Second-order statistical regression of kinetic time series

Raf Roelant[1], Denis Constales[2],
Roger Van Keer[2], Guy B. Marin[1]

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

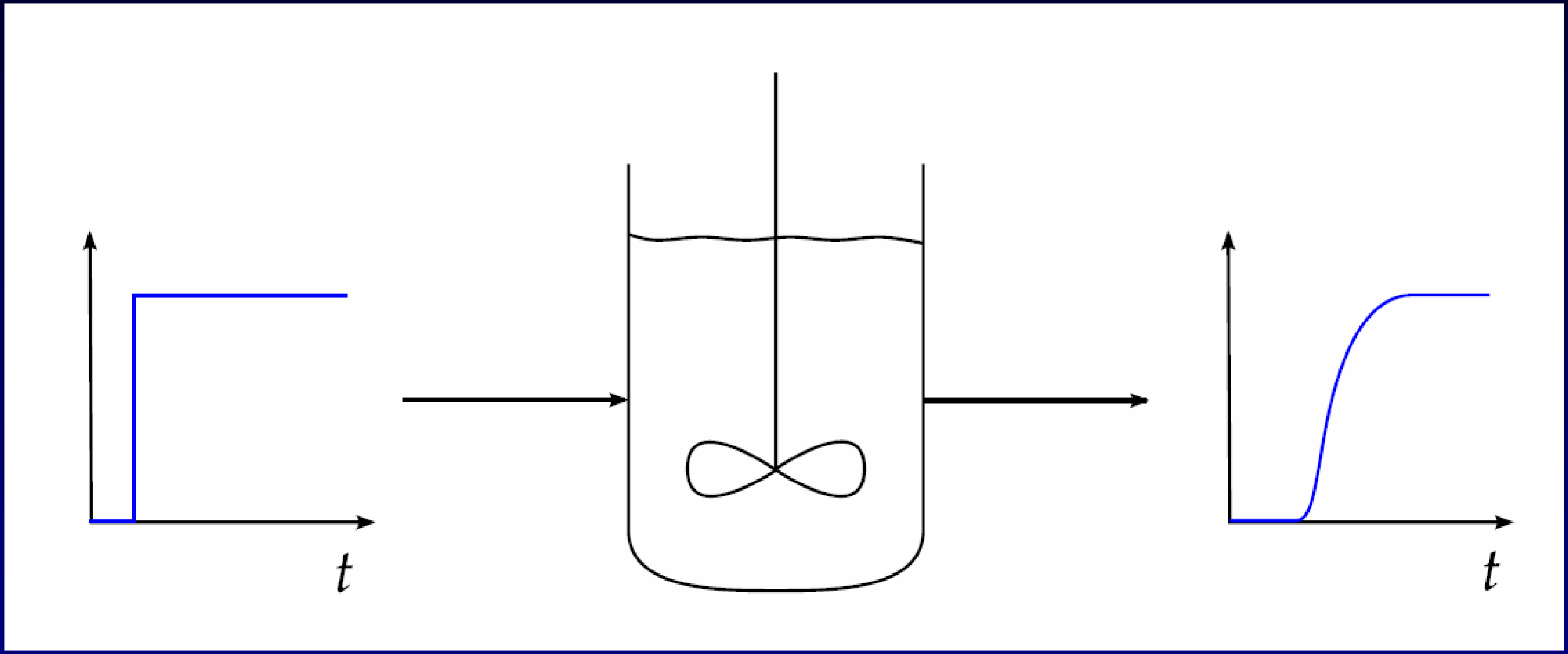
[2]Research group NfaM^2,
Department of Mathematical Analysis,
Ghent University
Galglaan 2, B-9000 Gent, Belgium

20th Eurokin workshop, October 2nd 2007, Trondheim

Overview

- Introduction
- Mechanism elucidation by regression
- Second-order statistical regression
- Numerical experiments
- Application
- Conclusions

Introduction



Transient kinetic experiments record the variation of certain physical variables (concentrations, temperature, pressure) in response to a forced variation (pulse, step, ramp, oscillation) of another such variable.

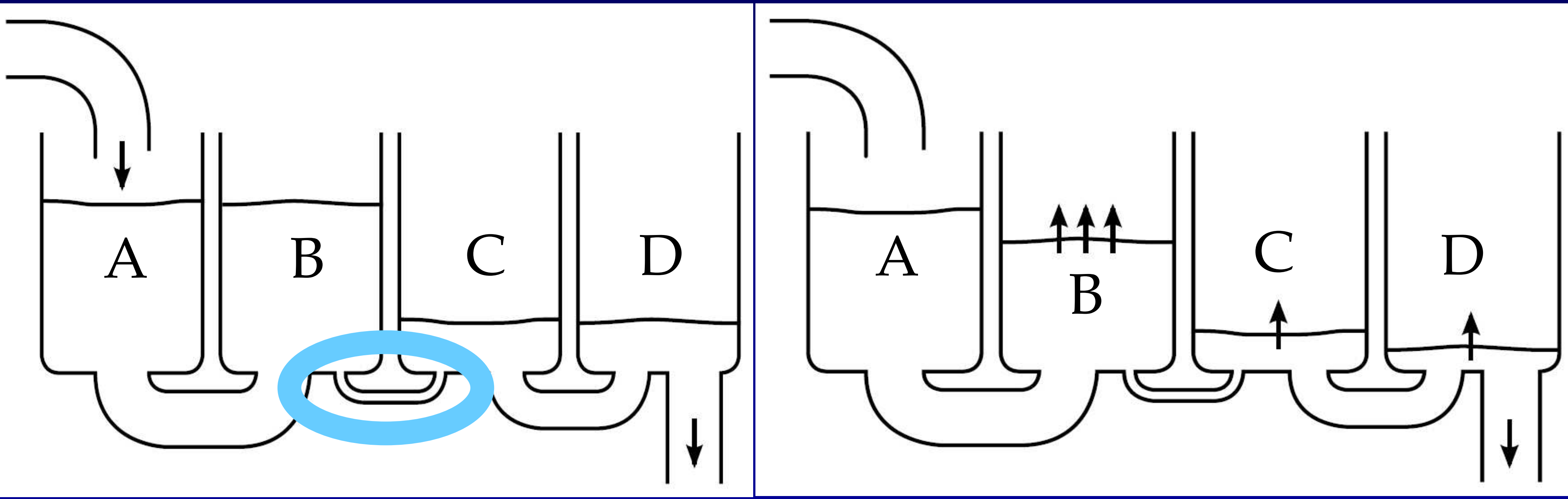
Introduction

Transient experiments outperform steady-state experiments in the determination of reaction mechanisms.

$$A \rightleftarrows B \rightleftarrows C \rightleftarrows D$$

steady-state experiment

transient experiment

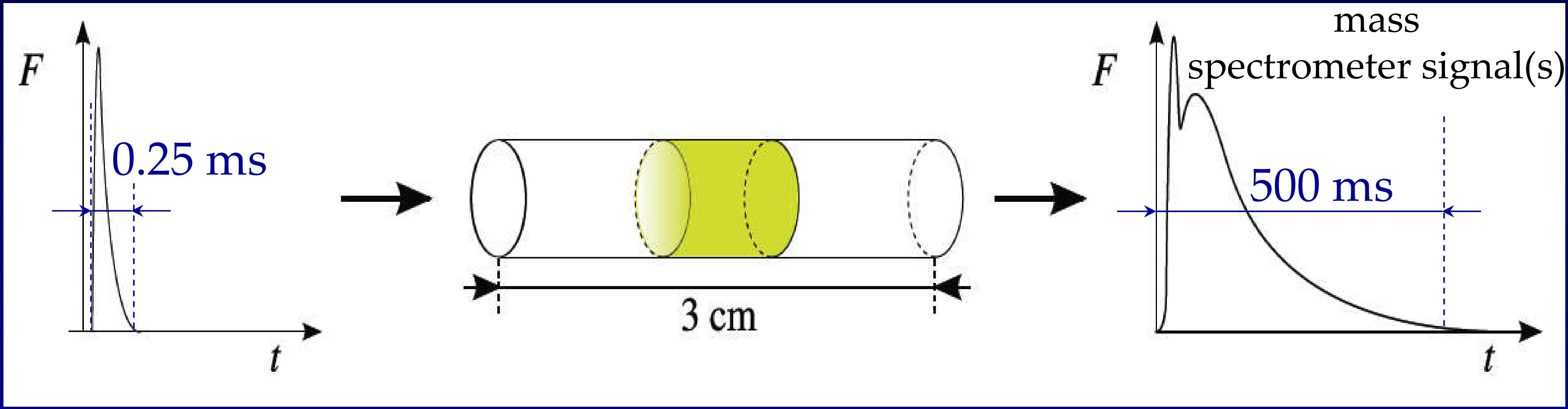


only rate-determining step manifested

after K. Tamaru (1964), Adv. Catal., 15, 65–90.

Introduction

temporal analysis of products (TAP)



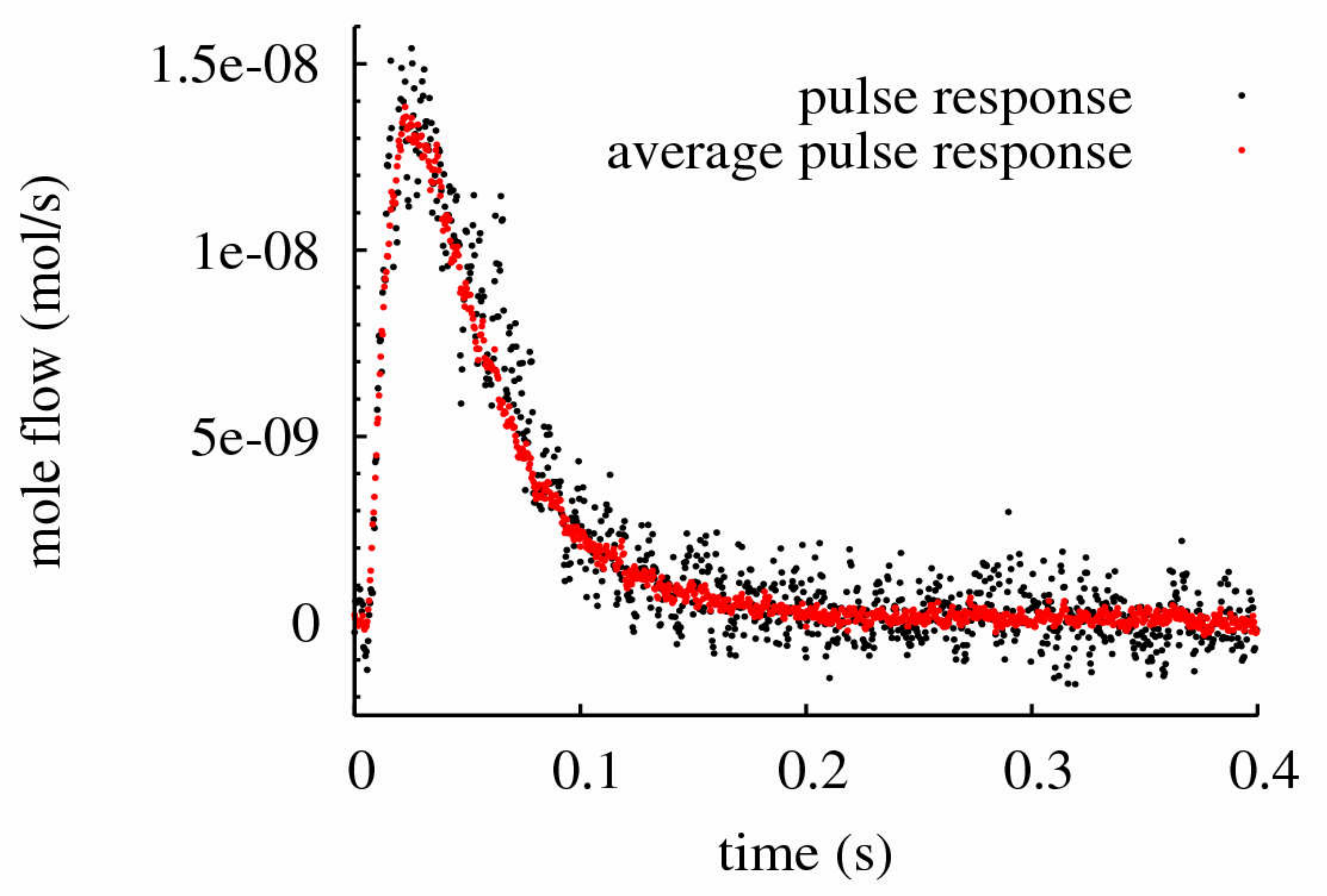
reactant
inlet pulse

catalytic reactor bed

outlet flux vs. time
of reactants and
products:
pulse responses

Introduction

Pulse responses saved and processed as time series.
Typically, 20 pulse responses averaged to increase signal-to-noise ratio.

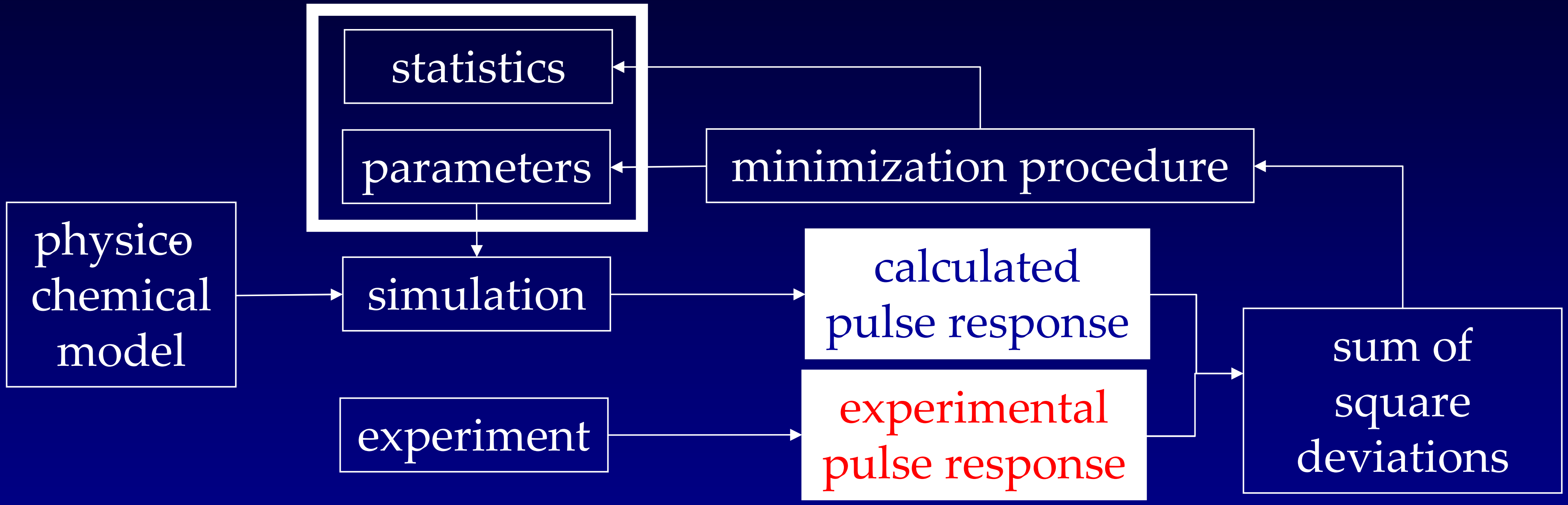


Overview

- Introduction
- Mechanism elucidation by regression
- Second-order statistical regression
- Numerical experiments
- Application
- Conclusions

Mechanism elucidation by regression

Validity of a postulated reaction mechanism verified quantitatively (fully reproducibly) through least-squares regression.



statistics

parameters

minimization procedure

physico-chemical model

simulation

calculated pulse response

experiment

experimental pulse response

sum of square deviations

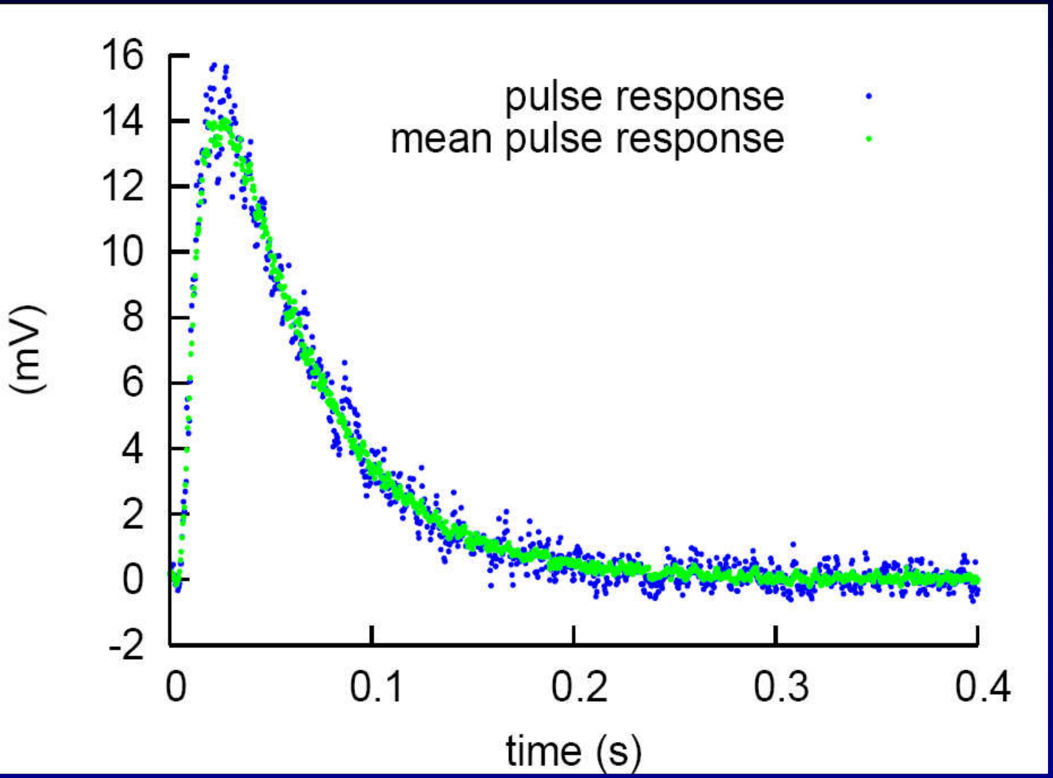
Physico-chemical model (including kinetic model involving assumptions about reaction mechanism) retained if

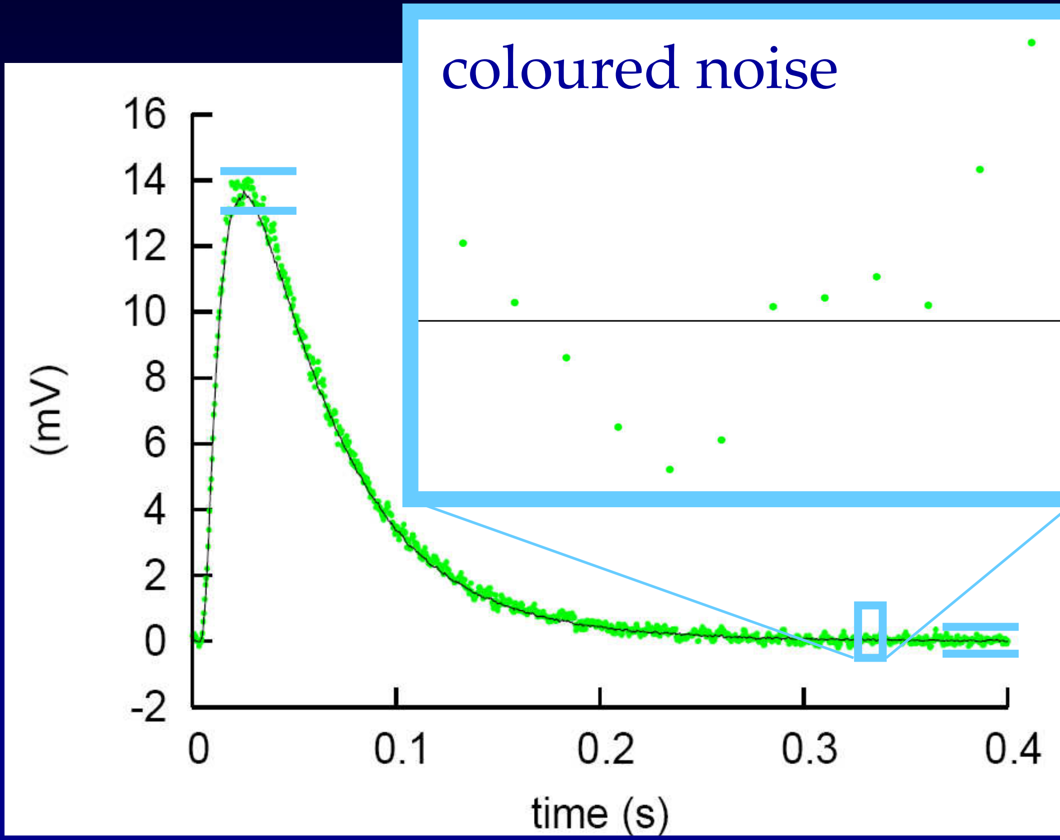
- all parameter estimates are physico-chemically meaningful;
- the regression is adequate, showing no lack of fit (F statistics);
- all estimates are significant (t statistics).

replicate data required

Mechanism elucidation by regression

Former strategy: Regression of the average of the replicate time series.





Least squares regression requires the errors to be

☐ Gaussian, ✓ (central limit theorem)
☐ Homoskedastic, ✗
☐ Uncorrelated. ✗

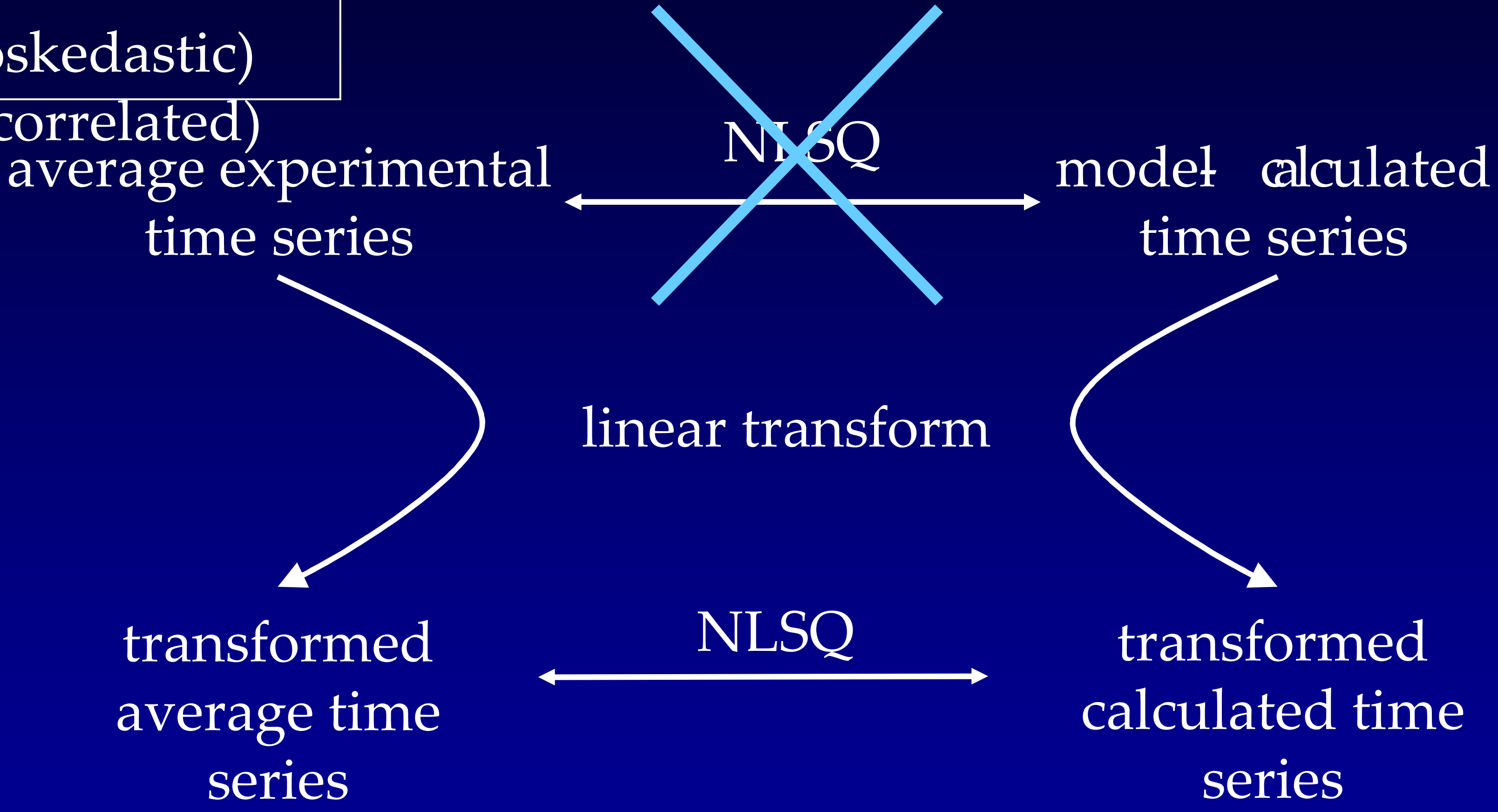
mechanism verification procedure not statistically sound!

Overview

- Introduction
- Mechanism elucidation by regression
- Second-order statistical regression
- Numerical experiments
- Application
- Conclusions

Second-order statistical regression

☑ G(aussian)
☒ H(omoskedastic)
☐ U(ncorrelated)



average experimental time series

~~NLSQ~~

model-calculated time series

linear transform

☑ G
☑ H
☑ U

transformed average time series

NLSQ

transformed calculated time series

Second-order statistical regression

☑ G
☒ H
☒ U

average experimental
time series

preconditioning transformation

☑ G
☒ H
☒ U

principal component analysis

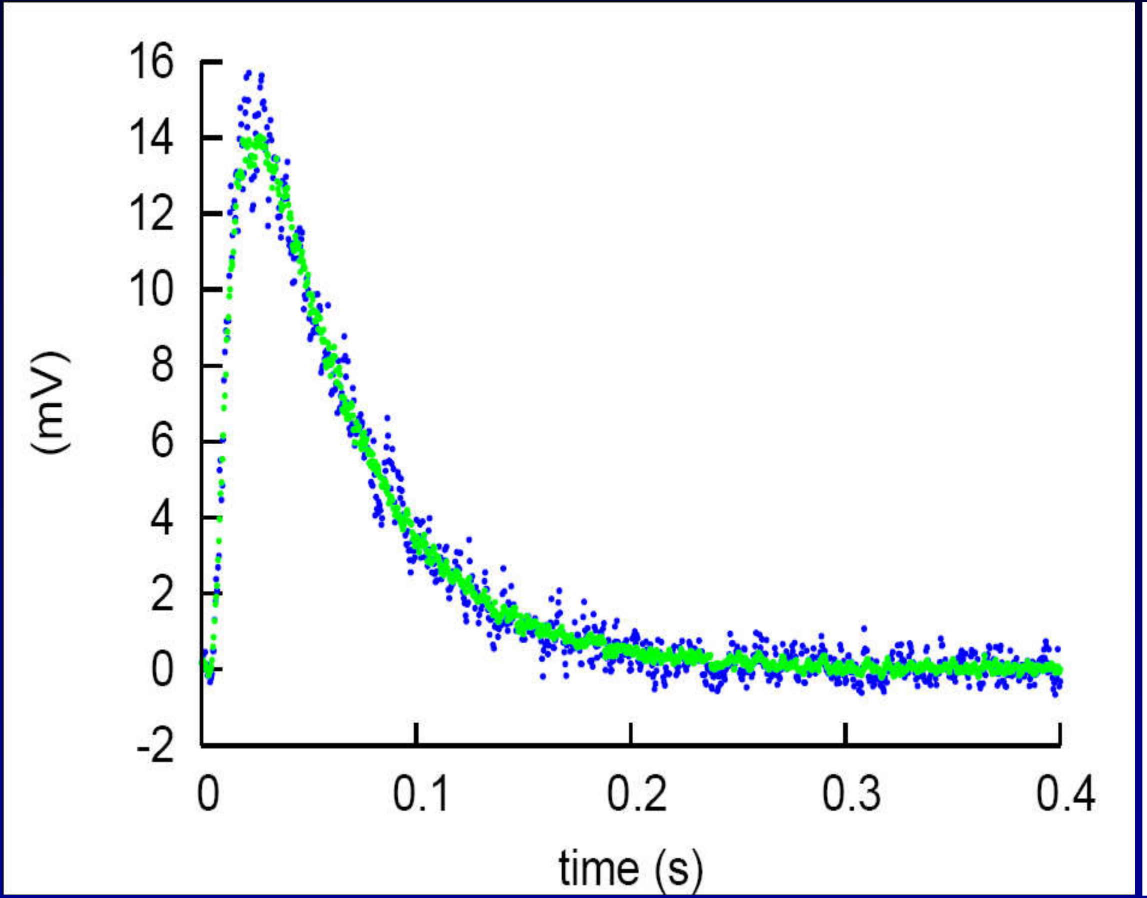
☑ G
☒ H
☑ U

rescaling

☑ G
☑ H
☑ U

transformed
average time
series

Second-order statistical regression



$$\mathbf{s}^{(r)} = \begin{bmatrix} s_1^{(r)} \\ s_2^{(r)} \\ s_3^{(r)} \\ \vdots \\ s_{n_t}^{(r)} \end{bmatrix} \qquad \bar{\mathbf{s}} = \begin{bmatrix} \bar{s}_1 \\ \bar{s}_2 \\ \bar{s}_3 \\ \vdots \\ \bar{s}_{n_t} \end{bmatrix}$$

r^{th} replicate average

Second-order statistical regression

$$\mathbf{D} = \begin{bmatrix} | & | & & | \\ \mathbf{s}^{(1)} & \mathbf{s}^{(2)} & \cdots & \mathbf{s}^{(n_r)} \\ | & | & & | \end{bmatrix}$$

data matrix

$$\mathbf{E} = \left[\begin{matrix} | \\ \mathbf{s}^{(1)} - \bar{\mathbf{s}} \\ | \end{matrix} \quad \boxed{\begin{matrix} | \\ \mathbf{s}^{(2)} - \bar{\mathbf{s}} \\ | \end{matrix}} \quad \cdots \quad \begin{matrix} | \\ \mathbf{s}^{(n_r)} - \bar{\mathbf{s}} \\ | \end{matrix} \right]$$

error matrix

estimated error (noise) in the 2nd replicate time series

Second-order statistical regression

Singular value decomposition of the error matrix: $\mathbf{E} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^T$

left singular vectors

positive singular values

right singular vectors

$$\mathbf{E} = \boxed{\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n_t} \\ | & | & & | \end{bmatrix}} \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{n_r-1} \\ \hdashline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \boxed{\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{n_t} \\ | & | & & | \end{bmatrix}^T}$$

$n_t \times n_t$ $\quad$ $n_t \times (n_r - 1)$ $\quad$ $(n_r - 1) \times (n_r - 1)$

The left and right singular vectors are orthonormal:

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_{n_r-1},\ \mathbf{V}^T\mathbf{V} = \mathbf{I}_{n_t}$$

The singular values are ordered:

$$s_1 > s_2 > \cdots > s_{n_r-1} > 0$$

Second-order statistical regression

It can be proven that the orthogonal projection of the n_t samples of the average time series on the first n_r–1 left singular vectors $\mathbf{u}_i$ of the error matrix **E** yields n_r–1 new numbers, (sample) principal components, expected to be uncorrelated:

$$\operatorname{cov}\left(\mathbf{u}_i^T \overline{\mathbf{s}}, \mathbf{u}_j^T \overline{\mathbf{s}}\right) \approx \begin{cases} \dfrac{s_i^2}{\left(n_r - 1\right)^2} & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

As $s_1 > s_2 > \cdots > s_{n_r - 1} > 0$, apparently,
most of the error is found to be parallel to $\mathbf{u}_1$, next parallel to $\mathbf{u}_2$, etc.

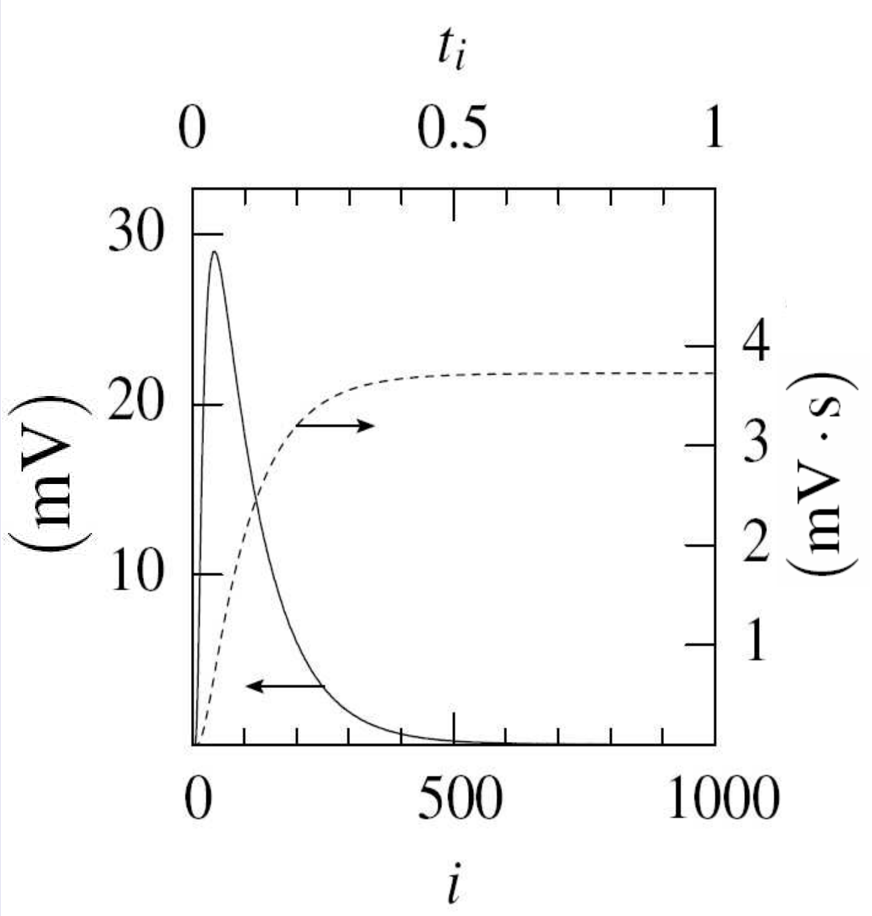
Second-order statistical regression

PCA: important dimensionality decrease from n_t to $n_r - 1$ →

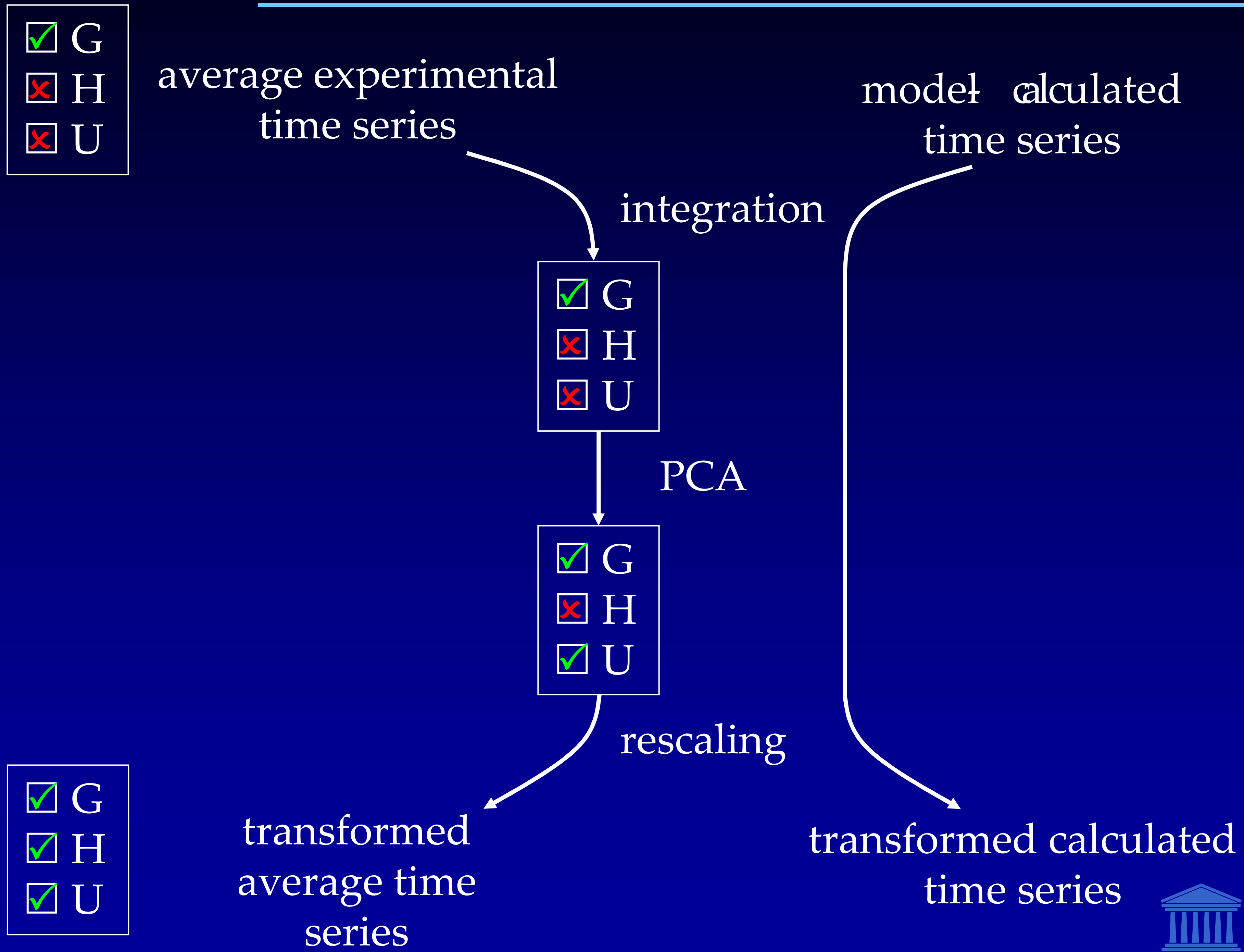
- no information lost w.r.t. the error
- quite some information lost w.r.t. the signal

To limit this loss to a minimum: preconditioning transformation:

(Discrete) integration of the time series



Second-order statistical regression



☑ G
☒ H
☒ U

average experimental
time series

model-calculated
time series

integration

☑ G
☒ H
☒ U

PCA

☑ G
☒ H
☑ U

rescaling

☑ G
☑ H
☑ U

transformed
average time
series

transformed calculated
time series

UNIVERSITEIT
GENT

Overview

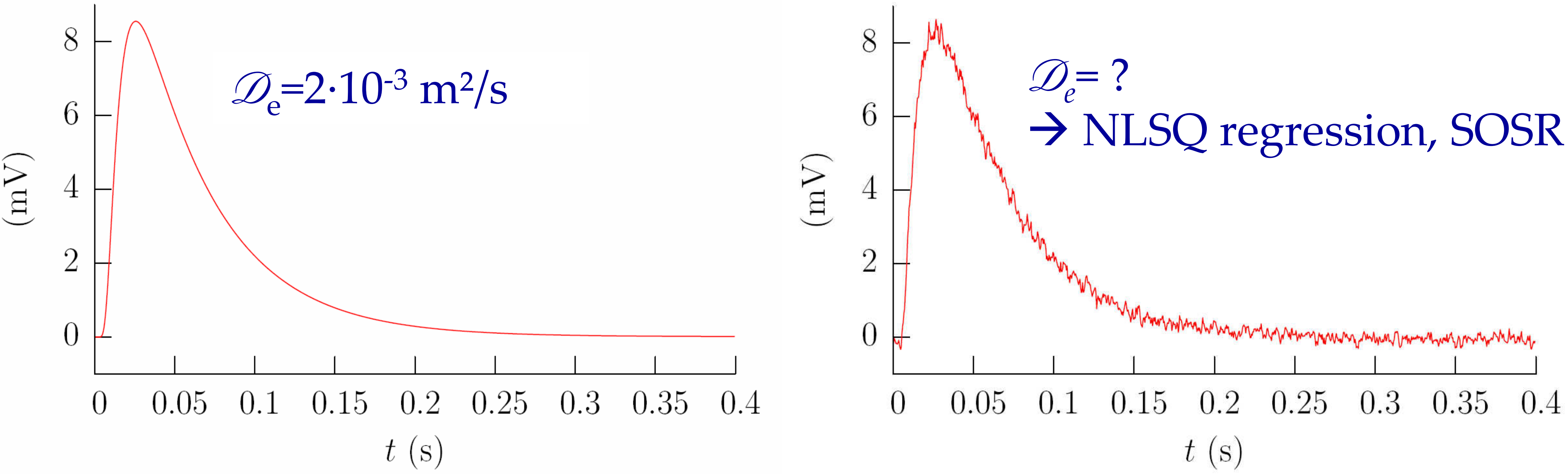
- Introduction
- Mechanism elucidation by regression
- Second-order statistical regression
- **Numerical experiments**
- Application
- Conclusions

UNIVERSITEIT GENT

Numerical experiments

1. Single-response experiment.

Simulation of a simple diffusion TAP experiment.



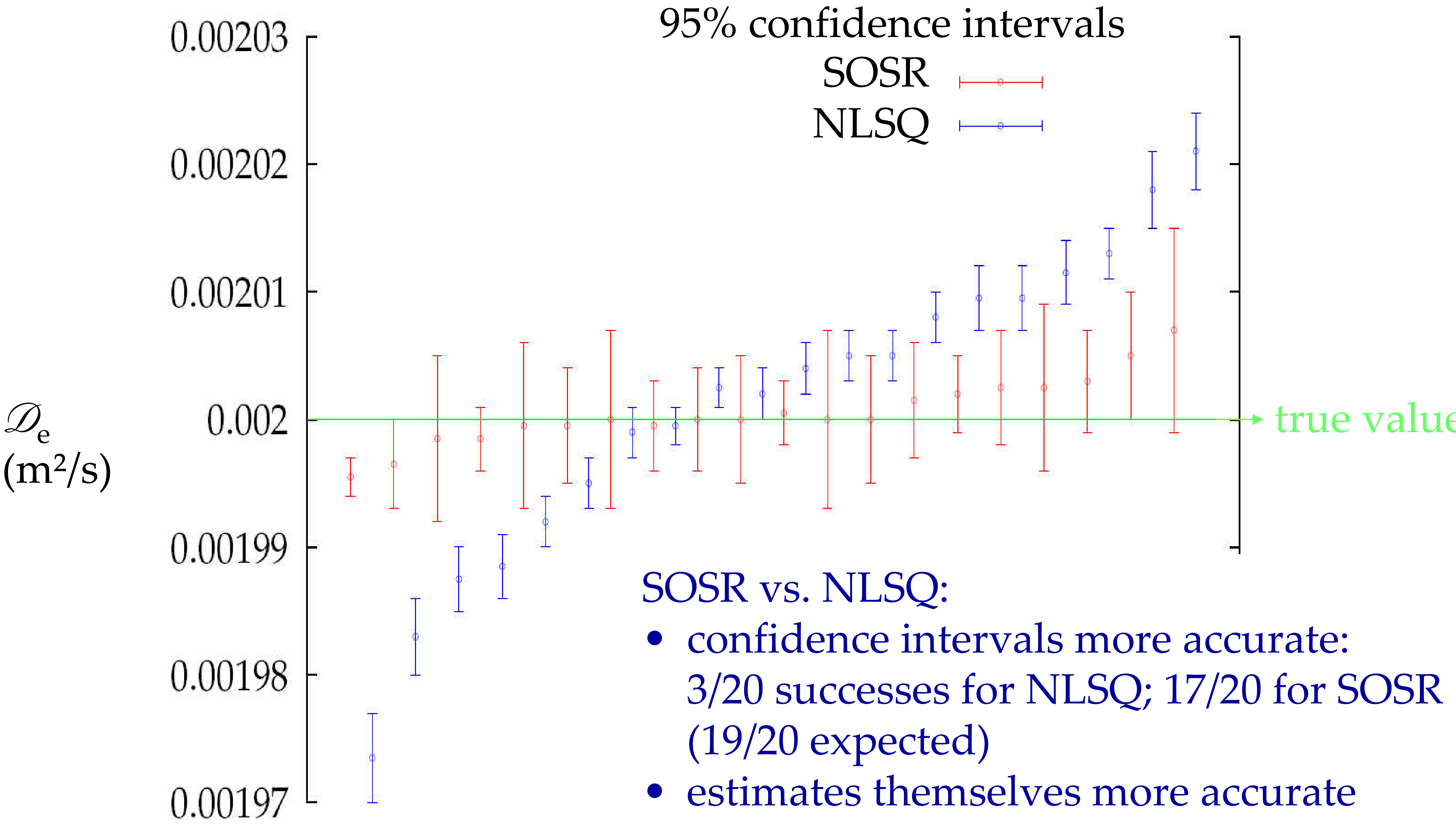
$\mathcal{D}_e = 2 \cdot 10^{-3}$ m²/s

$\mathcal{D}_e = ?$

→ NLSQ regression, SOSR

Typical error superposed in 20 × 20 replicates: autocorrelated Gaussian noise, 50 Hz oscillation with variable amplitude, random baseline shift, intensity variability.

Numerical experiments



SOSR vs. NLSQ:

- confidence intervals more accurate: 3/20 successes for NLSQ; 17/20 for SOSR (19/20 expected)
- estimates themselves more accurate

Numerical experiments

2. Multiple-response experiment

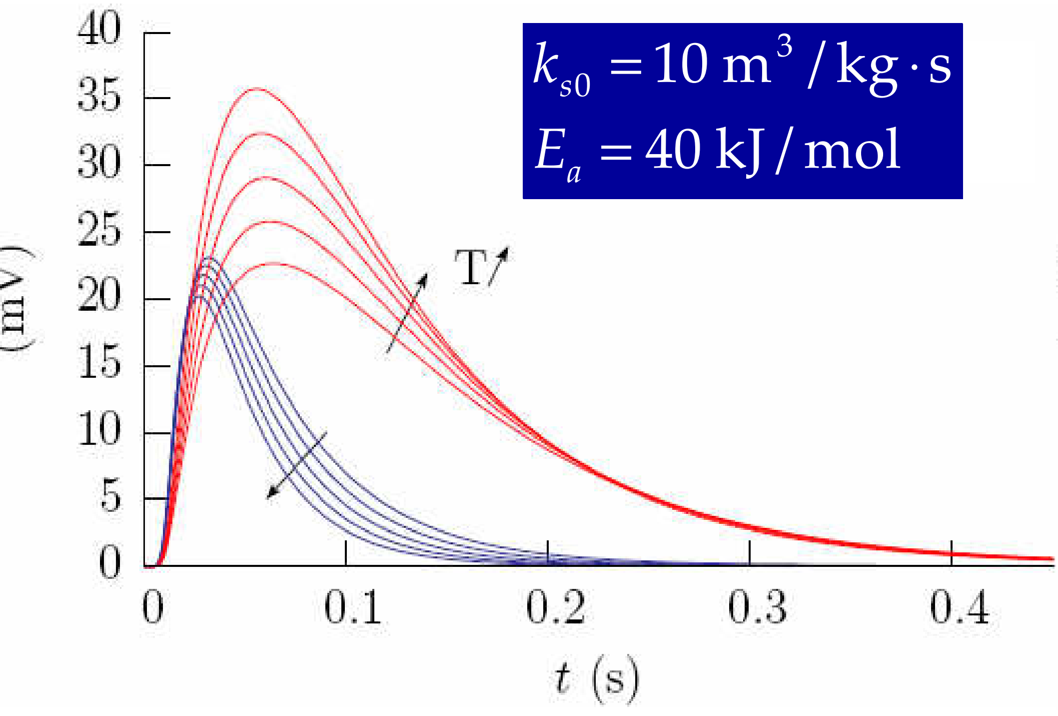
Simulation of a three-zone TAP-experiment. Propane is fed and is subject to total oxidation in the central, active, zone.

$$\boxed{C_3H_8 + 10\,O^* \xrightarrow{k} 3\,CO_2 + 4\,H_2O + 10\,^*}$$

$$r_s = k_s\, C_{C_3H_8}\,, \quad k_s = k_{s0} \cdot e^{-E_a/R\cdot T}$$

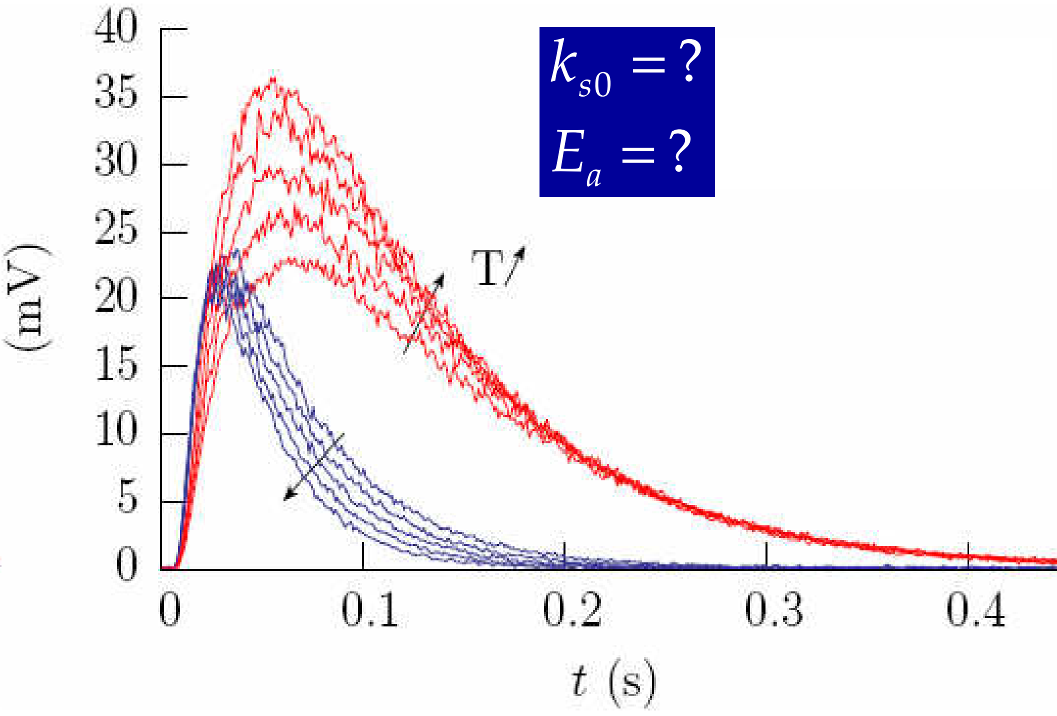
Numerical experiments

Responses from simulation at 450 °C, 475 °C, 500 °C, 525 °C and 550 °C, contaminated with typical TAP noise in 100 × 20 replicates.



$k_{s0} = 10\ \mathrm{m}^3/\mathrm{kg}\cdot\mathrm{s}$

$E_a = 40\ \mathrm{kJ/mol}$



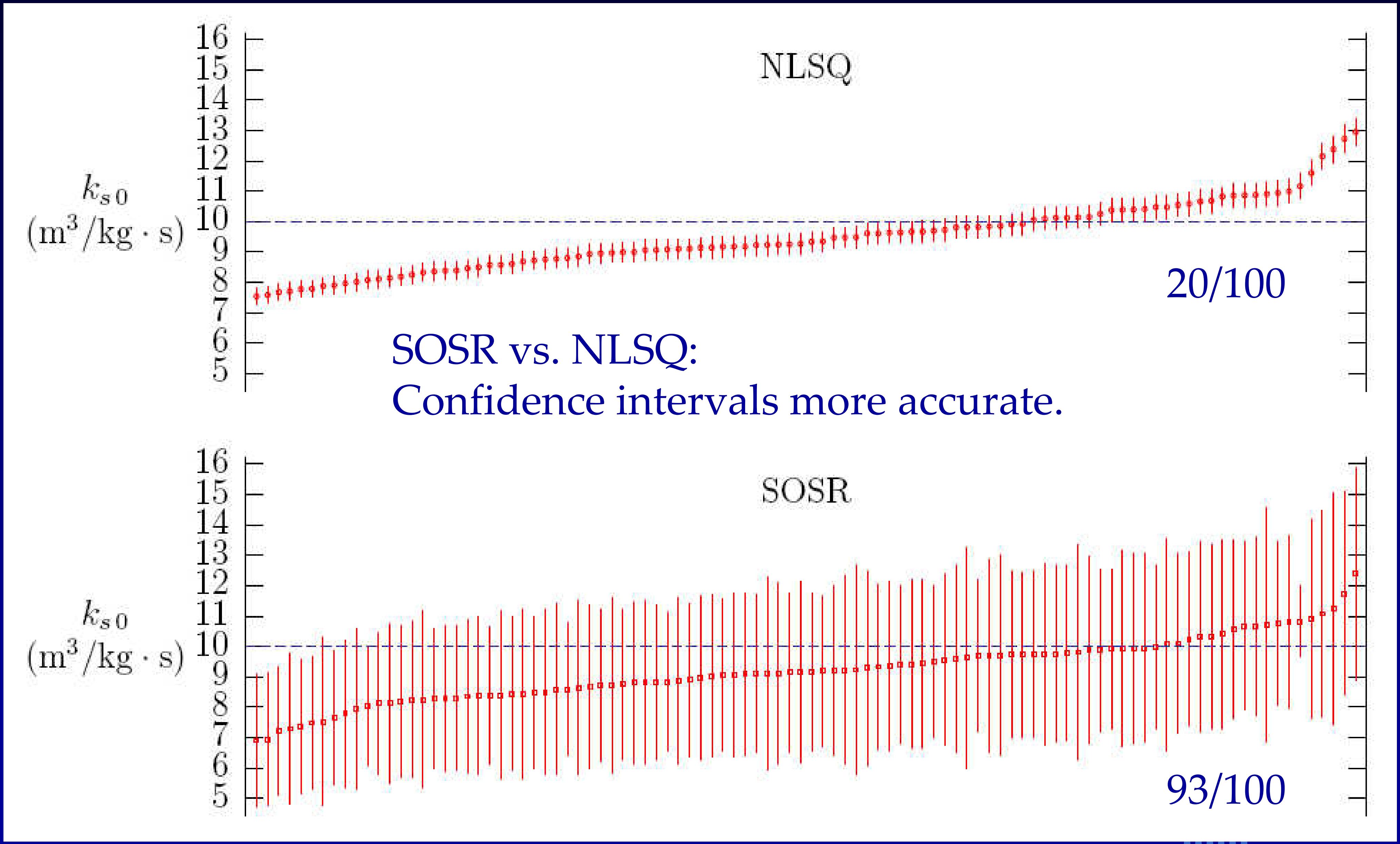
$k_{s0} = ?$

$E_a = ?$

→ NLSQ regression, SOSR

Numerical experiments

95 % confidence intervals



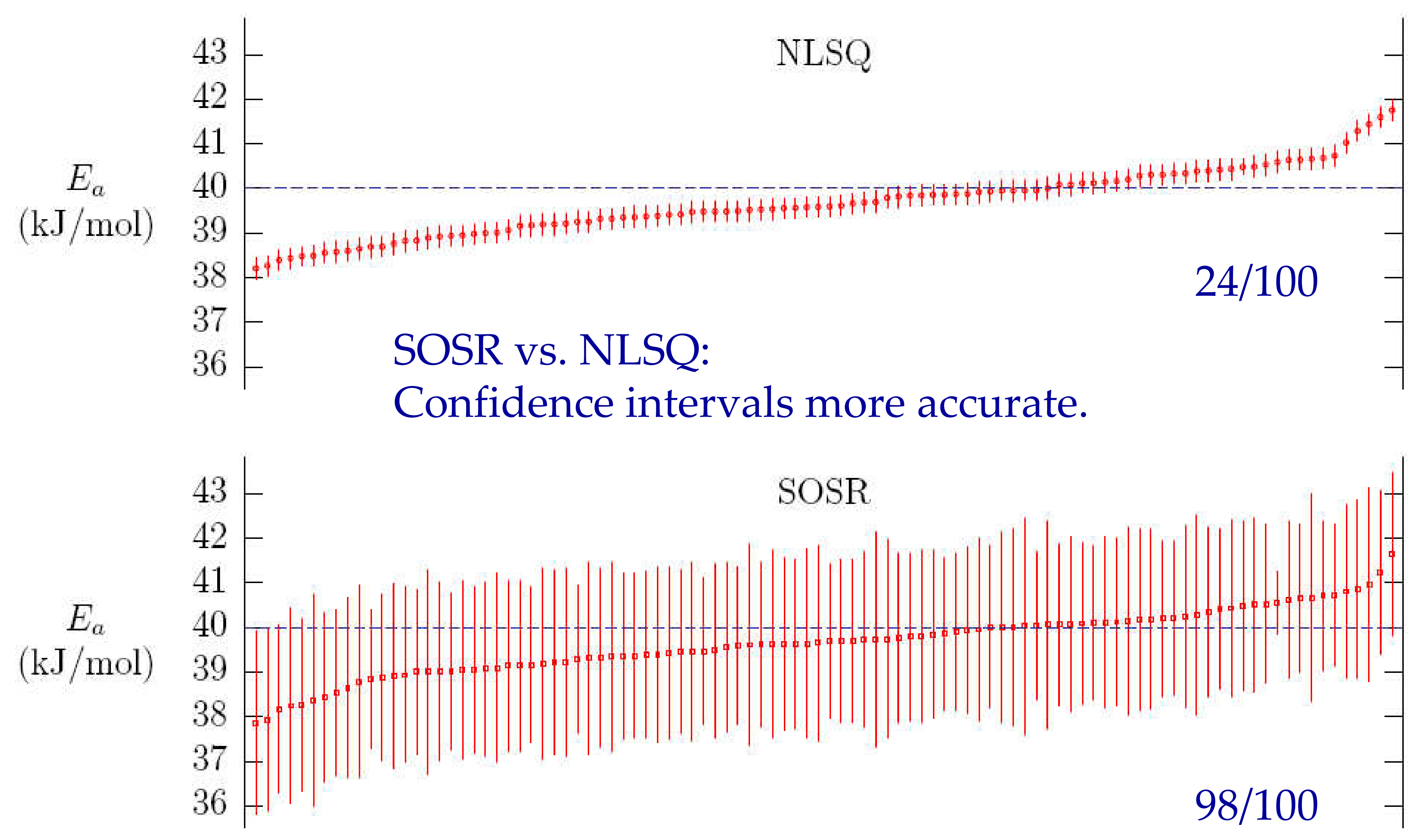
20/100

SOSR vs. NLSQ:
Confidence intervals more accurate.

93/100

Numerical experiments

95 % confidence intervals



24/100

SOSR vs. NLSQ:
Confidence intervals more accurate.

98/100

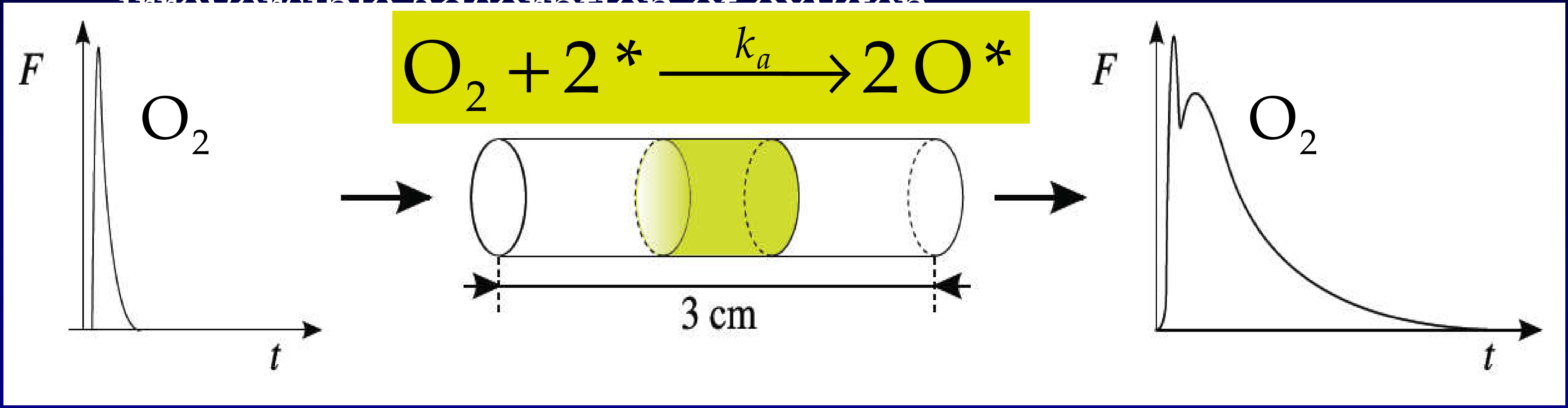
Overview

- Introduction
- Mechanism elucidation by regression
- Second-order statistical regression
- Numerical experiments
- Application
- Conclusions

UNIVERSITEIT GENT

Application

TAP-reactor with a central zone of V_2O_5/TiO_2 on quartz:
irreversible adsorption of oxygen

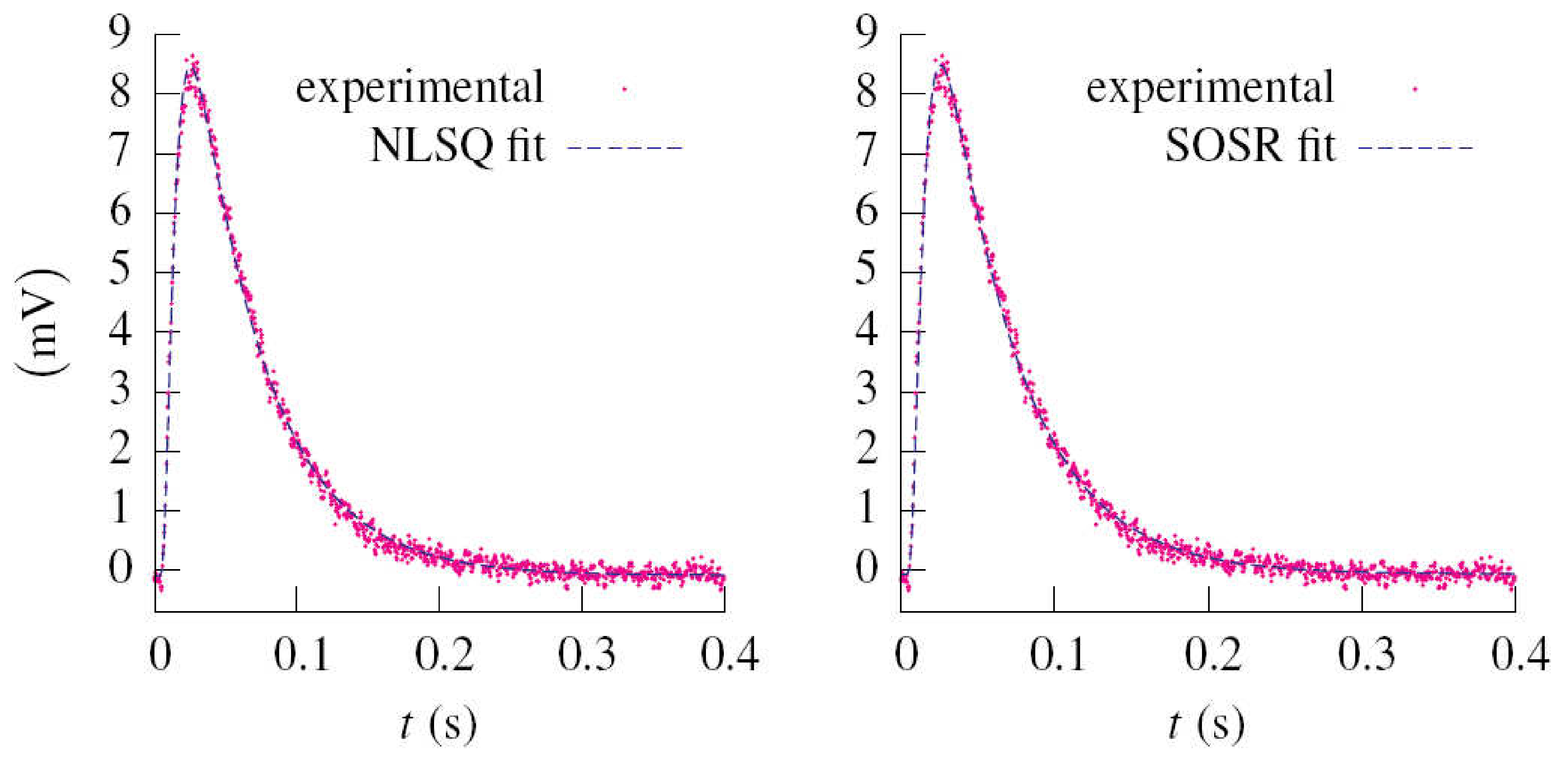


$$O_2 + 2\,* \xrightarrow{k_a} 2\,O*$$

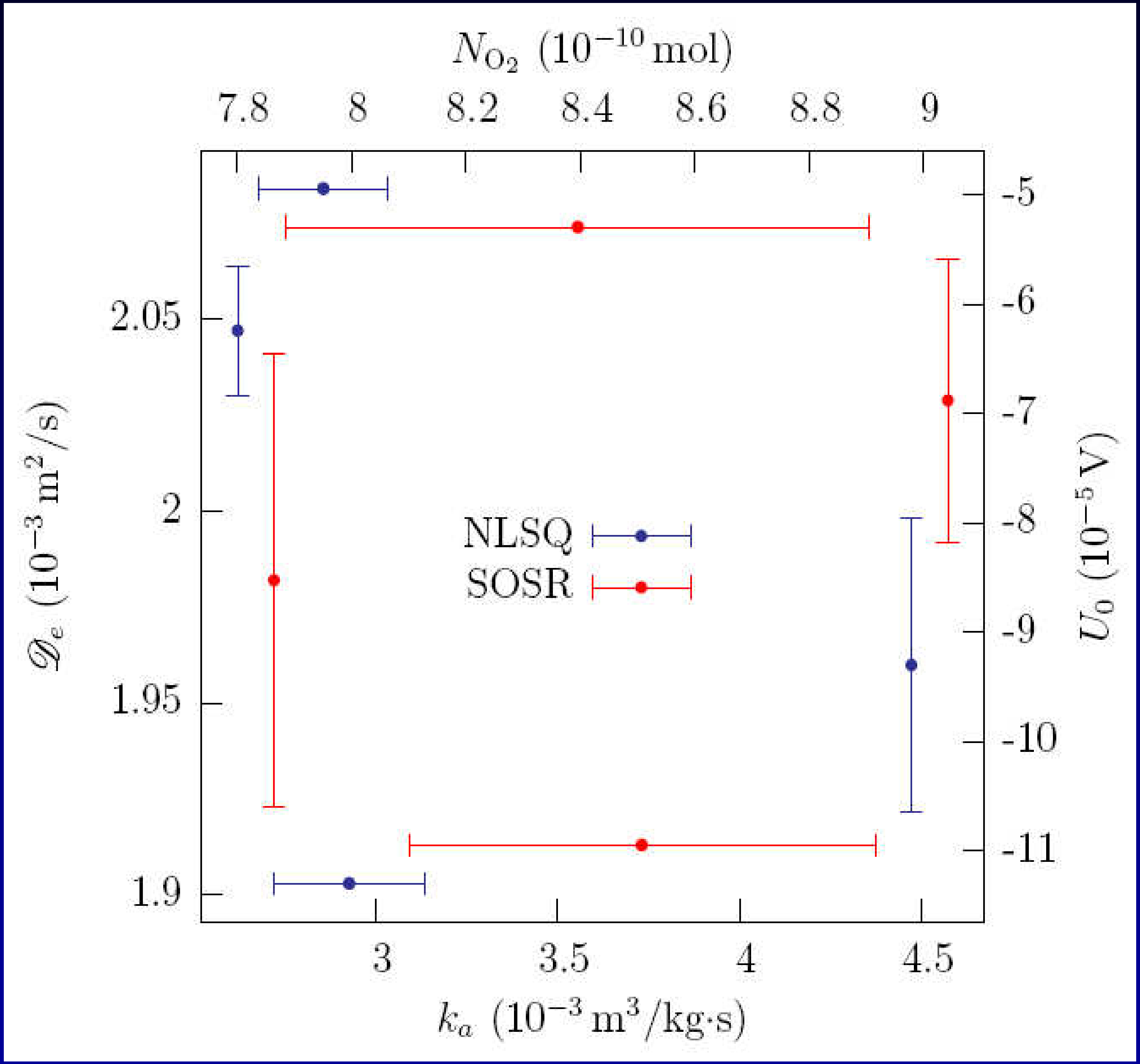
Parameters estimated:

k_a	specific adsorption rate coefficient ($m^3/kg \cdot s$)
$\mathscr{D}_e$	Knudsen diffusion coefficient (m^2/s)
N_{O2}	intensity of the pulse (mol)
U_0	baseline position (V)

Application

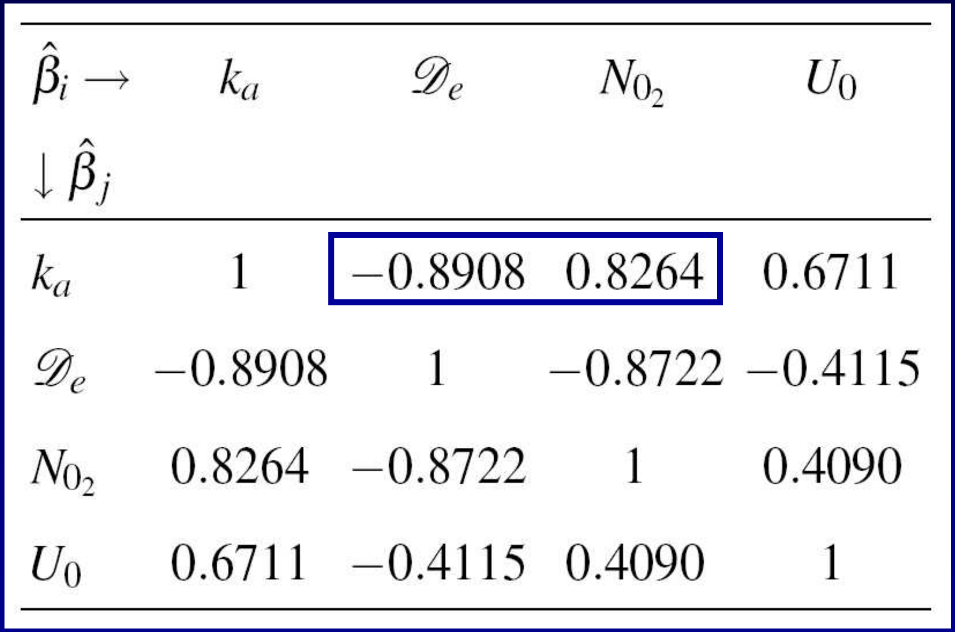


Application



Application

correlation matrix



$\hat{\beta}_i \rightarrow$ $\downarrow \hat{\beta}_j$	k_a	$\mathscr{D}_e$	N_{0_2}	U_0
k_a	1	−0.8908	0.8264	0.6711
$\mathscr{D}_e$	−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

Overview

- Introduction
- Mechanism elucidation by regression
- Second-order statistical regression
- Numerical experiments
- Application
- Conclusions

Conclusions

A second-order statistical regression was developed to regress time series from transient kinetic experiments with heteroskedastic and coloured noise.

Results: compared to NLSQ, increased accuracy of

- parameter estimates
- statistical information coming with the estimates

→ elucidation of reaction mechanism

Acknowledgments

Research Foundation – Flanders,
Special Research Fund of Ghent University