_{e,O_2}$, N_{O_2} and U_0 is advisable.

5.2.2 Interaction of propane with a copper/ceria catalyst

$CuO-CeO_2|\gamma-Al_2O_3$ was studied as a total oxidation catalyst for volatile organic compounds (VOCs). Propane, chosen as model VOC, was pulsed over a three-zone TAP reactor at 773 K. Part of the propane reacts with the catalyst surface oxygen to form carbon dioxide and water. The first and third zones contained quartz beads and had a length of 2.5 mm and 25.0 mm. The central zone contained 48.6 mg of catalyst and was 3.6 mm long. The interparticle porosity of all zones was 0.53. Ten replicate propane pulse responses were collected after oxidation of the catalyst surface. The calibration coefficient for propane was $4.6 \cdot 10^5$ V·s/mol.

After some regression attempts, the interaction between propane and the catalyst was found to be modelled well by reversible adsorption followed by an irreversible surface reaction. Schematically:

$$C_3H_8 + * \rightleftarrows C_3H_8^*, \tag{57}$$
$$C_3H_8^* \rightarrow \quad . \tag{58}$$

The steps following the latter do not affect the propane response. Adsorption (57) is pseudo first order in propane, in accordance with the assumption that the quantity of propane pulsed is a lot smaller than the number of active sites $*$. The specific rates r_a and r_d of propane adsorption and desorption (57), and r of reaction (58), all in mol/kg.s, are expressed as

$$r_a = k_a C_{C_3H_8}, \tag{59}$$
$$r_d = k_d C_{C_3H_8^*}, \tag{60}$$
$$r = k C_{C_3H_8^*}, \tag{61}$$

where $C_{C_3H_8}$ is the concentration of propane, in mol/m^3, and $C_{C_3H_8^*}$ the quantity of propane reversibly adsorbed per unit mass of catalyst, in mol/kg. The adsorption rate coefficient k_a is in m^3/kg.s. The desorption and reaction rate coefficients k_d and k are in Hz. All three rate coefficients were estimated by regression. The former example revealed that the simultaneous estimation of unknown physical variables has a negative effect on the width of the confidence intervals of the chemical parameters. For this reason, the (average) quantity of propane pulsed was determined independently by pulsing propane/krypton mixtures with known composition. The propane responses were collected in alternation with krypton responses. This allowed to determine the (average) quantity of propane pulsed to be 12 nmol. The propane pulse responses were also submitted to baseline correction before regression, eliminating the baseline position as an unknown parameter. As in the former example, the Knudsen diffusion coefficient $\mathscr{D}_{e,C_{C_3H_8}}$, in m^2/s, of propane could be assumed equal for all zones of the reactor. $\mathscr{D}_{e,C_{C_3H_8}}$ could not, however, be determined independently and therefore had to be estimated by regression simultaneous with k_a, k_d and k.

Figure 12 shows the average of the ten propane pulse responses and the correspond-

ing SOSR-fitted model-calculated response. The NLSQ-fitted pulse response is not depicted as it is again hardly distinguishable from the SOSR-fitted response. The parameter estimates resulting from both regressions are shown in Table 3. Again, the confidence intervals obtained from SOSR are wider than those from NLSQ regression. Table 3 also contains the F statistics for global significance of the regression. In both cases, the F value exceeds the 95 % quantile. However, for NLSQ regression, the difference is artificially high. This shows again that the F test and confidence intervals from NLSQ regression can mislead the experimentalist if the conditions 1 to 6 are not fulfilled.

Finally, Table 4 displays the binary correlation coefficients of the parameters estimated by SOSR. The strongest correlations exist between the rate coefficients of the reversible adsorption, k_a and k_d, and the reaction rate coefficient k. Though important, these correlations are still reasonably far from perfect. As was already pointed out in the former example, this is due to the transient nature of the experiment.

6   Conclusions

A sample principal component analysis (sPCA) of replicate data has lead to a maximum-likelihood parameter regression technique with higher accuracy for both the estimates and the corresponding statistics. The so-called second order statistical regression (SOSR) technique allows to relax several of the assumptions that underlie the classic regression techniques, in particular the assumptions of whiteness and homoskedasticity.

The proposed technique can be applied in a straightforward way to time series related to chemical reaction kinetics or other physico-chemical phenomena. Although based on second-order statistics, i.e. covariances, the procedure does not require the explicit calculation of the latter and is not computationally demanding.

7   Notation

Arabic symbol:

$\mathbf{0}_n$ zero column vector with n components, –

Roman symbols:

$\mathbf{b}$ vector of physico-chemical parameters to be varied during regression

b physico-chemical parameter, component of $\mathbf{b}$

21

C_A concentration of gas A, mol/m^3

C_A^* quantity of A adsorbed per unit mass of catalyst, mol/kg

$\mathscr{D}_{e,A}$ effective Knudsen diffusivity of gas A through a TAP reactor, m^2/s

$E[\ldots]$ expected value of a stochastic variable, vector or matrix

$F_{\chi:m,n}$ χ quantile of the F distribution with m and n degrees of freedom, –

$\mathbf{f}$ vector function of $\boldsymbol{\xi}$ and $\mathbf{b}$ to a model-calculated analogue of $\mathbf{y}$

$\mathbf{I}_n$ $(n \times n)$ unit matrix, –

k specific surface reaction rate coefficient, Hz

k_a specific adsorption rate coefficient, $m^3/kg.s$

k_d specific desorption rate coefficient, Hz

N_A quantity of gas A pulsed in a TAP experiment, mol

n number of components of $\mathbf{y}$, –

n''' total number of replicate experimental time series, –

n_e number of experiments carried out, not counting their replicates, –

n_t number of time samples in one time series, –

n_v number of variables recorded during a transient kinetic experiment, –

p number of unknown physico-chemical parameters, –

r specific rate of a surface reaction, mol/kg.s

r_a specific rate of adsorption, mol/kg.s

r_d specific rate of desorption, mol/kg.s

r_e number of replicates obtained at experimental conditions $\boldsymbol{\xi}_e$, –

s^2 population variance

$\mathbf{T}$ smoothing matrix, –

t time, s

$t_{\chi:n}$ χ quantile of the t distribution with n degrees of freedom, –

$\mathbf{U}$ matrix with population principal noise components as columns, –

$\hat{\mathbf{U}}$ matrix with sample principal noise components as columns, –

$\mathbf{u}$ population principal noise component, –

$\hat{\mathbf{u}}$ sample principal noise component, –

U_0	baseline position of a TAP pulse response, V
V	population variance matrix
$\hat{\mathbf{V}}$	sample variance matrix
V	physical variable monitored during a transient experiment
y	vector representing one or more experimental time series
$\bar{\mathbf{y}}$	average of replicate **y**'s
z_χ	χ quantile of the unit normal z distribution, –

Greek symbols:

$\boldsymbol{\beta}$	vector of true, unknown physico-chemical parameters
β	true, unknown physico-chemical parameter, component of $\boldsymbol{\beta}$
$\hat{\boldsymbol{\beta}}$	estimate of $\boldsymbol{\beta}$ obtained by regression
$\hat{\beta}$	estimate of β obtained by regression, component of $\hat{\boldsymbol{\beta}}$
Δt	sampling interval, s
$\boldsymbol{\varepsilon}$	vector representing the noise in one or more experimental time series
θ	angle between the experimental set and the space spanned by the sample principal noise components, –
$\boldsymbol{\Lambda}$	diagonal matrix of population principal variances
$\hat{\boldsymbol{\Lambda}}$	diagonal matrix of sample principal variances
ρ	binary correlation coefficient, –
σ^2	population variance
$\boldsymbol{\xi}$	vector of experimental conditions

Abbreviations:

NLSQ	nonlinear least-squares
pPC	population principal component
pPCA	population principal component analysis
SOSR	second-order statistical regression
sPC	sample principal component

sPCA   sample principal component analysis

TAP   temporal analysis of products

Column vectors are represented by small bold letters, e.g. $\mathbf{a}$. The ith element of $\mathbf{a}$ is represented by $(\mathbf{a})_i$. Matrices are represented by bold capitals, e.g. $\mathbf{A}$. To avoid overloading the reader with symbols, no symbolic difference was made in this paper between stochastic variables and the values they assume. The distinction should be clear from the context.

8 Acknowledgements

A   Least-squares regression: theory

Very generally, say $\mathbf{y} \in \mathbb{R}^{n \times 1}$ is a vector of empirical values and $\mathbf{f}(\mathbf{b})$ its analogue calculated from a model with parameters collected in the vector $\mathbf{b} \in \mathbb{R}^{p \times 1}$. Assume that the model applies with some real, unknown parameters $\boldsymbol{\beta}$:

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{62}$$

with $\boldsymbol{\varepsilon}$ a random error vector with multidimensional normal distribution with average zero ($\mathbf{0}_n$) and variance matrix $\mathbf{I}_n \sigma^2$:

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}_n, \mathbf{I}_n \sigma^2\right). \tag{63}$$

$\boldsymbol{\beta}$ is estimated as $\hat{\boldsymbol{\beta}}$, minimizing the residual sum of squares

$$S(\mathbf{b}) = \|\mathbf{y} - \mathbf{f}(\mathbf{b})\|^2. \tag{64}$$

The following important properties apply in good approximation if a first order Taylor expansion of $\mathbf{f}$ is valid in good approximation in a sufficiently large domain about $\boldsymbol{\beta}$:

$$\mathbf{f}(\boldsymbol{\beta} + \Delta \mathbf{b}) \approx \mathbf{f}(\boldsymbol{\beta}) + \mathbf{J}\Delta \mathbf{b}, \tag{65}$$

with $\mathbf{J}$ the Jacobian matrix

$$\mathbf{J} = \frac{\partial \mathbf{f}}{\partial \mathbf{b}}(\boldsymbol{\beta}) \in \mathbb{R}^{n \times p}. \tag{66}$$

Property 1 $\hat{\boldsymbol{\beta}}$ *is multidimensional normally distributed with mean $\boldsymbol{\beta}$ and variance matrix*

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = (\mathbf{J}^T \mathbf{J})^{-1} \sigma^2, \tag{67}$$

It follows that $\hat{\boldsymbol{\beta}}$ is an unbiased and maximum-likelihood estimator of $\boldsymbol{\beta}$. The Jacobian $\mathbf{J}$ is calculated from

$$\mathbf{J} \approx \frac{\partial \mathbf{f}}{\partial \mathbf{b}}(\hat{\boldsymbol{\beta}}), \tag{68}$$

and the error variance σ^2 is estimated as the mean squared residual:

$$s^2 = \frac{(\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\beta}}))^T (\mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\beta}}))}{n - p}. \tag{69}$$

The variance matrix $\mathbf{V}(\hat{\boldsymbol{\beta}})$ is estimated as

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = (\mathbf{J}^T \mathbf{J})^{-1} s^2. \tag{70}$$

Property 2 *The variable*

$$t_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{V}_{ii}(\hat{\boldsymbol{\beta}})}}, \tag{71}$$

where $\hat{V}_{ii}(\hat{\boldsymbol{\beta}})$ is the ith diagonal element of $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$, is distributed as Student's t distribution with $n - p$ degrees of freedom.

This allows to test if the real, unknown parameter β_i differs significantly from a certain value β_i^*. It is concluded with likelihood $1 - \alpha$ that it does not if

$$\left| \frac{\hat{\beta}_i - \beta_i^*}{\sqrt{\hat{V}_{ii}(\hat{\boldsymbol{\beta}})}} \right| < t_{1-\alpha/2:n-p}, \tag{72}$$

where the $t_{1-\alpha/2:n-p}$ quantile can be calculated with a suitable statistical package. If $\beta_i = \beta_i^*$, there is a probability $1 - \alpha$ that (72) leads to the correct conclusion, namely that β_i does not differ significantly from β_i^*. Typically, α is taken to be 5 %. This test can be used to verify whether a parameter differs significantly from a neutral (trivial) value (mostly 0), or, shortly, whether the parameter estimate is significant. (72) is readily transformed into

$$\hat{\beta}_i - \frac{\Delta \beta_i}{2} < \beta_i^* < \hat{\beta}_i + \frac{\Delta \beta_i}{2}, \tag{73}$$

representing the $1 - \alpha$ confidence interval for parameter β_i^*, with

$$\Delta\beta_i = 2\,t_{1-\alpha/2:n-p}\,\sqrt{\hat{V}_{ii}(\hat{\boldsymbol{\beta}})}. \tag{74}$$

In the frequently occurring case that

$$n - p \gg 1, \tag{75}$$

t_i in (71) can be considered unit normally distributed. Consequently, the approximation

$$t_{1-\alpha/2:n-p} \approx z_{1-\alpha/2} \tag{76}$$

can then be applied in equations (72) and (74).

Property 3 *The variable*

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{J}^T \mathbf{J} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{p\,s^2} \tag{77}$$

is distributed as Snedecor's F distribution with p and n − p degrees of freedom.

This allows to test whether $\boldsymbol{\beta}$ as a whole differs significantly from a certain vector of parameters $\boldsymbol{\beta^*}$. It is concluded with $1 - \alpha$ likelihood that it does not if

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta^*})^T \mathbf{J}^T \mathbf{J} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta^*})}{p\,s^2} < F_{1-\alpha:p,n-p}. \tag{78}$$

Mostly $\boldsymbol{\beta^*}$ is taken to be a vector of neutral parameters (mostly the null vector). (78) then means that the regression is globally insignificant. Inequality (78) implicitly defines the joint confidence region of all parameters, delimited by a hyperellipsoid. pF with F given by (77) can be considered χ^2 distributed with p degrees of freedom if (75) is valid. The approximation

$$F_{1-\alpha:p,n-p} \approx \frac{1}{p}\chi^2_{1-\alpha:p} \tag{79}$$

can then be applied in (78).

References

Bennett, C.O. (2000). Experiments and processes in the transient regime for heterogeneous catalysis. *Advances in Catalysis*, *44*, 329–416.

Biloen, P. (1983). Transient kinetic methods. *Journal of Molecular Catalysis*, *21*, 17–24.

Boggs, P.T., Byrd, R.H., Schnabel, R.B. (1987). A stable and efficient algorithm for nonlinear orthogonal distance regression. *SIAM Journal on Scientific and Statistical Computation*, *8*, 1052–1078.

Boudart, M. (1968). Kinetics of Chemical Processes. In N.R. Amundsen (Ed.), *Prentice-Hall International Series in the Physical and Chemical Engineering Sciences*. Englewood Cliffs: Prentice-Hall.

Cutlip, M.B., Yang, C.C., & Bennett, C.O. (1972). Parameter Estimation from Transient Rate Data. *AIChE Journal*, *18*, 1073–1076.

Constales, D., Yablonsky, G.S., Marin, G.B., & Gleaves, J.T. (2004). Multi-zone TAP-reactors theory and application. III Multi-response theory and criteria of instantaneousness. *Chemical Engineering Science*, *59*, 3725–3736.

Constales, D., Yablonsky, G.S., Marin, G.B., & Gleaves, J.T. (2006). Multi-zone TAP-reactors theory and application IV. Ideal and non-ideal boundary conditions. *Chemical Engineering Science*, *61*, 1878–1891.

Draper, N.R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). In V. Barnett et al. (Eds.), *Wiley Series in Probability and Statistics*. New York: Wiley.

Efstathiou, A.M., & Verykios, X.E. (1997). Transient methods in heterogeneous catalysis: Experimental features and application to study mechanistic aspects of the CH_4/O_2 (OCM), NH_3/O_2 and NO/He reactions. *Applied Catalysis A: General*, *151*, 109–166.

Froment, G.F., & Hosten, L.H. (1981). Catalytic Kinetics: Modelling. In Anderson, J.R., & Boudart, M. (Eds.), Catalysis – Science and Technology, vol. 2. Berlin: Springer-Verlag (chapter 3, pp. 97–170).

Furusawa, T., Suzuki, M., & Smith, J.M. (1976). Rate parameters in heterogeneous catalysis by pulse techniques. *Catalysis Reviews: Science and Engineering*, *13*, 43–76.

Gleaves, J.T., Ebner, J.R., & Kuechler, T.C. (1988). Temporal analysis of products (TAP) – a unique catalyst evaluation system with submillisecond time resolution. *Catalysis Reviews: Science and Engineering*, *30*, 49–116.

Gleaves, J.T., Yablonskii, G.S., Phanawadee, P., & Schuurman, Y. (1997). TAP-2: an interrogative kinetics approach. *Applied Catalysis A: General*, *160*, 55–88.

Graybill, F.A., & Iyer, H.K. (1994). Regression Analysis: Concepts and Applications. Belmont: Wadsworth.

Hindmarsh, A.C. (1983). ODEPACK, A Systematized Collection of ODE Solvers. In R.S. Stepleman et al. (Eds.), *Scientific Computing*. Amsterdam: North-Holland.

Hougen, O.A., & Watson, K.M. (1947). Chemical Process Principles. Part Three. Kinetics and Catalysis. New York: Wiley.

Hulburt, H.M., Kim, Y.G. (1966). Reaction mechanisms for engineering design. *Industrial and Engineering Chemistry*, *58*, *No. 9*, 20–31.

Jolliffe, I.T. (2004). Principal Component Analysis (2nd ed.). In P. Bickel et al. (Eds.), *Springer Series in Statistics*. New York: Springer.

Kittrell, J.R., Hunter, W.G., & Watson, C.C. (1965). Nonlinear least squares analysis of catalytic rate models. *AIChE Journal*, *11*, 1051–1057.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least-squares. *Quarterly of Applied Mathematics*, *2*, 164–168.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, *11*,

431–441.

Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., De Jong, S., Lewi, P.J., & Smeyers-Verbeke, J. (1997). Handbook of Chemometrics and Qualimetrics: Part A. In B.G.M. Vandeginste, & S.C. Rutan (Eds.), *Data Handling in Science and Technology*. Amsterdam: Elsevier.

Mills, P.L., & Lerou, J.J. (1993). Transient response methods for assisted design of gas phase heterogeneous catalysts: experimental techniques and mathematical modeling. *Reviews in Chemical Engineering*, *9*, 1–96.

Nijhuis, T.A., van den Broeke, L.J.P., van de Graaf, J.M., Kapteijn, F., Makkee, M., & Moulijn, J.A. (1997). Bridging the gap between macroscopic and NMR diffusivities. *Chemical Engineering Science*, *52*, 3401–3404.

Poelman, H., Sels, B.F., Olea, M., Eufinger, K., Paul, J.S., Moens, B., Sack, I., Balcaen, V., Bertinchamps, F., Gaigneaux, E.M., Jacobs, P.A., Marin, G.B., Poelman, D., & De Gryse, R. (2007). New supported vanadia catalysts for oxidation reactions prepared by sputter deposition. *Journal of Catalysis*, *245*, 156–172.

Renken, A. (1993). Transient operation for the purpose of modeling heterogeneous catalytic reactions. *International Chemical Engineering*, *33*, 61–71.

Roelant, R., Constales, D., Yablonsky, G.S., Van Keer, R., Rude, M.A., & Marin, G.B. (2007). Noise in temporal analysis of products (TAP) pulse responses. *Catalysis Today*, *121*, 269–281.

Schuurman, Y. (2007). Assessment of kinetic modeling procedures of TAP experiments. *Catalysis Today*, *121*, 187–196.

Seber, G.A.F., & Lee, A.J. (2003). Linear Regression Analysis (2nd ed.). In V. Barnett et al. (Eds.), *Wiley Series in Probability and Statistics*. New York: Wiley.

Seber, G.A.F., & Wild, C.J. (2003). Nonlinear Regression. In V. Barnett et al. (Eds.), *Wiley Series in Probability and Statistics*. New York: Wiley.

Tamaru, K. (1964). Adsorption measurements during surface catalysis. *Advances in Catalysis and Related Subjects*, *15*, 65–90.

Therrien, C.W. (1992). Discrete Random Signals and Statistical Signal Processing. In A.V. Oppenheim (Ed.), *Prentice Hall Signal Processing Series*. Englewood Cliffs: Prentice-Hall.

van der Linde, S.C., Nijhuis, T.A., Dekker, F.H.M., Kapteijn, F., & Moulijn, J.A. (1997). Mathematical treatment of transient kinetic data: combination of parameter estimation with solving the related partial differential equations. *Applied Catalysis A: General*, *151*, 27–57.

Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J. (1998). Handbook of Chemometrics and Qualimetrics: Part B. In B.G.M. Vandeginste, & S.C. Rutan (Eds.), *Data Handling in Science and Technology*. Amsterdam: Elsevier.

Yablonsky, G.S., Olea, M., & Marin, G.B. (2003). Temporal analysis of products: basic principles, applications and theory. *Journal of Catalysis*, *216*, 120–134.

Zamostny, P., & Belohlav, Z. (2002). Identification of kinetic models of heterogeneously catalyzed reactions. *Applied Catalysis A: General*, *225*, 291–299.
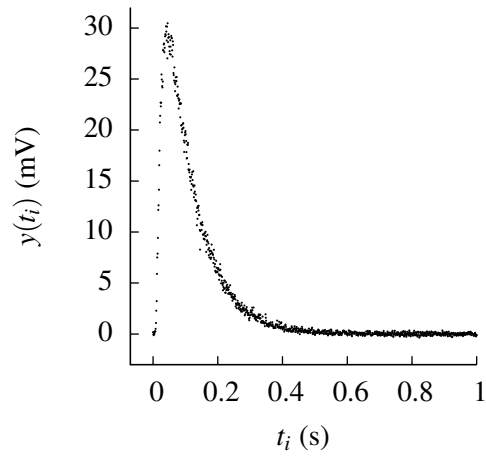
List of Figures

Figure 1. A typical TAP pulse response, consisting of a thousand samples: $i \in \{1, 2, \ldots, 1000\}$. For this response, argon was pulsed over quartz.
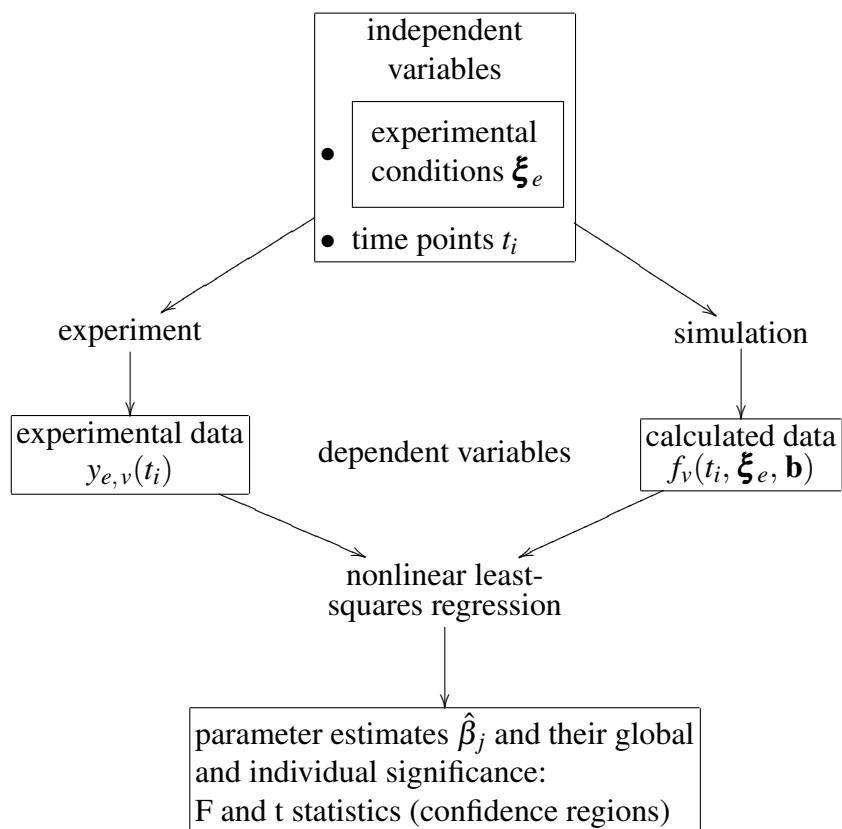
31

Figure 2. Overall scheme of the direct regression approach as a means to estimate physico-chemical parameters from transient kinetic data.
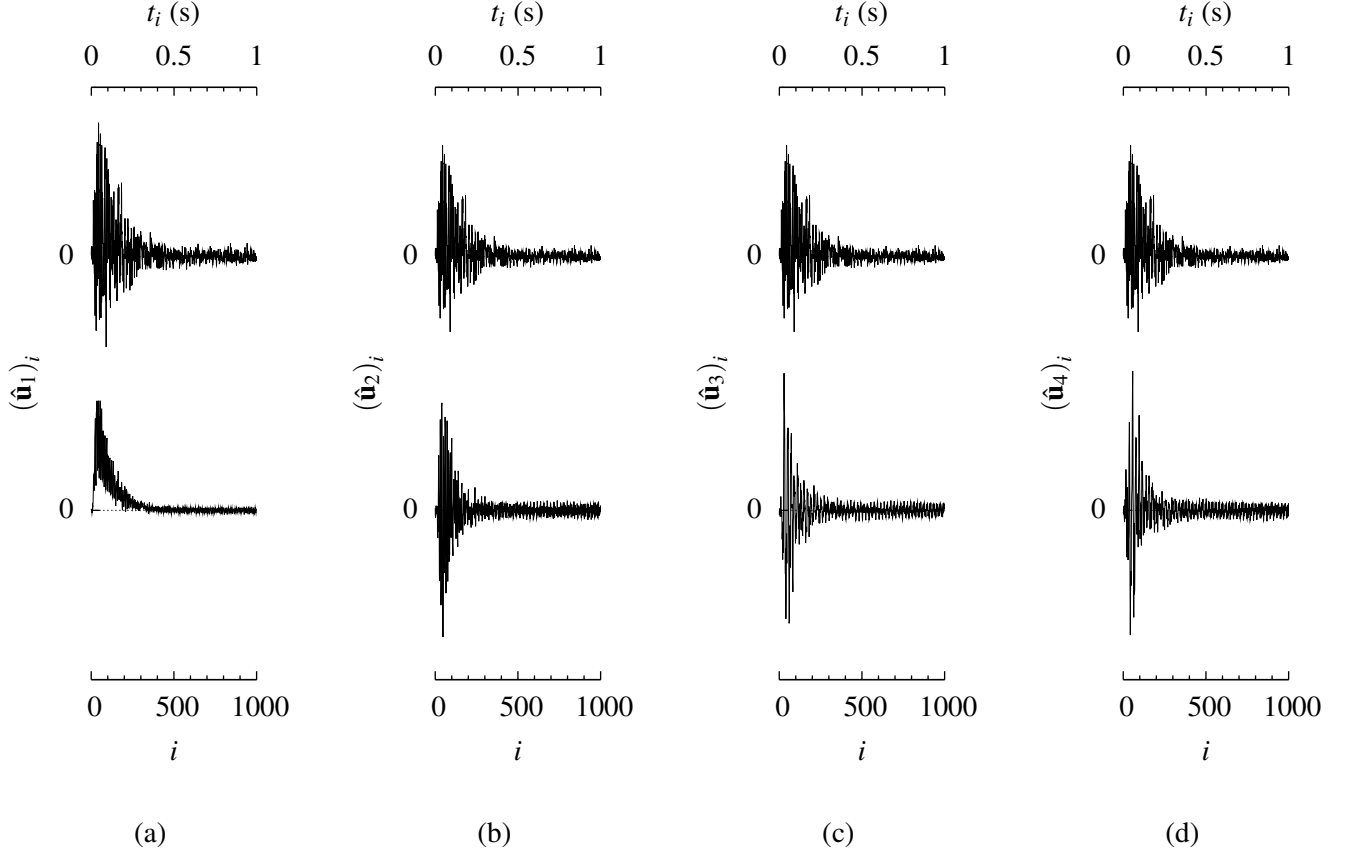
Figure 3. The first (a), second (b), third (c) and fourth (d) sample principal noise component calculated from a set of replicate experimental TAP pulse responses. $(\hat{\mathbf{u}}_k)_i$ refers to the ith element of the kth sample principal component $\hat{\mathbf{u}}_k$. The lower graphs show the sample principal components calculated from a set of eight hundred replicates. The upper graphs show those calculated from a reduced set of twenty of these replicates. In both cases, the absolute magnitudes of the $(\hat{\mathbf{u}}_k)_i$ are irrelevant because the vectors are normalized by construction: $\sum_{i=1}^{1000} (\hat{\mathbf{u}}_k)_i^2 = 1$. For this reason, a graduated y-axis is omitted from the graph.
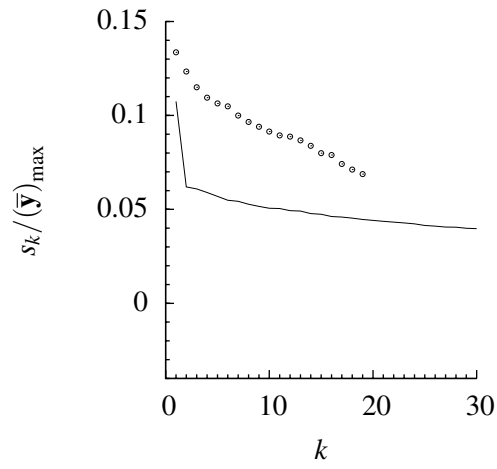
Figure 4. The noise's sample standard deviations s_k along the first few of its sample principal components, calculated from a set of replicate experimental TAP pulse responses and rendered dimensionless by dividing by the peak $(\bar{\mathbf{y}})_{max}$ of the average pulse response $\bar{\mathbf{y}}$. The full line connects the first thirty sample standard deviations calculated from a set of eight hundred replicates. The circles represent the nineteen sample standard deviations calculated from a reduced set of twenty replicates.

34
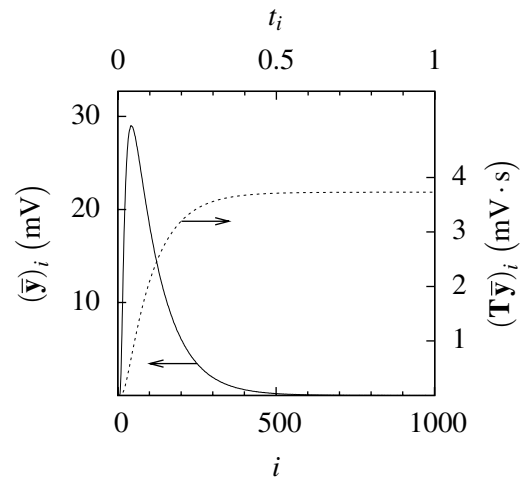
Figure 5. Average pulse response $\overline{\mathbf{y}}$ of a set of eight hundred replicated pulse responses and its Riemann cumulative $\mathbf{T}\overline{\mathbf{y}}$. The subscript i in $(\overline{\mathbf{y}})_i$ and $(\mathbf{T}\overline{\mathbf{y}})_i$ refers to the ith element of both vectors.
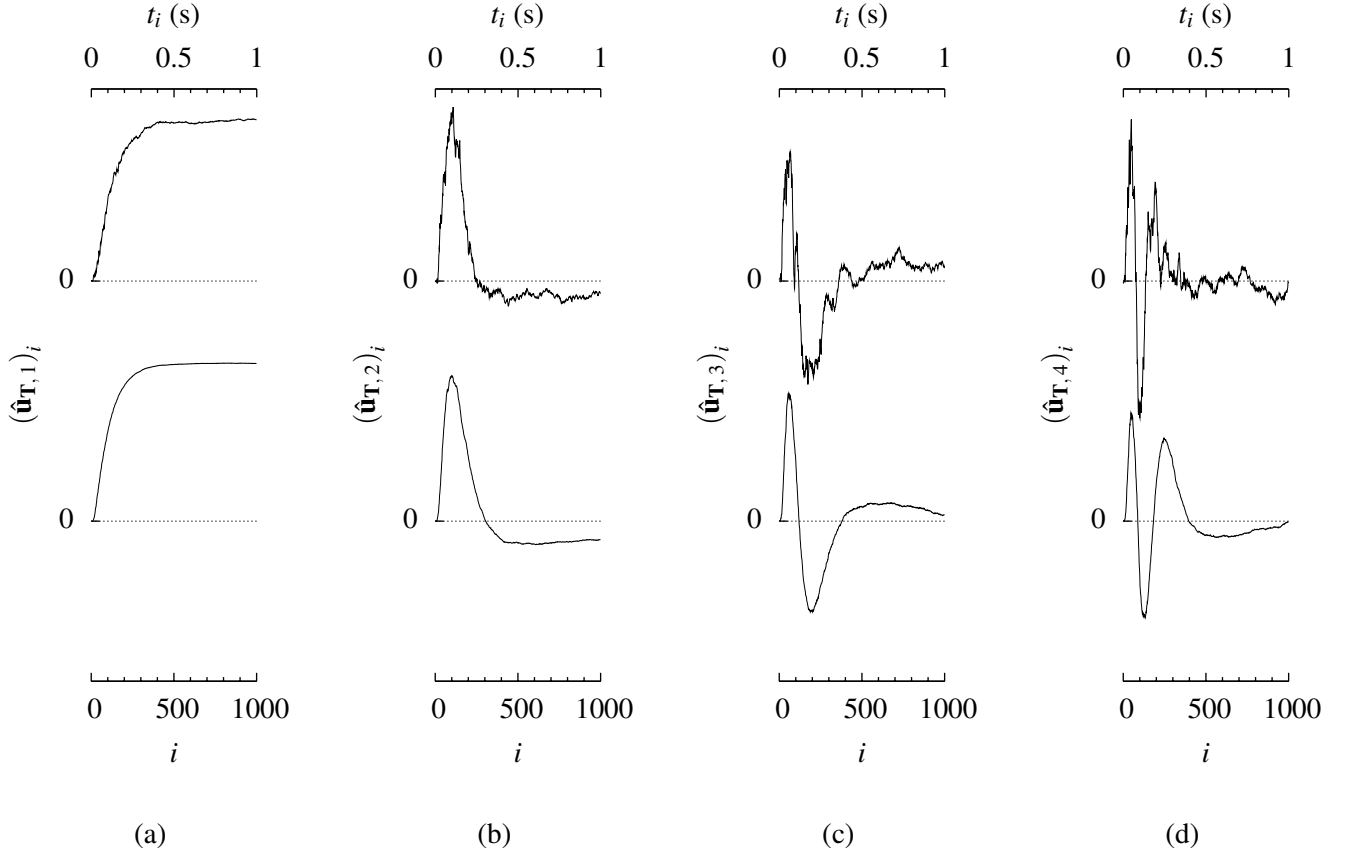
Figure 6. The first (a), second (b), third (c) and fourth (d) sample principal noise component of the Riemann cumulative of an experimental TAP pulse response of which eight hundred replicates were taken. $(\hat{\mathbf{u}}_{\mathbf{T},k})_i$ refers to the ith element of the kth sample principal component $\hat{\mathbf{u}}_{\mathbf{T},k}$. The lower graphs show the sample principal components estimated from the full set of replicates. The upper graphs show those estimated from a reduced set of twenty replicates.
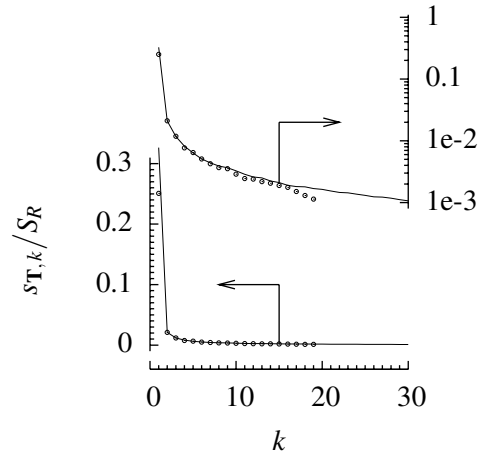
Figure 7. The sample standard deviations $s_{\mathbf{T},k}$ along the first few sample principal components of the noise in the Riemann cumulative $\mathbf{T}\bar{\mathbf{y}}$ of an experimental argon TAP pulse response. They were estimated from a set of replicates and rendered dimensionless by dividing by $S_R = \left(\mathbf{T}\bar{\mathbf{y}}\right)_{n_t} = \Delta t \cdot \sum_{j=1}^{n_t} (\bar{\mathbf{y}})_j$, the Riemann sum of the average. In both the linear and the semilogarithmic plot, the full line connects the first thirty standard deviations calculated from a set of eight hundred replicate experimental pulse responses. The circles represent the nineteen standard deviations calculated from a reduced set of twenty replicates.

37
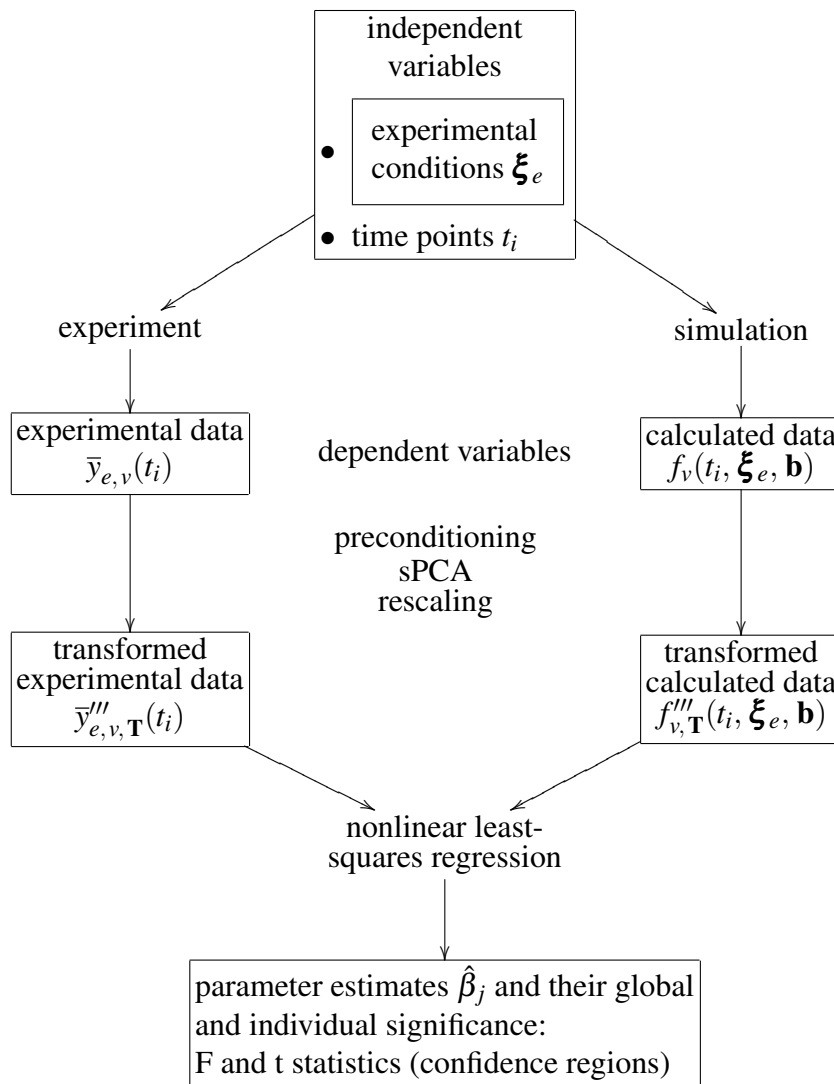
Figure 8. Overall scheme of the second-order statistical regression (SOSR) approach as a means to estimate physico-chemical parameters from transient kinetic data. SOSR performs a nonlinear least-squares (NLSQ) regression of linearly transformed data. The idea of the three-part transformation is to fulfil certain conditions for NLSQ regression as best as possible. It is inferred from replicate experimental time series.
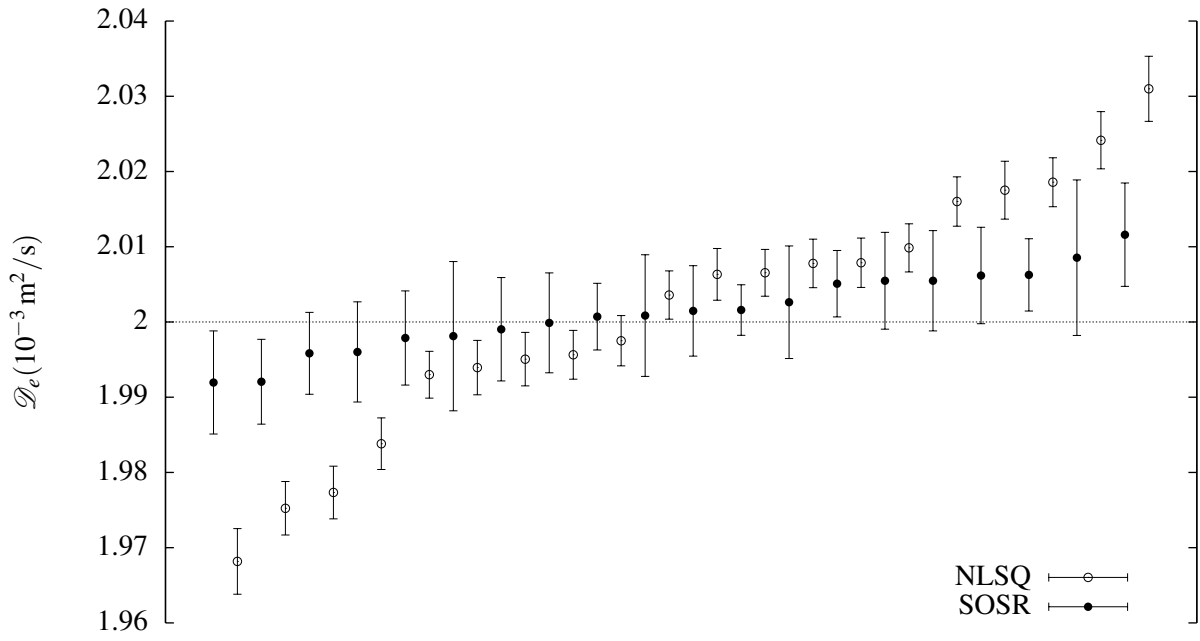
Figure 9. 95 % confidence intervals for the Knudsen diffusivity $\mathscr{D}_e$ estimated by NLSQ regression and SOSR, ranked from low to high. The data used were groups of twenty artificial TAP pulse responses, calculated from a model with $\mathscr{D}_e = 2 \cdot 10^{-3}\,\mathrm{m^2/s}$, with typical noise superposed.

39
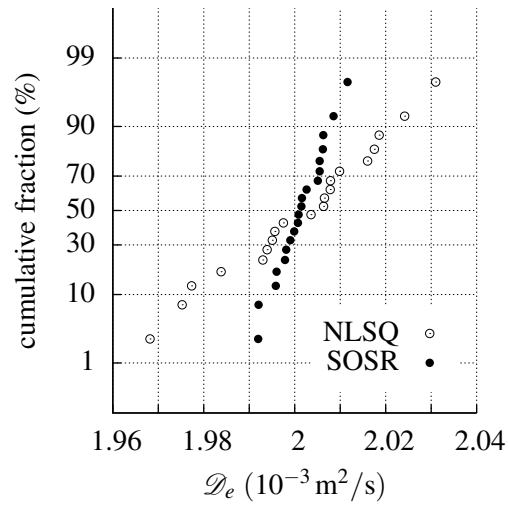
Figure 10. Normal probability graph of estimates of the Knudsen diffusivity $\mathscr{D}_e$, obtained by regression of different statistically homogeneous artificial TAP data sets. The fraction of estimates that fall below values indicated on the x-axis are presented on the y-axis, which has a normal probability scaling. If the estimates are normally distributed around the true value, straight lines through the point $(2 \cdot 10^{-3}\,\text{m}^2/\text{s}, 50\%)$ are expected.
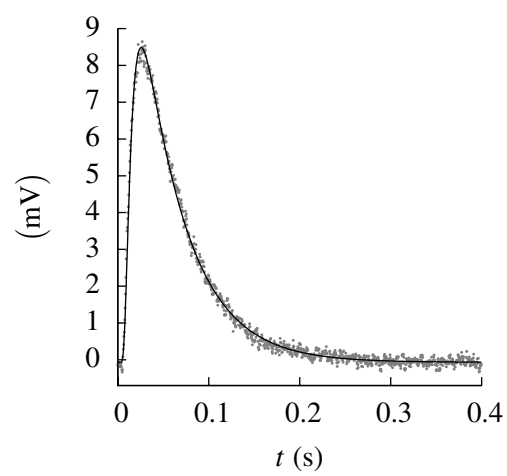
Figure 11. The average of twenty replicate experimental oxygen pulse responses over a three-zone TAP reactor with central vanadia/silica catalytic zone and the model-calculated analogue resulting from SOSR. The TAP reactor model used assumes irreversible adsorption of oxygen on the catalyst.
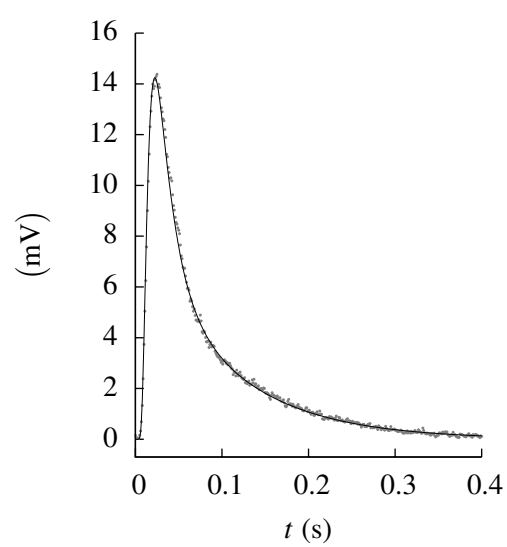
Figure 12. The average of ten replicate experimental propane pulse responses over a three-zone TAP reactor with central copper/ceria catalytic zone and model-calculated analogue resulting from SOSR. The interaction of propane with the catalyst is modelled as reversible adsorption followed by an irreversible surface reaction.

List of Tables

Table 1
Parameter estimates and approximate individual 95% probability level confidence limits by regression of a set of twenty replicate experimental oxygen TAP pulse responses. The regression was carried out by NLSQ regression, and by SOSR. The F values for global significance and corresponding 95% quantile are indicated for both approaches.

	NLSQ	SOSR
k_a (10^{-3} m^3/kg·s)	2.927 ± 0.206	3.731 ± 0.640
$\mathscr{D}_{e,\mathrm{O_2}}$ (10^{-3} m^2/s)	2.047 ± 0.017	1.982 ± 0.059
$N_{\mathrm{O_2}}$ (nmol)	0.7951 ± 0.0113	0.8396 ± 0.0510
U_0 (µV)	-93.00 ± 13.46	-68.79 ± 12.95
F	$9.142 \cdot 10^4$	$1.395 \cdot 10^3$
$F_{95\%}$	2.38	3.01

Table 2

Estimates of the binary correlation coefficients $\rho(\hat{\beta}_i, \hat{\beta}_j)$ between the parameter estimates obtained by SOSR of twenty replicate experimental oxygen TAP pulse responses.

$\hat{\beta}_i \rightarrow$ $\downarrow \hat{\beta}_j$	k_a	$\mathscr{D}_{e,O_2}$	N_{0_2}	U_0
k_a	1	-0.8908	0.8264	0.6711
$\mathscr{D}_{e,O_2}$	-0.8908	1	-0.8722	-0.4115
N_{0_2}	0.8264	-0.8722	1	0.4090
U_0	0.6711	-0.4115	0.4090	1

Table 3
Parameter estimates and approximate individual 95% probability level confidence limits by regression of a set of ten replicate experimental propane TAP pulse responses. The regression was carried out by NLSQ regression, and by SOSR. The F values for global significance and corresponding 95% quantile are indicated for both approaches.

	NLSQ	SOSR
k_a (10^{-3} m^3/kg·s)	271.0 ± 1.7	300.9 ± 53.2
k_d (Hz)	10.41 ± 0.17	12.2 ± 2.0
k (Hz)	9.74 ± 0.13	10.27 ± 0.49
$\mathscr{D}_{e,\mathrm{C_3H_8}}$ (10^{-3} m^2/s)	3.391 ± 0.007	3.520 ± 0.146
F	$3.080 \cdot 10^5$	615.2
$F_{95\%}$	2.38	5.19

Table 4

Estimates of the binary correlation coefficients $\rho\left(\hat{\beta}_i, \hat{\beta}_j\right)$ between the parameter estimates obtained by SOSR of ten replicate experimental propane TAP pulse responses.

$\hat{\beta}_i \rightarrow$ $\downarrow \hat{\beta}_j$	k_a	k_d	k	$\mathscr{D}_{e,C_3H_8}$
k_a	1	0.9139	0.7330	0.6335
k_d	0.9139	1	0.8812	0.6677
k	0.7330	0.8812	1	0.2913
$\mathscr{D}_{e,C_3H_8}$	0.6335	0.6677	0.2913	1