



# Extent-based kinetic identification using spectroscopic measurements and multivariate calibration<sup>☆</sup>



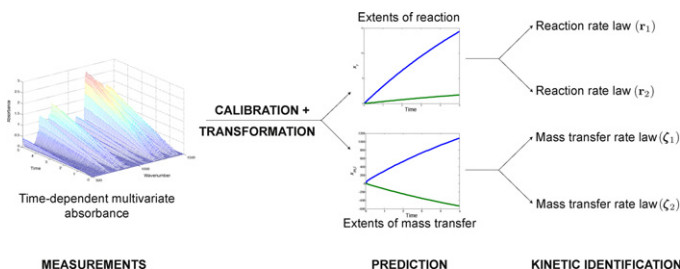
Julien Billeter<sup>\*</sup>, Sriniketh Srinivasan, Dominique Bonvin

Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne, Switzerland<sup>1</sup>

## HIGHLIGHTS

- ▶ Factorization of spectral data in RV- and RMV-forms.
- ▶ Prediction of extents of reaction and mass transfer from spectroscopic measurements.
- ▶ Reconstruction of all concentrations from the extents.
- ▶ Extent-based identification of each rate law independently of other rate processes.
- ▶ Illustration using simulated data from homogeneous and gas–liquid reaction systems.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 17 August 2012  
 Received in revised form  
 17 December 2012  
 Accepted 18 December 2012  
 Available online 28 December 2012

### Keywords:

Gas–liquid reactions  
 Reaction kinetics  
 Extents of reaction  
 Extents of mass transfer  
 Multivariate calibration  
 Spectroscopy

## ABSTRACT

Extent-based kinetic identification is a kinetic modeling technique that uses concentration measurements to compute extents and identify reaction kinetics by the integral method of parameter estimation. This article considers the case where spectroscopic data are used together with a calibration model to predict concentrations. The calibration set is assumed to be constructed from reacting data that include pairs of concentration and spectral data. Alternatively, one can use the concentration- and spectral contributions of the reactions and mass transfers, which are obtained by pretreatment in reaction- and mass-transfer-variant form. The extent-based kinetic identification using concentrations predicted from spectroscopic data is illustrated through the simulation of both a homogeneous and a gas–liquid reaction system.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

Dynamic models of reaction systems are very valuable for the monitoring, control and optimization of industrial chemical

processes [1]. Reaction models are typically built on first principles and describe the states (concentrations, volume and temperature) of the reaction system by means of differential equations expressing the conservation of mass and energy [2]. The identification of kinetic laws often represents the most challenging task of reaction modeling. This challenge is even bigger for gas–liquid systems because of the coupling between reactions and mass transfers.

The identification of kinetic models for complex reaction systems can be efficiently performed in an incremental way. Incremental identification involves two steps [3]: (i) the computation of the contribution of each reaction and each mass transfer in the

<sup>☆</sup> Paper presented at the XIII Conference on Chemometrics in Analytical Chemistry (CAC 2012), Budapest, Hungary, 25–29 June 2012.

<sup>\*</sup> Corresponding author.

E-mail addresses: [julien.billeter@epfl.ch](mailto:julien.billeter@epfl.ch) (J. Billeter), [sriniketh.srinivasan@epfl.ch](mailto:sriniketh.srinivasan@epfl.ch) (S. Srinivasan), [dominique.bonvin@epfl.ch](mailto:dominique.bonvin@epfl.ch) (D. Bonvin).

<sup>1</sup> <http://la.epfl.ch>.

**List of symbols***Scalars*

$\lambda$	dimensionless discounting variable
$a$	specific interfacial area [1/distance]
$H$	number of calibration pairs
$H_s$	Henry constant of the species $s$
$k_i$	rate constant of the $i$ th reaction
$k_s$	molar transfer coefficient of the species $s$ [distance/time]
$m$	mass
$m_0$	initial mass
$P$	total pressure
$p$	number of independent inlet streams
$p_m$	number of mass transfers
$R$	number of independent reactions
$S$	number of species
$T$	temperature
$t$	time
$u$	mass flow rate [mass/time]
$V$	volume
$V_{tot}$	volume of the gas–liquid reactor
$V_0$	initial volume
$W$	number of wavelengths/wavenumbers
$\varepsilon$	number of latent variables
$\mathcal{R}$	ideal gas constant

*Vectors*

$\delta_m$	$p_m$ -dimensional vector of auxiliary variables [mass]
$\zeta$	$p_m$ -dimensional vector of mass transfer rates [mass/time]
$\mathbf{c}$	$S$ -dimensional vector of concentrations [mole/volume]
$\mathbf{c}_0$	$S$ -dimensional vector of initial concentrations [mole/volume]
$\mathbf{n}$	$S$ -dimensional vector of numbers of moles
$\mathbf{n}_0$	$S$ -dimensional vector of initial numbers of moles
$\mathbf{p}$	$S$ -dimensional vector of partial pressures
$\mathbf{q}_0$	$S$ -dimensional projection vector
$\mathbf{r}$	$R$ -dimensional vector of reaction rates [mole/(volume time)]
$\mathbf{u}$	$p$ -dimensional vector of mass flow rates [mass/time]
$\mathbf{x}_r$	$R$ -dimensional vector of extents of reaction [mole]
$\mathbf{x}_{in}$	$p$ -dimensional vector of extents of inlet [mass]
$\mathbf{x}_m$	$p_m$ -dimensional vector of extents of mass transfer [mass]
$\boldsymbol{\theta}$	vector of adjustable parameters
$\mathbf{a}$	$W$ -dimensional absorbance spectrum
$\mathbf{a}_0$	$W$ -dimensional initial absorbance spectrum

*Matrices*

$\mathbf{0}$	vector or matrix of zeros
$\mathbf{1}$	vector or matrix of ones
$\mathbf{A}_{cal}$	$(H \times W)$ -dimensional matrix of calibration spectra
$\mathbf{A}_{in}$	$(p \times W)$ -dimensional inlet spectra
$\mathbf{A}_m$	$(p_m \times W)$ -dimensional mass-transfer spectra
$\mathbf{C}_{cal}$	$(H \times S_c)$ -dimensional matrix of calibration concentrations
$\mathbf{N}$	$(R \times S)$ -dimensional matrix of stoichiometry
$\mathbf{M}_{in,0}$	$(S \times p)$ -dimensional projection matrix
$\mathbf{M}_{m,0}$	$(S \times p_m)$ -dimensional projection matrix
$\mathbf{M}_w$	$S$ -dimensional diagonal matrix of molecular weights [mass/mole]

$\mathbf{E}$	$(S_a \times W)$ -dimensional matrix of pure component spectra
$\check{\mathbf{E}}_m$	$(S \times p_m)$ -dimensional matrix of transferring species
$\mathbf{F}$	$(S_c \times W)$ -dimensional calibration matrix
$\mathbf{S}$	$\varepsilon$ -dimensional diagonal matrix of principal component factors
$\mathbf{S}_0$	$(S \times R)$ -dimensional projection matrix
$\mathbf{U}$	$(H \times \varepsilon)$ -dimensional matrix of principal component vectors
$\bar{\mathbf{V}}$	$(W \times \varepsilon)$ -dimensional matrix of principal component vectors
$\mathbf{W}_{in}$	$(S \times p)$ -dimensional matrix of inlet composition [mole/mass]
$\check{\mathbf{W}}_{in}$	$(S \times p)$ -dimensional matrix of inlet weight fractions
$\mathbf{W}_m$	$(S \times p_m)$ dimensional matrix of mass transfer [mole/mass]

*Operators and other symbols*

CAL	construction of calibration model
$F$	phase, $F \in \{G, L\}$
$\text{Im}(\cdot)$	image (row space) of a matrix
$\mathcal{L}$	linear transformation to extents
$\mathfrak{R}$	reconstruction of concentrations from extents
$\text{rank}(\cdot)$	rank of a matrix
RMV	transformation to RMV-form
RV	transformation to RV-form
$\mathbb{S}$	set of experimental conditions

*Subscripts*

$(\cdot)_a$	related to absorbing species
$(\cdot)_c$	related to calibrated species
$(\cdot)_{cal}$	related to calibration experiment
$(\cdot)_f$	related to $f$ phase, $f \in \{g, l\}$
$(\cdot)_g$	related to gas phase
$(\cdot)_{in}$	related to inlet flows
$(\cdot)_l$	related to liquid phase
$(\cdot)_m$	related to mass transfers
$(\cdot)_{out}$	related to outlet flows
$(\cdot)_r$	related to reactions
$(\cdot)_u$	related to uncalibrated species

*Superscripts*

$(\cdot)$	time derivative
$(\hat{\cdot})$	predicted quantity
$(\cdot)^*$	equilibrium quantity
$(\cdot)^+$	Moore–Penrose pseudo-inverse of a matrix
$(\cdot)^{MV}$	mass-transfer-variant form
$(\cdot)^{RMV}$	reaction- and mass-transfer-variant form
$(\cdot)^{RV}$	reaction-variant form
$(\cdot)^T$	transpose of a vector or matrix

form of a rate or an extent and (ii) the identification of the corresponding rate law and rate parameters by comparing computed and modeled rates or extents. Owing to the decoupling of the various rates and extents, each reaction and each mass transfer can be dealt with individually, unlike in the simultaneous identification, where an overall model including all the reactions and mass transfers has to be postulated [4–6]. Hence, the reaction and mass-transfer rate parameters estimated by the incremental methods are largely uncorrelated. Note that the incremental methods require that there are at least as many measurements as there are rates and extents to be computed. This requirement can be difficult to meet for systems with numerous reactions. The kinetic modeling

of biological processes, for example, requires adapting the incremental methods to model a range of feasible rates and extents of reaction [7].

Two different incremental methods have been proposed in the literature. The differential method involves the computation of reaction and mass transfer *rates* by differentiating measured time-dependent profiles [8]. In the integral method, the reaction and mass-transfer contributions are computed as *extents* of reaction and mass transfer, which are obtained via algebraic transformation – and not differentiation – of measured data [9,10]. Depending on whether the contributions are expressed as rates or extents, these methods are labeled ‘rate-based’ or ‘extent-based’ identification methods. Although computationally less expensive, the rate-based method has the drawback of introducing a bias due to the numerical differentiation step [6].

The extent-based identification method has been developed in the context of measured concentrations. Other sources of information, such as gas consumption [10] and calorimetry [11], can be added to augment the rank of the data matrix, that is, to increase the number of measured quantities. In practice, concentration measurements require sampling, sample preparation and subsequent offline analysis, which is expensive in terms of both cost and time. In comparison, the recent development of Process Analytical Technologies (PAT) has opened up new avenues for exploiting spectroscopic data, which indirectly provide non-destructive concentration measurements of the species in-situ/online during the course of a reaction [12]. Accurate spectroscopic measurements are available at low cost and short sampling times, and are free of any delay. In addition, most spectroscopic measurements can be considered linear with respect to the concentrations of the absorbing species (Beer’s law). Note that the technology of spectrometers and fiber-optic probes has improved significantly in recent years. Furthermore, numerical methods facilitate the analysis of large sets of multivariate data. For these reasons, online spectroscopic measurements are nowadays widely used, probably more so than direct concentration measurements [13,14].

In absorption spectroscopy, calibration models based on Principal Component Regression (PCR) [15] or Partial Least Squares (PLS) [16] are used to predict the concentrations of absorbing species from an absorbance spectrum measured in a sample [17,18]. Calibration involves two steps. In the calibration step, the calibration model is built using pairs of measured concentrations and absorbance spectra. This construction requires defining a certain number of latent variables that represent abstract building blocks. In the prediction step, the calibration model is applied to the absorbance spectrum measured in a new sample to predict the concentrations of the calibrated species.

Since calibration models require that the new spectrum lies in the space spanned by the calibration set (space-inclusion condition), it is important to properly design the calibration set [19]. This condition implies that the experimental conditions used for the calibration set should mimic as much as possible the experimental conditions of the prediction experiment, inter alia, in terms of concentration and temperature ranges [20,21]. One approach consists in preparing calibration samples of various independent compositions and measuring the absorbance spectrum of each sample. The resulting non-reacting calibration data are sometimes qualified as ‘static’ in the sense that the composition of each calibration sample does not vary with time [22]. This way of constructing the calibration set has the drawback that it requires numerous calibration mixtures and that none of the calibration samples contains the short-lived absorbing intermediates that are often found in real samples taken during the course of a reaction. This can cause a more or less severe violation of the space-inclusion condition. An alternative approach consists in taking calibration samples during the course of a preliminary

experiment, based on the chemical process at hand, and measuring pairs of concentrations and absorbance spectra for each of these samples. With such a design, the intermediate species that are produced by the reactions are included, and the need to vary them externally disappears [19]. Such a ‘dynamic’ calibration system is particularly suited for industry, where reference samples can be taken during the course of the reaction and used for calibration.

When reacting calibration data are used, the calibration model can be constructed on two types of calibration pairs. A first possibility consists in taking measured concentrations and absorbance spectra as calibration pairs; in such a case, the calibration model predicts concentrations from a new spectrum. Alternatively, the calibration model can be constructed on pairs of concentration and spectroscopic measurements containing only the reaction and mass-transfer contributions, that is, the contribution of the inlet and outlet flows is removed from the measurements. The calculation of these pairs requires the pretreatment of concentrations and spectroscopic data in so-called *reaction- and mass-transfer-variant (RMV) form*. If gas-phase measurements can be used to compute the concentration and spectroscopic contributions of the mass transfers, these contributions can be removed, thus resulting in calibration pairs in *reaction-variant (RV) form*. A calibration model in RMV- or RV-form uses a new spectrum also pretreated in that form to predict concentrations in the same form. The pros and cons of using these forms for calibration have been discussed in [19].

In this study, the concentrations are predicted from measured absorbance spectra via multivariate calibration. The calibration model is constructed from reacting data collected in a previous experiment. The predicted concentrations subsequently feed the incremental extent-based identification method. Kinetic modeling based on spectroscopic data using this combination of calibration and identification methods proceeds in four steps: (i) in the *calibration step*, a calibration model is constructed from pairs of measured concentrations and absorbance spectra, (ii) in the *prediction step*, the calibration model is used to predict the concentrations of the calibrated species from a new absorbance spectrum, (iii) in the *transformation step*, the predicted concentrations are transformed into extents, and (iv) in the *identification step*, each extent is used individually to model the corresponding rate expressions.

The paper is organized as follows. Section 2 reviews the extent-based identification of reaction kinetics from measured concentrations. Section 3 discusses the extensions that are necessary to predict the extents and all concentrations from spectroscopic measurements. Although all developments are proposed for the general case of gas–liquid reaction systems, a short remark is provided at the end of each subsection indicating the simplified form that corresponds to the case of homogeneous reaction systems. Section 4 illustrates the proposed kinetic identification approach via the simulation of two case studies, namely, the acetoacetylation of pyrrole in an open homogeneous stirred-tank reactor and the chlorination of butanoic acid in an open gas–liquid reactor. Finally, Section 5 concludes the paper.

## 2. Extent-based kinetic identification from measured concentrations

This section describes the extent-based kinetic identification approach. The entry point is a generic model consisting of the differential mole balance equations for the system. Next, a linear transformation computes the extents of reaction and mass transfer from measured concentrations. Finally, the rate expressions and corresponding rate parameters are identified from the computed extents.

## 2.1. Mole balance equations

Consider a general reaction system comprised of the two phases  $G$  (gas phase) and  $L$  (liquid phase), with  $S_g$  species in the gas phase and  $S_l$  species in the liquid phase.<sup>2</sup> The two phases are connected by  $p_m$  mass transfers involving  $S_m$  species transferring between the phases. Let  $p_g$  and  $p_l$  be the numbers of inlets in the phases  $G$  and  $L$ , respectively, and let each phase have one outlet. Assuming that (i) both phases are homogeneous, (ii) the reactions take place in the phase  $L$  only and (iii) mass-transfer phenomena can be described by the film theory with no accumulation in the film, the differential mole balance equations for the gas and liquid phases can be formulated as:

$$\dot{\mathbf{n}}_l(t) = \mathbf{N}^T V_l(t) \mathbf{r}(t) + \mathbf{W}_{m,l} \boldsymbol{\zeta}(t) + \mathbf{W}_{in,l} \mathbf{u}_{in,l}(t) - \frac{u_{out,l}(t)}{m_l(t)} \mathbf{n}_l(t), \quad \mathbf{n}_l(0) = \mathbf{n}_{l0} \quad (1a)$$

$$\dot{\mathbf{n}}_g(t) = -\mathbf{W}_{m,g} \boldsymbol{\zeta}(t) + \mathbf{W}_{in,g} \mathbf{u}_{in,g}(t) - \frac{u_{out,g}(t)}{m_g(t)} \mathbf{n}_g(t), \quad \mathbf{n}_g(0) = \mathbf{n}_{g0} \quad (1b)$$

where  $\mathbf{n}_f$  is the  $S_f$ -dimensional vector of numbers of moles in the  $F$  phase,  $f \in \{g, l\}$  and  $F \in \{G, L\}$ ,  $\mathbf{N}$  the  $R \times S_l$  stoichiometric matrix,  $V_l$  the volume of the liquid phase,  $\mathbf{W}_{in,f} = \mathbf{M}_{w,f}^{-1} \check{\mathbf{W}}_{in,f}$  the  $S_f \times p_f$  inlet matrix expressing the composition of the inlets to the  $F$  phase,  $\mathbf{M}_{w,f}$  the  $S_f$ -dimensional diagonal matrix of molecular weights, and  $\check{\mathbf{W}}_{in,f} = [\check{\mathbf{w}}_{in,f}^1 \cdots \check{\mathbf{w}}_{in,f}^{p_f}]$  with  $\check{\mathbf{w}}_{in,f}^k$  being the  $S_f$ -dimensional vector of weight fractions of the  $k$ th inlet to the  $F$  phase with  $k = 1, \dots, p_f$ ,  $\mathbf{u}_{in,f}$  the  $p_f$ -dimensional inlet mass flow rate to the  $F$  phase,  $u_{out,f}$  the outlet mass flow rate in the  $F$  phase,  $m_f$  the mass of the  $F$  phase, and  $\mathbf{n}_{f0}$  the vector of initial numbers of moles in the  $F$  phase. The  $p_m$  mass transfers are treated as pseudo inlets with unknown flow rates and involve  $\mathbf{W}_{m,f}$  and  $\boldsymbol{\zeta}$ , where  $\mathbf{W}_{m,f} = \mathbf{M}_{w,f}^{-1} \check{\mathbf{E}}_{m,f}$  is the  $S_f \times p_m$  mass-transfer matrix,  $\check{\mathbf{E}}_{m,f} = [\check{\mathbf{e}}_{m,f}^1 \cdots \check{\mathbf{e}}_{m,f}^{p_m}]$  with  $\check{\mathbf{e}}_{m,f}^j$  being the  $S_f$ -dimensional vector with the elements corresponding to the  $j$ th transferring species equal to unity and the other elements equal to zero, and  $\boldsymbol{\zeta}$  is the  $p_m$ -dimensional vector of mass-transfer rates expressed in mass/time units and defined positively (+) when the mass transfer occurs from  $G$  to  $L$ .

### Special case

For homogeneous reaction systems, there is a single phase and the differential mole balance equations are given by Eq. (1a), without the subscript  $(\cdot)_l$  and without the mass-transfer term  $\mathbf{W}_{m,l} \boldsymbol{\zeta}(t)$ .

## 2.2. Computation of extents from numbers of moles

The method to compute the extents from the measured numbers of moles has been described by Bonvin and co-workers [23,24]. The method involves a linear transformation that is applicable to all reaction systems for which the three assumptions described in Section 2.1 are valid. The standard linear transformation is summarized in the next subsection and then adapted to deal with the numbers of moles in RMV-, MV- and RV-forms in the subsequent subsections.

### 2.2.1. Standard transformation

This linear transformation decomposes the  $S_f$ -dimensional vector of numbers of moles in the  $F$  phase into the following contributions:  $R$  extents of reaction  $\mathbf{x}_r(t)$  (in mole units) in the liquid

phase, as well as  $p_m$  extents of mass transfer  $\mathbf{x}_{m,f}(t)$  (in mass units),  $p_f$  extents of inlet flow  $\mathbf{x}_{in,f}(t)$  (in mass units) and a scalar dimensionless discounting variable  $\lambda_f(t)$  in each phase. This transformation can be described mathematically as:

$$\begin{aligned} \mathbf{x}_r(t) &= \mathbf{S}_{l0}^T \mathbf{n}_l(t) \\ \begin{bmatrix} \mathbf{x}_{m,f}(t) \\ \mathbf{x}_{in,f}(t) \\ \lambda_f(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{M}_{m,f0}^T \\ \mathbf{M}_{in,f0}^T \\ \mathbf{q}_{f0}^T \end{bmatrix} \mathbf{n}_f(t) \end{aligned} \quad (2)$$

As the reactions are assumed to take place in the  $L$  phase, the term  $\mathbf{x}_r(t) = \mathbf{S}_{l0}^T \mathbf{n}_l(t)$  applies only to this phase. Eq. (2) is applicable under the conditions,  $\text{rank}([\mathbf{N}^T \mathbf{W}_{m,l} \mathbf{W}_{in,l} \mathbf{n}_{l0}]) = R + p_m + p_l + 1$  for the  $L$  phase and  $\text{rank}([\mathbf{W}_{m,g} \mathbf{W}_{in,g} \mathbf{n}_{g0}]) = p_m + p_g + 1$  for the  $G$  phase. The procedure for constructing the matrices  $\mathbf{S}_{l0}$  ( $S_l \times R$ ),  $\mathbf{M}_{m,f0}$  ( $S_f \times p_m$ ),  $\mathbf{M}_{in,f0}$  ( $S_f \times p_f$ ), and the  $S_f$ -dimensional vector  $\mathbf{q}_{f0}$  has been outlined in [23].

Application of the linear transformation (2) to  $\mathbf{n}_f(t)$  transforms the differential mole balance equations (1a) and (1b) in the following system of differential equations:

$$\dot{\mathbf{x}}_r(t) = V_l(t) \mathbf{r}(t) - \frac{u_{out,l}(t)}{m_l(t)} \mathbf{x}_r(t), \quad \mathbf{x}_r(0) = \mathbf{0}_R, \quad (3a)$$

$$\dot{\mathbf{x}}_{m,f}(t) = \boldsymbol{\zeta}(t) - \frac{u_{out,f}(t)}{m_f(t)} \mathbf{x}_{m,f}(t), \quad \mathbf{x}_{m,f}(0) = \mathbf{0}_{p_m}, \quad (3b)$$

$$\dot{\mathbf{x}}_{in,f}(t) = \mathbf{u}_{in,f}(t) - \frac{u_{out,f}(t)}{m_f(t)} \mathbf{x}_{in,f}(t), \quad \mathbf{x}_{in,f}(0) = \mathbf{0}_{p_f}, \quad (3c)$$

$$\dot{\lambda}_f(t) = -\frac{u_{out,f}(t)}{m_f(t)} \lambda_f(t), \quad \lambda_f(0) = 1 \quad (3d)$$

Eqs. (3a)–(3c) define so-called vessel extents, which represent the material that is still in the phase  $F$  at time  $t$ . The negative term on the right-hand side accounts for the effect of the outlet flow.

The vectors of numbers of moles can be reconstructed from the extents as follows:

$$\mathbf{n}_l(t) = \mathbf{N}^T \mathbf{x}_r(t) + \mathbf{W}_{m,l} \mathbf{x}_{m,l}(t) + \mathbf{W}_{in,l} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0} \lambda_l(t) \quad (4a)$$

$$\mathbf{n}_g(t) = -\mathbf{W}_{m,g} \mathbf{x}_{m,g}(t) + \mathbf{W}_{in,g} \mathbf{x}_{in,g}(t) + \mathbf{n}_{g0} \lambda_g(t) \quad (4b)$$

### 2.2.2. Transformation from RMV-form

When the rank conditions associated with Eq. (2) are not satisfied, the linear transformation cannot be used. However, if the inlet and outlet flow rates  $\mathbf{u}_{in,l}(t)$  and  $u_{out,l}(t)$ , and the mass  $m_l(t)$  are known in the liquid phase, the  $p_l$  extents of inlet flow  $\mathbf{x}_{in,l}(t)$  and the discounting variable  $\lambda_l(t)$  can be calculated by integration of the differential equations (3c) and (3d). This allows rearranging Eq. (4a) in a reaction- and mass-transfer-variant (RMV) form:

$$\mathbf{n}_l^{RMV}(t) := \mathbf{n}_l(t) - \mathbf{W}_{in,l} \mathbf{x}_{in,l}(t) - \mathbf{n}_{l0} \lambda_l(t) = [\mathbf{N}^T \quad \mathbf{W}_{m,l}] \begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{m,l}(t) \end{bmatrix} \quad (5)$$

The extents of reaction and mass transfer in the liquid phase are computed by taking the Moore–Penrose pseudo-inverse of the matrix  $[\mathbf{N}^T \quad \mathbf{W}_{m,l}]$ , under the condition that this matrix is of rank  $R + p_m$ .

### 2.2.3. Transformations from MV- and RV-forms

Similarly, for the gas phase, if the quantities  $\mathbf{u}_{in,g}(t)$ ,  $u_{out,g}(t)$  and  $m_g(t)$  are known,  $\mathbf{x}_{in,g}(t)$  and  $\lambda_g(t)$  can be calculated by integration of Eqs. (3c) and (3d). Eq. (4b) can then be rearranged in a mass-transfer-variant (MV) form:

$$\mathbf{n}_g^{MV}(t) := \mathbf{n}_g(t) - \mathbf{W}_{in,g} \mathbf{x}_{in,g}(t) - \mathbf{n}_{g0} \lambda_g(t) = -\mathbf{W}_{m,g} \mathbf{x}_{m,g}(t) \quad (6)$$

<sup>2</sup> Note that  $G$  and  $L$  can also refer to two distinct liquid phases.

and  $\mathbf{x}_{m,g}(t)$  can be calculated using the pseudo-inverse of  $\mathbf{W}_{m,g}$ , provided this matrix is of full rank  $p_m$ .

Once the extents of mass transfer have been computed from the gas phase, they can be converted to extents of mass transfer in the liquid phase using the relationship

$$\mathbf{x}_{m,l}(t) = \mathbf{x}_{m,g}(t) - \delta_m(t) \quad (7)$$

with the auxiliary variables  $\delta_m(t)$  obtained by integration of the following differential equations:

$$\dot{\delta}_m(t) = -\frac{u_{out,l}(t)}{m_l(t)}\delta_m(t) + \left(\frac{u_{out,l}(t)}{m_l(t)} - \frac{u_{out,g}(t)}{m_g(t)}\right)\mathbf{x}_{m,g}(t), \quad \delta_m(0) = \mathbf{0}_{p_m} \quad (8)$$

When the extents of mass transfer  $\mathbf{x}_{m,l}(t)$  have been inferred, Eq. (4a) can be rearranged in a reaction-variant (RV) form, and the extents of reaction can be calculated by means of the pseudo-inverse of  $\mathbf{N}^T$ , under the condition that  $\text{rank}(\mathbf{N}) = R$ :

$$\mathbf{n}_l^{RV}(t) := \mathbf{n}_l(t) - \mathbf{W}_{m,l}\mathbf{x}_{m,l}(t) - \mathbf{W}_{in,l}\mathbf{x}_{in,l}(t) - \mathbf{n}_{l0}\lambda_l(t) = \mathbf{N}^T\mathbf{x}_r(t) \quad (9)$$

#### Special case

For homogeneous reaction systems, the mass-transfer terms vanish and Eqs. (3b) and (4b) drop out. Eqs. (2)–(9) are written without the subscript  $(\cdot)_l$  and are adapted as follows: the matrix  $\mathbf{M}_{m,f0}^T$  is omitted in Eq. (2) and the rank condition for the linear transformation becomes  $\text{rank}([\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0]) = R + p + 1$ ; the mass-transfer term  $\mathbf{W}_{m,l}\mathbf{x}_{m,l}(t)$  is removed from Eq. (4a), which allows writing Eq. (9) as  $\mathbf{n}^{RV}(t) := \mathbf{n}(t) - \mathbf{W}_{in}\mathbf{x}_{in}(t) - \mathbf{n}_0\lambda(t) = \mathbf{N}^T\mathbf{x}_r(t)$ .

### 2.3. Kinetic identification based on computed extents

In the second step, the rate laws and the corresponding rate parameters are estimated using the extents of reaction and mass transfer computed in the first step. For each reaction and mass transfer, a least-squares regression problem is formulated, which involves the extent values computed in the previous step and a postulated rate law with unknown parameters to model the extent as a function of time. The rate parameters are adjusted so as to minimize the difference between the computed and modeled extents.

For the  $i$ th extent of reaction, the modeled extents of reaction are obtained from Eq. (3a) with a rate law of the form  $r_i(\mathbf{c}_i(t), \boldsymbol{\theta}_{r,i})$ , where  $\mathbf{c}_i(t) = \mathbf{n}_i(t)/V_i(t)$  and the vector  $\boldsymbol{\theta}_{r,i}$  represents a set of adjustable parameters such as rate constants and activation energies. If the measurements  $\mathbf{c}_g(t) = \mathbf{n}_g(t)/V_g(t)$  are available in the gas phase, the modeled extents of the  $j$ th mass transfer can be obtained from Eq. (3b) with a postulated rate law of the form  $\zeta_j(\mathbf{c}_i(t), \mathbf{c}_g(t), \boldsymbol{\theta}_{m,j})$ , which depends on the adjustable parameters  $\boldsymbol{\theta}_{m,j}$  that include mass-transfer coefficients.

Because of the incremental nature of the extent-based kinetic identification, each reaction and each mass transfer can be dealt with individually, that is, independently of the other reactions and mass transfers. Hence, an imperfect model (or even the absence of a model) for some of the extents has no impact on the modeling of the other extents. In particular, the extents of reaction can be modeled without having to model the extents of mass transfer.

#### Special case

Homogeneous reaction systems do not involve mass transfer. For these reaction systems, the interest of the incremental method lies in the fact that the effect of each reaction can be decoupled from the other reactions.

### 3. Extent-based prediction of concentrations

The approach of the previous section is now extended to the case where concentration measurements are not directly available, but

instead they are predicted from measured absorbance spectra using multivariate calibration.

#### 3.1. Factorization of spectroscopic data from reaction systems

The starting point of this quantitative analysis of spectroscopic data is Beer's law. Let assume that an absorbance signal is measured in the phase  $L$  by a spectrometer, which produces data that have a linear response with respect to concentrations. Let  $\mathbf{a}(t)$  be a  $W$ -dimensional absorbance spectrum measured for a unit pathlength at  $W$  wavelengths at time  $t$ . Beer's law allows writing:

$$\mathbf{a}(t) = \mathbf{E}^T \mathbf{c}_{l,a}(t) \quad (10)$$

where  $\mathbf{c}_{l,a}(t)$  is the  $S_a$ -dimensional vector of concentrations and  $\mathbf{E}$  is the  $(S_a \times W)$  matrix containing the pure component spectra (molar absorptivities) of the  $S_a$  absorbing species ( $S_a \leq S_l$ ). The subscript  $(\cdot)_a$  indicates a quantity associated with the absorbing species.

##### 3.1.1. Standard factorization

The reconstruction of the numbers of moles  $\mathbf{n}_l(t)$  from the extents  $\mathbf{x}_r(t)$ ,  $\mathbf{x}_{m,l}(t)$ ,  $\mathbf{x}_{in,l}(t)$  and the variable  $\lambda_l(t)$  given in Eq. (4a) can be used to compute the concentrations  $\mathbf{c}_{l,a}(t)$  of the  $S_a$  absorbing species at time  $t$ :

$$\mathbf{c}_{l,a}(t) = \frac{1}{V_l(t)} \left( \mathbf{N}_a^T \mathbf{x}_r(t) + \mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t) + \mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,a} \lambda_l(t) \right) \quad (11)$$

Combining Eqs. (10) and (11), together with  $\mathbf{n}_{l0,a} = \mathbf{c}_{l0,a} V_{l0}$ , leads to the so-called *factorization of spectroscopic data* [19,25] given by:

$$\mathbf{a}(t) = \frac{1}{V_l(t)} \left( \mathbf{E}^T \mathbf{N}_a^T \mathbf{x}_r(t) + \mathbf{A}_m^T \mathbf{x}_{m,l}(t) + \mathbf{A}_{in}^T \mathbf{x}_{in,l}(t) + \mathbf{a}_0 V_{l0} \lambda_l(t) \right) \quad (12)$$

with  $\mathbf{A}_m^T = \mathbf{E}^T \mathbf{W}_{m,l,a}$  the  $(W \times p_m)$  spectra of the species transferring between the two phases (mass-transfer spectra),  $\mathbf{A}_{in}^T = \mathbf{E}^T \mathbf{W}_{in,l,a}$  the  $(W \times p_l)$  spectra of the inlet flows (inlet spectra) and  $\mathbf{a}_0 = \mathbf{E}^T \mathbf{c}_{l0,a}$  the  $W$ -dimensional initial spectrum.

##### 3.1.2. Factorization in RMV-form

Similarly to Eq. (5), if the inlet and outlet flow rates  $\mathbf{u}_{in,l}(t)$  and  $u_{out,l}(t)$  and the mass  $m_l(t)$  of the reaction mixture are known, the extents  $\mathbf{x}_{in,l}(t)$  and the variable  $\lambda_l(t)$  can be calculated by integration of the differential equations (3c) and (3d). It follows that, if  $V_l(t)$ ,  $\mathbf{W}_{in,l,a}$  and  $\mathbf{n}_{l0,a}$  are also known, Eq. (11) can be written in RMV-form as:

$$\begin{aligned} \mathbf{c}_{l,a}^{RMV}(t) &:= \mathbf{c}_{l,a}(t) - \frac{1}{V_l(t)} \left( \mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,a} \lambda_l(t) \right) \\ &= \frac{1}{V_l(t)} \left[ \mathbf{N}_a^T \mathbf{W}_{m,l,a} \right] \begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{m,l}(t) \end{bmatrix} \end{aligned} \quad (13)$$

In addition, if the initial spectrum  $\mathbf{a}_0$  and the inlet spectra  $\mathbf{A}_{in}$  are known, Eq. (12) can be rearranged in RMV-form that accounts only for the spectroscopic contributions of the reactions and mass transfers:

$$\begin{aligned} \mathbf{a}^{RMV}(t) &:= \mathbf{a}(t) - \frac{1}{V_l(t)} \left( \mathbf{A}_{in}^T \mathbf{x}_{in,l}(t) + \mathbf{a}_0 V_{l0} \lambda_l(t) \right) \\ &= \frac{1}{V_l(t)} \left[ \mathbf{E}^T \mathbf{N}_a^T \mathbf{A}_m^T \right] \begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{m,l}(t) \end{bmatrix} \end{aligned} \quad (14)$$

### 3.1.3. Factorization in RV-form

When the  $p_m$  extents of mass transfer are calculated from the numbers of moles measured in the gas phase, the concentration and spectroscopic contributions of the mass transfers can be subtracted from Eqs. (13) and (14), which leads to the following equations:

$$\begin{aligned} \mathbf{c}_{l,a}^{RV}(t) &:= \mathbf{c}_{l,a}(t) - \frac{1}{V_l(t)} (\mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,a} \lambda_l(t) + \mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t)) \\ &= \frac{1}{V_l(t)} \mathbf{N}_a^T \mathbf{x}_r(t) \end{aligned} \quad (15)$$

$$\begin{aligned} \mathbf{a}^{RV}(t) &:= \mathbf{a}(t) - \frac{1}{V_l(t)} (\mathbf{A}_{in}^T \mathbf{x}_{in,l}(t) + \mathbf{a}_0 V_{l0} \lambda_l(t) + \mathbf{A}_m^T \mathbf{x}_{m,l}(t)) \\ &= \frac{1}{V_l(t)} \mathbf{E}^T \mathbf{N}_a^T \mathbf{x}_r(t) \end{aligned} \quad (16)$$

Eqs. (15) and (16) represent the RV-forms of Eqs. (11) and (12), respectively.

#### Special case

For homogeneous reaction systems, Eq. (15) simplifies since  $\mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t) = \mathbf{0}_{S_a}$ . It follows that  $\mathbf{a}^{RV}(t)$  is identical to the expression that holds when all transferring species are spectroscopically silent (non-absorbing), namely, when  $\mathbf{A}_m^T = \mathbf{0}_{W \times p_m}$ .

## 3.2. Prediction of calibrated concentrations from spectroscopy

Concentrations in reaction systems can be predicted from measured absorbance spectra using multivariate calibration. The standard approach for building a calibration model consists in using measured concentrations and absorbance spectra as calibration pairs. Other choices are possible and result in different calibration models.

### 3.2.1. Standard calibration

The calibration proceeds in two steps. In the calibration step, a calibration model is constructed using  $\varepsilon$  latent variables. The calibration set consists of  $H$  pairs of vectors that include  $W$ -dimensional absorbance spectra  $\mathbf{a}_{cal}(t)$  and  $S_c$ -dimensional concentrations  $\mathbf{c}_{l,cal}(t)$ , with  $t$  one of the  $H$  sampling times and  $S_c$  the number of species available for calibration ( $S_c \leq S_l$ ). The subscript  $(\cdot)_{cal}$  denotes a quantity associated with the calibration experiment, while the subscript  $(\cdot)_c$  indicates a quantity associated with the  $S_c$  calibrated species. The  $H$  pairs of vectors  $\mathbf{a}_{cal}(t)$  and  $\mathbf{c}_{l,cal}(t)$  are transposed and stacked vertically in time-resolved matrices  $\mathbf{A}_{cal}(H \times W)$  and  $\mathbf{C}_{l,cal}(H \times S_c)$ , respectively. A calibration model can be expressed by the calibration matrix  $\mathbf{F}$  of dimensions  $(S_c \times W)$  comprising  $S_c$  rows of prognostic vectors of dimension  $W$  that describe the relation between the pairs  $(\mathbf{c}_{l,cal}(t), \mathbf{a}_{cal}(t))$ . This is illustrated in the central part of Fig. 1.

The method for constructing the matrix  $\mathbf{F}$  depends on the calibration technique used. With PCR, the matrix  $\mathbf{A}_{cal}$  is first approximated as  $\bar{\mathbf{U}} \bar{\mathbf{S}} \bar{\mathbf{V}}^T$ , where  $\bar{\mathbf{U}} (H \times \varepsilon)$  and  $\bar{\mathbf{V}} (W \times \varepsilon)$  contain  $\varepsilon$  principal component vectors,  $\bar{\mathbf{S}} (\varepsilon \times \varepsilon)$  is a  $\varepsilon$ -dimensional diagonal matrix of principal component factors, and  $\varepsilon$  is the number of principal components (also called latent variables). The calibration matrix  $\mathbf{F}$  is calculated as  $\mathbf{C}_{l,cal}^T \bar{\mathbf{U}} \bar{\mathbf{S}}^{-1} \bar{\mathbf{V}}^T$ . With PLS, the calculation of the calibration matrix  $\mathbf{F}$  is more complex and calls for an iterative scheme [18].

In the prediction step, the calibration model is used to predict the  $S_c$  calibrated concentrations  $\hat{\mathbf{c}}_{l,c}(t)$  from the  $W$ -dimensional absorbance spectrum  $\mathbf{a}(t)$  measured at time  $t$  of a prediction (or test) experiment:

$$\hat{\mathbf{c}}_{l,c}(t) = \mathbf{F} \mathbf{a}(t) \quad (17)$$

### 3.2.2. Calibration in RMV-form

When calibration models are constructed using only the spectral contributions of the reactions and mass transfers, the pairs of calibration vectors  $(\mathbf{c}_{l,cal}^{RMV}(t), \mathbf{a}_{cal}^{RMV}(t))$  are given by Eqs. (13) and (14). Note that the pretreatment of Eqs. (13) and (14) requires the knowledge of the inlets  $(\mathbf{u}_{in,l,cal}(t), \mathbf{W}_{in,l,cal}, \mathbf{A}_{in,cal})$ , the outlet  $(u_{out,l,cal}(t))$  and the initial conditions  $(\mathbf{n}_{l0,cal}, \mathbf{a}_{0,cal})$ , as well as the mass and volume  $(m_{l,cal}(t), V_{l,cal}(t))$  of the calibration mixture as illustrated in the upper part of Fig. 1. The construction of the calibration matrix  $\mathbf{F}^{RMV}$  is similar to the standard calibration, with the matrices  $\mathbf{A}_{cal}$  and  $\mathbf{C}_{l,cal}$  replaced by  $\mathbf{A}_{cal}^{RMV} (H \times W)$  and  $\mathbf{C}_{l,cal}^{RMV} (H \times S_c)$ , respectively.

To predict the concentrations of the calibrated species, the spectroscopic contributions of the reactions and mass transfers are computed using Eq. (14) from the absorbance spectrum  $\mathbf{a}(t)$  measured for a prediction experiment. The resulting vector  $\mathbf{a}^{RMV}(t)$  is used by the calibration model to predict the  $S_c$  concentrations in RMV-form,  $\hat{\mathbf{c}}_{l,c}^{RMV}(t)$ , for all time instants of the prediction experiment.

### 3.2.3. Calibration in RV-form

If gas-phase measurements can be used to compute the extents of mass transfer in the liquid phase according to Eqs. (6)–(8), the calibration model can be built on the reaction contribution only by removing the contribution due to mass transfer. The pairs of calibration vectors  $(\mathbf{c}_{l,cal}^{RV}(t), \mathbf{a}_{cal}^{RV}(t))$  are obtained by pretreatment using Eqs. (15) and (16), which requires the knowledge of the experimental conditions at all calibration times (lower part of Fig. 1). The calibration matrix  $\mathbf{F}^{RV}$  is constructed in a similar way as for standard calibration, with the matrices  $\mathbf{A}_{cal}$  and  $\mathbf{C}_{l,cal}$  replaced by  $\mathbf{A}_{cal}^{RV} (H \times W)$  and  $\mathbf{C}_{l,cal}^{RV} (H \times S_c)$ , respectively.

In the prediction step, the absorbance spectrum  $\mathbf{a}(t)$  measured during a prediction experiment is transformed to the RV-form  $\mathbf{a}^{RV}(t)$  using Eq. (16). This spectroscopic contribution is used to predict the  $S_c$  concentrations in RV-form,  $\hat{\mathbf{c}}_{l,c}^{RV}(t)$ . Note that this procedure is also used when there are no transferring species in the calibration set.

#### Special case

For homogeneous reaction systems, standard calibration is based on the pairs  $(\mathbf{c}_{cal}(t), \mathbf{a}_{cal}(t))$ . For calibration in RV-form,  $\mathbf{c}_{cal}^{RV}(t)$  is obtained using Eq. (15) without the term  $\mathbf{W}_{m,l,a} \mathbf{x}_{m,l}(t)$ , while  $\mathbf{a}_{cal}^{RV}(t)$  results from Eq. (16) without the term  $\mathbf{A}_m^T \mathbf{x}_{m,l}(t)$ . In the prediction step, the pretreated vector  $\mathbf{a}^{RV}(t)$  is then used to predict the concentrations  $\hat{\mathbf{c}}_c^{RV}(t)$  associated with the reactions.

## 3.3. Extent-based prediction of all concentrations

The calibration models  $\mathbf{F}$ ,  $\mathbf{F}^{RMV}$  and  $\mathbf{F}^{RV}$  can predict the concentrations of the calibrated species only. However, for kinetic identification, one may need the concentrations of all (calibrated as well as uncalibrated) species, which can be achieved in different ways as illustrated in Fig. 2. The key idea is to predict the various extents and reconstruct all the concentrations using additional information involving the experimental conditions.

### 3.3.1. Prediction over $\hat{\mathbf{c}}_{l,c}(t)$

The most straightforward way to predict  $\hat{\mathbf{c}}_l(t)$  from  $\mathbf{a}(t)$  is to use the calibration model  $\mathbf{F}$  to predict  $\hat{\mathbf{c}}_{l,c}(t)$ . Then, with the knowledge of the volume  $V_l(t)$ ,  $S_c$  numbers of moles can be computed as  $\hat{\mathbf{n}}_{l,c}(t) = \hat{\mathbf{c}}_{l,c}(t) V_l(t)$  and transformed to extents according to Eq. (2) written for the  $S_c$  species. This requires the matrices of stoichiometry  $\mathbf{N} (R \times S_l)$ , mass transfer  $\mathbf{W}_{m,l} (S_l \times p_m)$ , inlet composition  $\mathbf{W}_{in,l} (S_l \times p_l)$  and the  $S_l$ -dimensional vector of initial conditions  $\mathbf{n}_{l0}$  to be reduced to  $\mathbf{N}_c (R \times S_c)$ ,  $\mathbf{W}_{m,l,c} (S_c \times p_m)$ ,  $\mathbf{W}_{in,l,c} (S_c \times p_l)$  and  $\mathbf{n}_{l0,c}$  ( $S_c$ -dimensional vector).

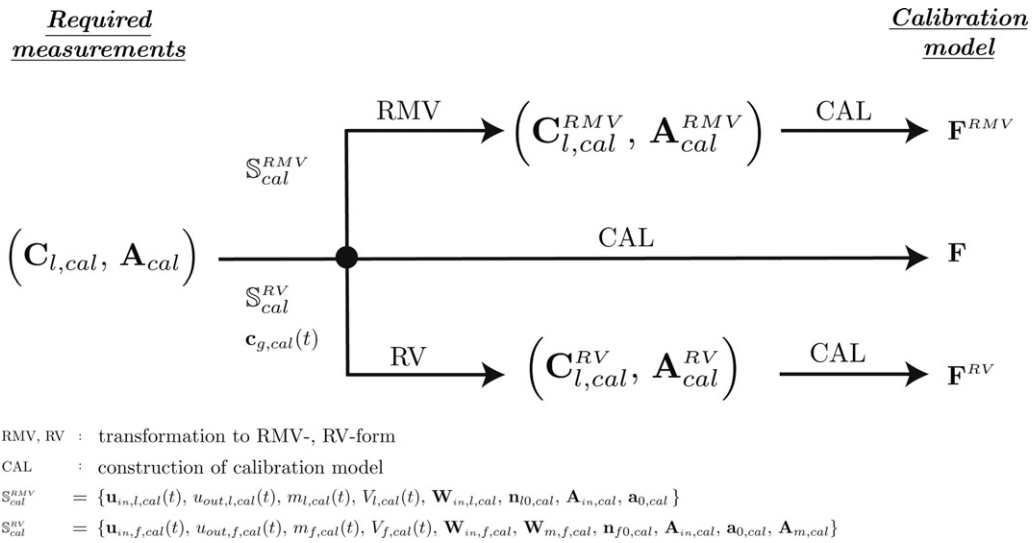


Fig. 1. Construction of the calibration model for standard calibration (middle), calibration in RMV-form (top) and calibration in RV-form (bottom).

The linear transformation from the measured spectrum to the predicted extents, which encompasses Paths 1 and 4 in Fig. 2, is:

$$\begin{bmatrix} \hat{x}_r(t) \\ \hat{x}_{m,l}(t) \\ \hat{x}_{in,l}(t) \\ \hat{\lambda}_l(t) \end{bmatrix} = \begin{bmatrix} S_{l0,c}^T \\ M_{m,l0,c}^T \\ M_{in,l0,c}^T \\ Q_{l0,c}^T \end{bmatrix} F a(t) V_l(t) \quad (18)$$

The procedure for constructing the matrices  $S_{l0,c}$  ( $S_c \times R$ ),  $M_{m,l0,c}$  ( $S_c \times p_m$ ),  $M_{in,l0,c}$  ( $S_c \times p_l$ ), and the  $S_c$ -dimensional vector  $Q_{l0,c}$  remains identical to the method proposed in [23], except that the total number of species  $S_l$  is replaced by the number of calibrated

species  $S_c$ . Consequently, the condition for using the linear transformation of Eq. (18) becomes  $\text{rank}([N_c^T W_{m,l,c} W_{in,l,c} n_{l0,c}]) = R + p_m + p_l + 1$ , which also requires  $S_c \geq R + p_m + p_l + 1$ . This condition can be more or less restrictive depending on the choice of the set of calibrated species  $S_c$ .

The  $S_u = S_l - S_c$  uncalibrated species can be reconstructed from the various extents as follows:

$$\hat{c}_{l,u}(t) = \frac{1}{V_l(t)} (N_u^T \hat{x}_r(t) + W_{m,l,u} \hat{x}_{m,l}(t) + W_{in,l,u} \hat{x}_{in,l}(t) + n_{l0,u} \hat{\lambda}_l(t)) \quad (19)$$

where the subscript  $(\cdot)_u$  indicates a quantity associated with the uncalibrated species.

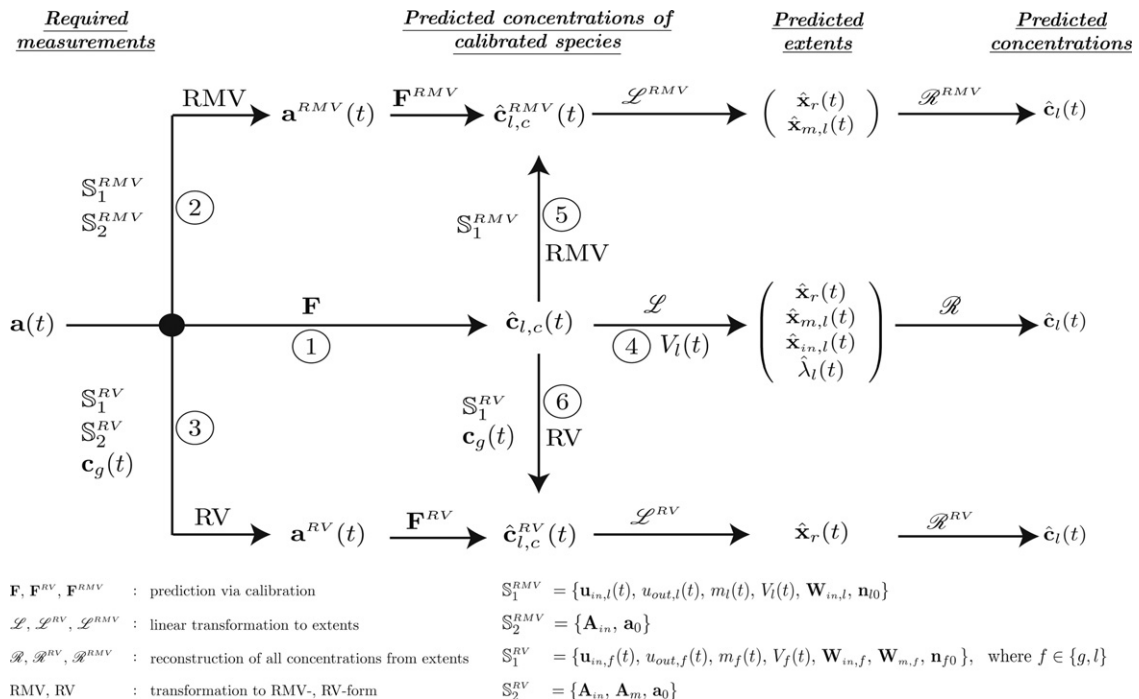


Fig. 2. Prediction of all concentrations from a measured absorbance spectrum.

### 3.3.2. Prediction over $\hat{\mathbf{c}}_{i,c}^{RMV}(t)$

If the condition,  $\text{rank}([\mathbf{N}_c^T \mathbf{W}_{m,l,c} \mathbf{W}_{in,l,c} \mathbf{n}_{l0,c}]) = R + p_m + p_l + 1$ , is not satisfied, an alternative consists in rewriting the predicted concentrations in RMV-form using Eq. (13), for which additional experimental conditions need to be known. This corresponds to Path 5 in Fig. 2. Note that  $\hat{\mathbf{c}}_{i,c}^{RMV}(t)$  can also be obtained using the calibration model in RMV-form (Path 2).

Under the condition,  $\text{rank}([\mathbf{N}_c^T \mathbf{W}_{m,l,c}]) = R + p_m$ , which requires  $S_c \geq R + p_m$ , the extents of reaction and mass transfer can be calculated as:

$$\begin{bmatrix} \hat{\mathbf{x}}_r(t) \\ \hat{\mathbf{x}}_{m,l}(t) \end{bmatrix} = [\mathbf{N}_c^T \mathbf{W}_{m,l,c}]^+ \hat{\mathbf{c}}_{i,c}^{RMV}(t) V_l(t) \quad (20)$$

The calibrated and uncalibrated concentrations are reconstructed from the predicted extents  $\hat{\mathbf{x}}_r(t)$  and  $\hat{\mathbf{x}}_{m,l}(t)$  and the variables  $\mathbf{x}_{in,l}(t)$  and  $\lambda_l(t)$  that are computed by integration of Eqs. (3c) and (3d) using known experimental conditions:

$$\hat{\mathbf{c}}_{i,c}(t) = \hat{\mathbf{c}}_{i,c}^{RMV}(t) + \frac{1}{V_l(t)} (\mathbf{W}_{in,l,c} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,c} \lambda_l(t)) \quad (21)$$

$$\hat{\mathbf{c}}_{i,u}(t) = \frac{1}{V_l(t)} (\mathbf{N}_u^T \hat{\mathbf{x}}_r(t) + \mathbf{W}_{m,l,u} \hat{\mathbf{x}}_{m,l}(t) + \mathbf{W}_{in,l,u} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,u} \lambda_l(t)) \quad (22)$$

### 3.3.3. Prediction over $\hat{\mathbf{c}}_{i,c}^{RV}(t)$

The predicted concentrations  $\hat{\mathbf{c}}_{i,c}(t)$  can also be transformed to RV-form using Eq. (15), for which additional gas-phase measurements and experimental conditions need to be known (Path 6). Note that  $\hat{\mathbf{c}}_{i,c}^{RV}(t)$  can also be obtained using the calibration model in RV-form (Path 3).

If  $\text{rank}(\mathbf{N}_c) = R$ , which requires  $S_c \geq R$ , the extents of reaction can then be computed as:

$$\hat{\mathbf{x}}_r(t) = \mathbf{N}_c^{T+} \hat{\mathbf{c}}_{i,c}^{RV}(t) V_l(t) \quad (23)$$

The calibrated and uncalibrated concentrations are reconstructed from the predicted extents of reaction  $\hat{\mathbf{x}}_r(t)$ , the variables  $\mathbf{x}_{in,l}(t)$  and  $\lambda_l(t)$  that are computed by integration of Eqs. (3c) and (3d) using known experimental conditions and the extents of mass transfer  $\mathbf{x}_{m,l}(t)$  that are obtained from gas-phase measurements and Eqs. (6)–(8):

$$\hat{\mathbf{c}}_{i,c}(t) = \hat{\mathbf{c}}_{i,c}^{RV}(t) + \frac{1}{V_l(t)} (\mathbf{W}_{m,l,c} \mathbf{x}_{m,l}(t) + \mathbf{W}_{in,l,c} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,c} \lambda_l(t)) \quad (24)$$

$$\hat{\mathbf{c}}_{i,u}(t) = \frac{1}{V_l(t)} (\mathbf{N}_u^T \hat{\mathbf{x}}_r(t) + \mathbf{W}_{m,l,u} \mathbf{x}_{m,l}(t) + \mathbf{W}_{in,l,u} \mathbf{x}_{in,l}(t) + \mathbf{n}_{l0,u} \lambda_l(t)) \quad (25)$$

### 3.3.4. Choice of calibration model

Which path to choose in Fig. 2 depends on whether or not the spectrum  $\mathbf{a}(t)$  of the prediction experiment lies in the subspace spanned by the calibration set. This space-inclusion condition has been studied in [19,25]. For Path 1, the space-inclusion condition  $\mathbf{a}(t) \in \text{Im}(\mathbf{A}_{cal})$  can be verified by checking that the Euclidean norm  $\epsilon(\mathbf{a}(t), \mathbf{A}_{cal})$  of the projection error of  $\mathbf{a}(t)$  on the row space of  $\mathbf{A}_{cal}$  is negligible. Such a norm can be computed as  $\epsilon(\mathbf{a}(t), \mathbf{A}_{cal}) = \|\mathbf{a}^T(t)(\mathbf{I}_W - \mathbf{A}_{cal}^+ \mathbf{A}_{cal})\|_2$ . If this condition is not satisfied, Path 2 can be followed, which implies that the calibration and linear transformation steps are performed in RMV-form (see Fig. 2). For Path 2, the space-inclusion condition becomes  $\mathbf{a}^{RMV}(t) \in \text{Im}(\mathbf{A}_{cal}^{RMV})$  and can be verified by calculating  $\epsilon(\mathbf{a}^{RMV}(t), \mathbf{A}_{cal}^{RMV})$ . If the space-inclusion condition is still not fulfilled, additional independent measurements,

such as gas-phase measurements, are required to take Path 3 and use a calibration model in RV-form. For such a calibration, the space-inclusion condition is  $\mathbf{a}^{RV}(t) \in \text{Im}(\mathbf{A}_{cal}^{RV})$  and can be checked by computing  $\epsilon(\mathbf{a}^{RV}(t), \mathbf{A}_{cal}^{RV})$ . If none of these space-inclusion conditions can be satisfied, the calibration set must be re-designed and it might be necessary to choose another set of  $S_c$  calibrated species.

### 3.3.5. Choice of linear transformation

There are three possible paths to predict the concentrations of the calibrated species, namely to  $\hat{\mathbf{c}}_{i,c}(t)$ ,  $\hat{\mathbf{c}}_{i,c}^{RMV}(t)$  and  $\hat{\mathbf{c}}_{i,c}^{RV}(t)$ . Which one to choose depends on the conditions for applying the linear transformation of Eqs. (18), (20) and (23). If  $\text{rank}([\mathbf{N}_c^T \mathbf{W}_{m,l,c} \mathbf{W}_{in,l,c} \mathbf{n}_{l0,c}]) = R + p_m + p_l + 1$ , Path 4 can be taken. This allows predicting the  $R$  extents of reaction, the  $p_m$  extents of mass transfer, the  $p_l$  extents of inlet flow and the discounting variable  $\lambda_l$  via Eq. (18). If this condition is not fulfilled, a treatment of  $\hat{\mathbf{c}}_{i,c}(t)$  to RMV- or RV-form is required. If  $\text{rank}([\mathbf{N}_c^T \mathbf{W}_{m,l,c}]) = R + p_m$ , the predicted concentrations  $\hat{\mathbf{c}}_{i,c}(t)$  can be rearranged to RMV-form along Path 5 and used to predict the  $R$  extents of reaction and the  $p_m$  extents of mass transfer via Eq. (20). Otherwise, additional independent measurements, such as gas-phase measurements, are necessary to follow Path 6 and construct  $\hat{\mathbf{c}}_{i,c}^{RV}(t)$  so that the  $R$  extents of reaction can be predicted by Eq. (23), under the condition that  $\text{rank}(\mathbf{N}_c) = R$ . If none of these rank conditions are met, another set of calibrated species has to be selected and the calibration set re-designed.

#### Special case

For homogeneous reaction systems, all the terms related to mass transfer are removed from Eqs. (18)–(25). In particular, the rank condition for the linear transformation of Path 4 reduces to  $\text{rank}([\mathbf{N}_c^T \mathbf{W}_{in,c} \mathbf{n}_{l0,c}]) = R + p + 1$ . When this latter condition is not satisfied, the extents of reaction can be computed on the basis of the predicted concentrations in RV-form according to Eq. (23), and the concentrations of all species can be reconstructed using Eqs. (24) and (25), without the mass-transfer terms  $\mathbf{W}_{m,l,c} \mathbf{x}_{m,l}(t)$  and  $\mathbf{W}_{m,l,u} \mathbf{x}_{m,l}(t)$ .

## 4. Case studies

Two case studies based on simulated data are presented. The first study considers the acetoacetylation of pyrrole in a stirred-tank reactor and illustrates the extent-based kinetic identification of a homogeneous reaction system from spectroscopic data. The second study, which considers the chlorination of butanoic acid in a controlled gas–liquid stirred-tank reactor, describes the use of extent-based kinetic identification from absorbance measurements in the liquid phase and concentration measurements in the gas phase.

### 4.1. Acetoacetylation of pyrrole

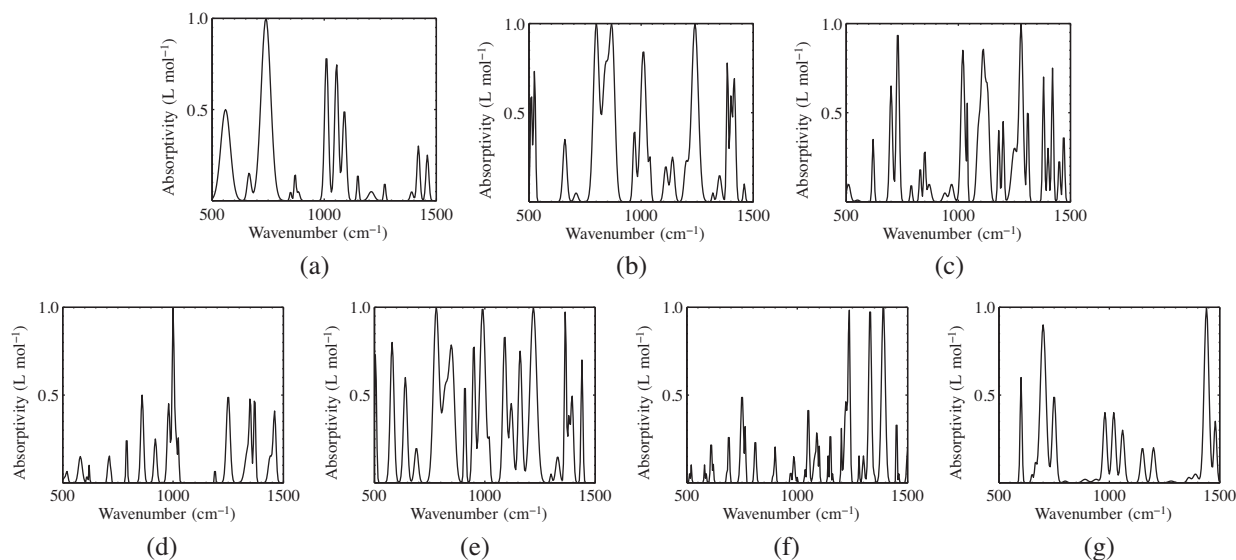
The acetoacetylation of pyrrole is a homogeneous reaction system that comprises  $R=4$  reactions involving  $S=7$  species.<sup>3</sup>

#### 4.1.1. Reaction system

In the first reaction (R1), pyrrole A reacts with diketene B to produce 2-acetoacetyl pyrrole C. Simultaneously, diketene undergoes multiple parallel reactions, including polymerization (R2) into dehydroacetic acid D, oligomerization (R3) forming oligomer E, and further reaction (R4) with acetoacetyl pyrrole leading to by-product F. Reactions R1, R2 and R4 are catalyzed by pyridine G. Toluene is

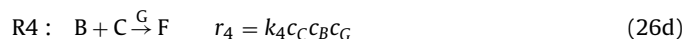
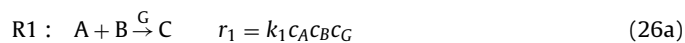
<sup>3</sup> The subscript ( $\cdot$ )<sub>l</sub> is omitted in this case study with a single phase.





**Fig. 3.** Acetoacetylation of pyrrole: Simulated mid-IR (500–1500  $\text{cm}^{-1}$ ) pure component spectra (absorptivities) of (a) pyrrole, (b) diketene, (c) 2-acetoacetyl pyrrole, (d) dehydroacetic acid, (e) oligomer, (f) by-product and (g) pyridine, normalized to a unit peak height.

used as solvent for this reaction system. The reactions are represented by the following stoichiometric and kinetic expressions:



where  $k_1 = 0.0530 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$ ,  $k_2 = 0.1280 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$ ,  $k_3 = 0.0280 \text{ min}^{-1}$  and  $k_4 = 0.0030 \text{ L}^2 \text{ mol}^{-2} \text{ min}^{-1}$  are rate constants reported in [26]. The stoichiometric matrix  $\mathbf{N}$  for the set of all species  $\{A, B, C, D, E, F, G\}$  reads:

$$\mathbf{N} = \begin{bmatrix} -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (27)$$

#### 4.1.2. Simulated measurements

All simulated experiments are conducted in an open isothermal unsteady-state reactor at  $50^\circ\text{C}$ . The reactor is initially loaded with  $V_0 = 1 \text{ L}$  of a solution of pyrrole in toluene and pure diketene is fed ( $p=1$ ) at a constant mass flow rate with the inlet composition  $\mathbf{w}_{in} = [0 \text{ } M_{w,B}^{-1} \check{w}_{in,B} \text{ } 0 \text{ } 0 \text{ } 0 \text{ } 0 \text{ } 0]^T$ , where  $M_{w,B} = 84.08 \text{ g mol}^{-1}$  and  $\check{w}_{in,B} = 1$ . Since the density and the volume of the reaction mixture are assumed to be constant throughout the course of the reaction,  $u_{out}(t) = u_{in}(t)$ .

The concentrations  $\mathbf{c}(t) = \mathbf{n}(t)/V(t)$  are obtained by integration of the differential mole balances in Eq. (1a). Each concentration is corrupted by additive zero-mean normally distributed noise with standard deviation corresponding to 1 % of the maximal concentration. Two other sets of concentration data are also generated using noise levels of 2% and 5%. Absorbance spectra are assumed to be measured in the mid-IR region (500–1500  $\text{cm}^{-1}$ ) with 4  $\text{cm}^{-1}$  resolution, thus leading to  $W=250$  wavenumbers monitored simultaneously. The spectrum of the solvent (toluene) is assumed to be treated as background spectrum. All species involved in the reactions are absorbing, so that  $S_a = S$  and  $\mathbf{c}_a(t) = \mathbf{c}(t)$ . The absorbance spectra  $\mathbf{a}(t)$  are generated using Eq. (10) from the simulated concentrations (corrupted with 1%, 2% and 5% noise) and

the  $(7 \times 250)$ -dimensional pure component spectra  $\mathbf{E}$  chosen to resemble the true ones. The pure component spectra are shown in Fig. 3.

Since the quantities  $u_{in}(t)$ ,  $u_{out}(t)$ ,  $m(t)$  and  $V(t)$  are known, the extent of inlet flow  $x_{in}(t)$  and the discounting variable  $\lambda(t)$  can be computed from Eqs. (3c) and (3d). Under the condition that all experimental conditions (including the inlet spectrum of diketene) are known, the concentration and spectral contributions associated with the inlet and outlet can be removed from the vectors  $\mathbf{c}(t)$  and  $\mathbf{a}(t)$  to generate  $\mathbf{c}^{RV}(t)$  and  $\mathbf{a}^{RV}(t)$  using Eqs. (15) and (16). The noise levels of 1%, 2% and 5% in  $\mathbf{c}(t)$  lead to noise levels of the same standard deviation in  $\mathbf{c}^{RV}(t)$ . The resulting noise levels in  $\mathbf{a}(t)$  and  $\mathbf{a}^{RV}(t)$  are 1.5%, 2.6% and 5.6%, respectively.

#### 4.1.3. Calibration experiment

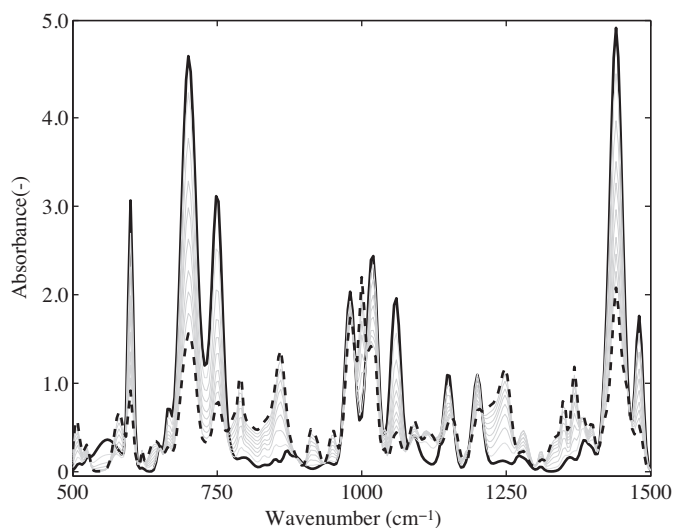
In the calibration experiment, the reactor is initially loaded with a solution of composition  $\mathbf{c}_{0,cal} = [0.92 \text{ } 0.07 \text{ } 0.10 \text{ } 0.02 \text{ } 0.0 \text{ } 4.50]^T \text{ mol L}^{-1}$  of mass 919.98 g. Pure diketene is fed at the constant mass flow rate of  $u_{in} = 3.65 \text{ g min}^{-1}$ , and the reactor content is continuously withdrawn at the same flow rate. The concentrations and absorbance spectra are generated according to the procedure described in Section 4.1.2.

The calibration set is built by taking  $H=15$  absorbance spectra at the equally spaced times  $t=0, 10, \dots, 140 \text{ min}$ , that is, with a sampling time of 10 min. In parallel, it is assumed that samples are taken from the reaction mixture and concentrations of all species except the catalyst are determined by an offline measurement technique. This sampling procedure results in 15 pairs of calibration vectors ( $\mathbf{c}_{cal}(t)$ ,  $\mathbf{a}_{cal}(t)$ ) for the  $S_c = 6$  species  $\{A, B, C, D, E, F\}$ . As all experimental conditions (including the inlet spectrum) are assumed to be known, the vectors  $\mathbf{c}_{cal}^{RV}(t)$  and  $\mathbf{a}_{cal}^{RV}(t)$  are also available.

The  $(6 \times 250)$ -dimensional calibration matrices  $\mathbf{F}$  and  $\mathbf{F}^{RV}$  are constructed by PCR using the pairs of calibration vectors ( $\mathbf{c}_{cal}(t)$ ,  $\mathbf{a}_{cal}(t)$ ) and ( $\mathbf{c}_{cal}^{RV}(t)$ ,  $\mathbf{a}_{cal}^{RV}(t)$ ), respectively. One extra latent variable is used to account for the effect of noise, that is,  $\varepsilon = 7$  latent variables instead of  $S_c = 6$  for  $\mathbf{F}$ , and  $\varepsilon = 5$  latent variables instead of  $R = 4$  for  $\mathbf{F}^{RV}$ .

#### 4.1.4. Prediction experiment

In the prediction experiment, the reactor is initially loaded with a solution of composition  $\mathbf{c}_0 = [0.72 \text{ } 0.09 \text{ } 0.10 \text{ } 0.02 \text{ } 0.0 \text{ } 5.00]^T$



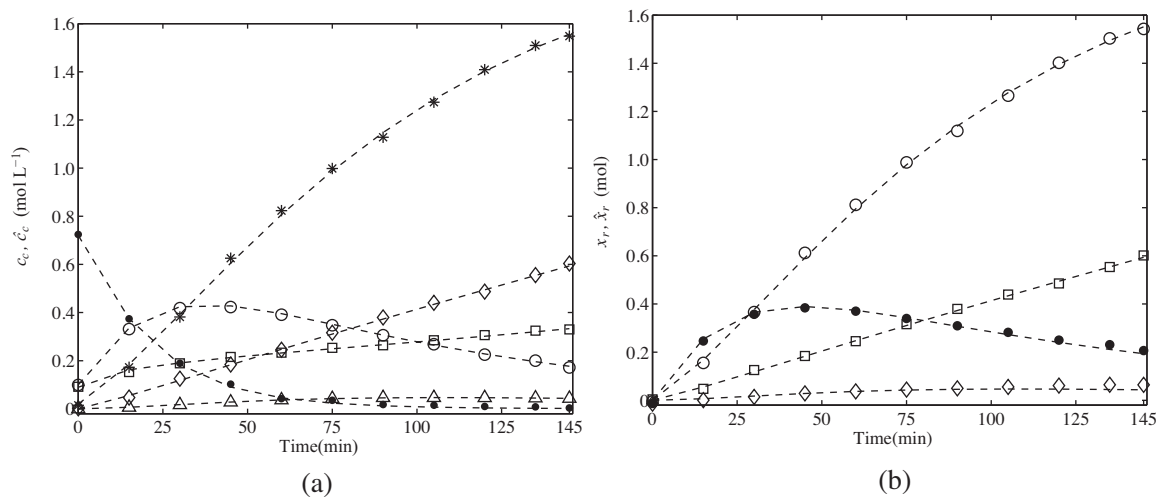
**Fig. 4.** Acetoacetylation of pyrrole: Mid-IR absorbance spectra  $\mathbf{a}(t)$  with 1.5% noise, measured every minute but plotted every 15 min. The black continuous and dashed lines indicate the first and the last spectra, respectively.

mol L<sup>-1</sup> of mass 923.46 g. Pure diketene is fed at the constant mass flow rate of  $u_{in} = 4.40$  g min<sup>-1</sup>, and the reactor content is continuously withdrawn at the same flow rate. The absorbance spectra are generated according to the procedure described in Section 4.1.2.

The prediction set is built by taking absorbance spectra every 1 min for a duration of 145 min. The resulting measured absorbance spectra with 1.5% noise are shown in Fig. 4. As all experimental conditions (including the inlet spectrum) are assumed to be known, the vector  $\mathbf{a}^{RV}(t)$  is also available.

Since  $\text{rank}([\mathbf{N}_c^T \mathbf{w}_{in,c} \mathbf{n}_{0,c}]) = R + p + 1 = 6$ , the rank condition is fulfilled for the linear transformation of Eq. (18) so that Paths 1 and 4 can be followed to the various extents and the concentrations  $\hat{c}_c(t)$  via the calibration model **F**. Fig. 5 shows the excellent agreement between simulated and predicted concentrations and extents of reaction for the case of absorbance spectra corrupted with 1.5% noise.

Two other routes exist in Fig. 2 for predicting the  $R=4$  extents of reaction. By following Paths 1 and 6, the concentrations  $\hat{c}_c(t)$  are predicted via the calibration model **F** and are rearranged in



**Fig. 5.** Prediction experiment for acetoacetylation of pyrrole: (a) Noise-free simulated concentrations (dashed line) and noisy concentrations predicted via Path 1 (PCR model) and Path 4 for species A (●), B (□), C (○), D (\*) and F (△) in the presence of 1.5% measurement noise; (b) noise-free simulated (dashed line) and noisy predicted extents of reaction for R1 (●), R2 (○), R3 (□) and R4 (◇), plotted every 15 min.

**Table 1**

Rate constants for the acetoacetylation of pyrrole: Simulated  $k_i$  and estimated  $\hat{k}_i$  rate constants obtained by fitting the predicted extents of reaction  $\hat{x}_r(t)$  obtained by PCR via both **F** and **F<sup>RV</sup>**. The rate constants for reactions R1, R2 and R4 are given in L<sup>2</sup> mol<sup>-2</sup> min<sup>-1</sup> and in min<sup>-1</sup> for reaction R3. The corresponding 99% confidence intervals are also given.

Reaction	Simulation $k_i$	Estimation via <b>F</b> $\hat{k}_i$ [99% C.I.]	Estimation via <b>F<sup>RV</sup></b> $\hat{k}_i$ [99% C.I.]
R1	0.0530	0.0531 [0.0529, 0.0532]	0.0528 [0.0526, 0.0530]
R2	0.1280	0.1281 [0.1276, 0.1286]	0.1277 [0.1274, 0.1280]
R3	0.0280	0.0280 [0.0279, 0.0281]	0.0280 [0.0278, 0.0281]
R4	0.0030	0.0030 [0.0030, 0.0030]	0.0030 [0.0030, 0.0030]

the RV-form  $\hat{c}_c^{RV}(t)$  using Eq. (15). Alternatively, the concentrations in RV-form  $\hat{c}_c^{RV}(t)$  can be obtained via Path 3 by rearranging the absorbance spectra  $\mathbf{a}(t)$  in the RV-form  $\mathbf{a}^{RV}(t)$  using Eq. (16) and applying the calibration model **F<sup>RV</sup>**. Both these routes involve the subsequent use of Eq. (23) to predict the extents of reaction. For the sake of concision, only Path 3 will be shown in the present case study.

#### 4.1.5. Kinetic identification

For kinetic identification based on the predicted extents of reaction  $\hat{x}_r(t)$ , the concentration  $\hat{c}_i(t)$  of the remaining  $S_{ii} = 1$  uncalibrated species **G** has to be reconstructed using Eq. (19) (Paths 1 and 4) or Eq. (25) (Path 3). This step is particularly important in this study because the rate laws of reactions R1, R2 and R4 involve the catalyst concentration, which is not in the calibration set.

For each of the four reactions, a rate law is postulated and a least-squares problem is formulated to estimate the rate parameters that minimize the difference between the extents predicted by calibration and those modeled by the postulated rate law, as described in Section 2.3. The identification is performed using the extents of reaction predicted via both **F** (Paths 1 and 4) and **F<sup>RV</sup>** (Path 3). As shown in Table 1, individual fitting of the extents of reaction obtained from absorbance spectra with 1.5% noise leads to accurate rate constant estimates, regardless of the calibration procedure (via **F** or **F<sup>RV</sup>**) that is used. Very similar results have been obtained using PLS (not shown in Table 1).

To study the effect of the noise level on the estimated rate constants in Table 1, a sensitivity analysis was conducted. For that purpose, sets of calibration and prediction data corrupted with higher noise levels were used. This analysis shows that (i) the value

**Table 2**

Identification of the first reaction in the acetoacetylation of pyrrole for the case of 1.5% noise: Sum of squared residuals (ssq) corresponding to various postulated rate laws for the least-squares regression of the extent of reaction R1, predicted by PCR via **F** (Paths 1 and 4) and **F<sup>RV</sup>** (Path 3). The rate law  $r_1 = k_{CAcBcG}$  is the one used to generate the measured data.

Rate law	Estimation via <b>F</b> ssq	Estimation via <b>F<sup>RV</sup></b> ssq
$k_{CAcBcG}$	0.0025	0.0048
$k_{CAcB}$	0.0510	0.0481
$k_{CAcG}$	0.0458	0.0502
$k_{CBcG}$	2.7712	2.7913

of the estimated parameters (bias) can change significantly with the level of noise, and (ii) in average, the range of the confidence intervals are multiplied by 2 and 8 when the noise level is raised from 1.5% to 2.6% and 5.6%, respectively, regardless of the calibration procedure used. As a consequence, with higher noise levels, some of the 99% confidence intervals do not include the true (simulated) rate constants.

To demonstrate the robustness of the identification approach, the absorbance spectra corrupted with 1.5% noise were used and different (correct as well as incorrect) rate laws were tested for the first reaction. Table 2 shows that the predicted extents of reaction lead to unambiguous identification of the correct rate laws. This is also the case for the other reactions and when PLS is used (not shown in Table 2).

#### 4.2. Chlorination of butanoic acid

The chlorination of butanoic acid is a gas–liquid reaction system involving  $R=2$  reactions in the liquid phase that both consume chlorine dissolved from the gas phase.

##### 4.2.1. Reaction system

The first reaction between butanoic acid (BA) and chlorine ( $\text{Cl}_2$ ) produces  $\alpha$ -mono-chloro-butanoic acid (MBA) and hydrochloric acid (HCl), while the second reaction leads to the side product  $\alpha$ -di-chloro-butanoic acid (DBA) and HCl:



Due to its high volatility, HCl is found in both phases. Ethanol (EtOH) is used as solvent for the liquid phase. Air is initially present in the gas phase and is considered as insoluble in the liquid phase. Hence, this reaction system has the  $S_l = 6$  species {BA,  $\text{Cl}_2$ , MBA, HCl, DBA, EtOH} in the liquid phase and the  $S_g = 3$  species { $\text{Cl}_2$ , HCl, air} in the gas phase. Both reactions are catalyzed by the product MBA. The rate laws for the reactions R1 and R2 and the mass transfers of  $\text{Cl}_2$  and HCl are<sup>4</sup>:

$$r_1 = k_1 c_{l,BA} c_{l,\text{Cl}_2} \sqrt{c_{l,MBA}} \quad (29a)$$

$$r_2 = k_2 c_{l,BA} c_{l,\text{Cl}_2}^2 \sqrt{c_{l,MBA}} \quad (29b)$$

$$\zeta_{\text{Cl}_2} = k_{\text{Cl}_2} a V_l M_{w,\text{Cl}_2} (c_{l,\text{Cl}_2}^* - c_{l,\text{Cl}_2}) \quad (29c)$$

$$\zeta_{\text{HCl}} = k_{\text{HCl}} a V_l M_{w,\text{HCl}} (c_{l,\text{HCl}}^* - c_{l,\text{HCl}}), \quad (29d)$$

where  $k_1 = 1.358 (\text{m}^3)^{3/2} \text{kmol}^{-3/2} \text{s}^{-1}$  and  $k_2 = 0.136 (\text{m}^3)^{5/2} \text{kmol}^{-5/2} \text{s}^{-1}$  are rate constants,  $k_{\text{Cl}_2} = 0.666 \times 10^{-4} \text{m s}^{-1}$  and  $k_{\text{HCl}} = 0.845 \times 10^{-4} \text{m s}^{-1}$  are molar transfer coefficients,  $a = 254.9 \text{m}^{-1}$  is the specific interfacial area, and  $M_{w,\text{Cl}_2} =$

$70.91 \text{kg kmol}^{-1}$  and  $M_{w,\text{HCl}} = 36.46 \text{kg kmol}^{-1}$  are the molecular weights of  $\text{Cl}_2$  and HCl. The equilibrium concentrations of  $\text{Cl}_2$  and HCl at the interface are calculated as  $c_{l,\text{Cl}_2}^* = p_{\text{Cl}_2} / H_{\text{Cl}_2}$  and  $c_{l,\text{HCl}}^* = p_{\text{HCl}} / H_{\text{HCl}}$ , where  $p_{\text{Cl}_2} = c_{g,\text{Cl}_2} \mathcal{R} T$  and  $p_{\text{HCl}} = c_{g,\text{HCl}} \mathcal{R} T$  are the partial pressures expressed in bar, with  $\mathcal{R} = 0.0831 \text{m}^3 \text{bar K}^{-1} \text{kmol}^{-1}$  the ideal gas constant and  $T$  the temperature in Kelvin, and  $H_{\text{Cl}_2} = 70.33 \text{bar m}^3 \text{kmol}^{-1}$  and  $H_{\text{HCl}} = 70.33 \text{bar m}^3 \text{kmol}^{-1}$  are the Henry constants of  $\text{Cl}_2$  and HCl. Note that the rate laws of reactions R1 and R2 correspond to simplified rate expressions of the ones published in [23].

The stoichiometric matrix **N** for the set of all liquid species reads:

$$\mathbf{N} = \begin{bmatrix} -1 & -1 & 1 & 1 & 0 & 0 \\ -1 & -2 & 0 & 1 & 2 & 0 \end{bmatrix} \quad (30)$$

The mass-transfer matrices  $\mathbf{W}_{m,l}$  and  $\mathbf{W}_{m,g}$  correspond to the set { $\text{Cl}_2$ , HCl} of  $S_m = 2$  transferring species and reads:

$$\mathbf{W}_{m,l} = \begin{bmatrix} 0 & 0 \\ M_{w,\text{Cl}_2}^{-1} & 0 \\ 0 & 0 \\ 0 & M_{w,\text{HCl}}^{-1} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{W}_{m,g} = \begin{bmatrix} M_{w,\text{Cl}_2}^{-1} & 0 \\ 0 & M_{w,\text{HCl}}^{-1} \\ 0 & 0 \end{bmatrix} \quad (31)$$

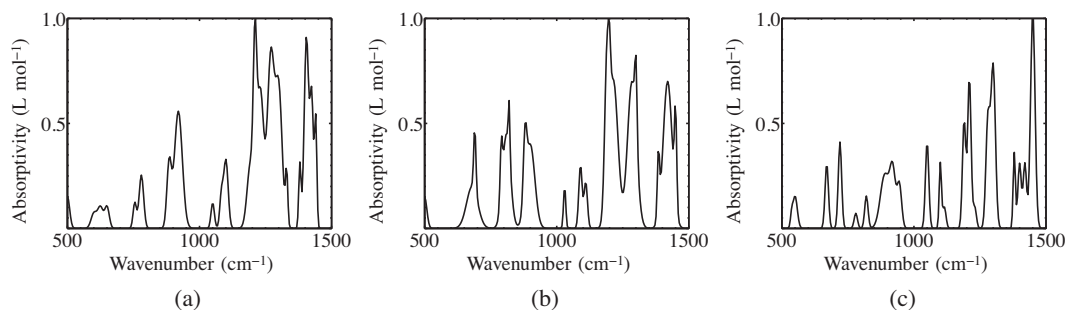
##### 4.2.2. Simulated measurements

All experiments are conducted at  $130^\circ \text{C}$  in an open unsteady-state gas–liquid reactor of total volume  $V_{\text{tot}} = 9 \text{m}^3$ . The reactor, which initially contains air, is loaded with  $m_{l0} = 4600 \text{kg}$  of a liquid solution containing ethanol as solvent. Depending on the experiments, butanoic acid is either present in the initial load or fed ( $p_l = 1$ ) at a constant mass flow rate  $u_{in,l}$  with composition  $\mathbf{w}_{in,l} = [M_{w,BA}^{-1} \check{w}_{in,l,BA} \ 0 \ 0 \ 0 \ 0 \ 0]^T$ , where  $M_{w,BA} = 88.11 \text{kg kmol}^{-1}$  and  $\check{w}_{in,l,BA} = 1$ . The outlet mass flow rate  $u_{out,l}(t)$  is regulated in real-time to keep the liquid mass constant at  $4600 \text{kg}$  throughout the reaction. The gas phase is fed with pure chlorine ( $p_g = 1$ ) with composition  $\mathbf{w}_{in,g} = [M_{w,\text{Cl}_2}^{-1} \check{w}_{in,g,\text{Cl}_2} \ 0 \ 0]^T$ , where  $\check{w}_{in,g,\text{Cl}_2} = 1$ . The inlet mass flow rate  $u_{in,g}(t)$  is also manipulated in real-time to keep the total pressure  $P(t) = \mathbf{1}_S^T \mathbf{c}_g(t) \mathcal{R} T$  constant. Gas is withdrawn from the reactor at the constant mass flow rate  $u_{out,g} = 3600 \text{kg h}^{-1}$ . The density of the liquid reaction mixture is assumed to vary with composition. The volume  $V_l(t)$  is calculated as the sum of the individual volumes of all pure compounds present in the liquid. As the reactor volume is constant, the volume of the gas phase is  $V_g(t) = V_{\text{tot}} - V_l(t)$ .

The concentrations  $\mathbf{c}_l(t) = \mathbf{n}_l(t) / V_l(t)$  and  $\mathbf{c}_g(t) = \mathbf{n}_g(t) / V_g(t)$  are obtained by integration of the differential mole balances in Eqs. (1a) and (1b). Zero-mean normally distributed noise with standard deviation of 1% relative to the species maximal concentration is added to each concentration measurement in the liquid and gas phases. Two other sets of concentration data are also generated using noise levels of 2% and 5%.

In the liquid phase, absorbance spectra are assumed to be measured in the mid-IR region ( $500\text{--}1500 \text{cm}^{-1}$ ) with  $4 \text{cm}^{-1}$  resolution, thus leading to  $W = 250$  wavenumbers recorded simultaneously. The spectrum of the solvent (ethanol) is assumed to be subtracted from all absorbance spectra. Chlorine and hydrochloric acid are considered non-absorbing and thus there are  $S_a = 3$  absorbing species {BA, MBA, DBA}. The absorbance spectra  $\mathbf{a}(t)$  are generated by Eq. (10) from the computed concentrations  $\mathbf{c}_{l,a}(t)$  (corrupted with 1%, 2% and 5% noise) and the  $(3 \times 250)$ -dimensional pure component spectra **E** chosen to resemble the true ones. The pure component spectra are shown in Fig. 6.

<sup>4</sup> Note that the rate of mass transfer of the species  $s$  is negative if  $c_{l,s} > c_{l,s}^*$ .



**Fig. 6.** Chlorination of butanoic acid: Simulated mid-IR (500–1500  $\text{cm}^{-1}$ ) pure component spectra (absorptivities) of (a) butanoic acid, (b)  $\alpha$ -mono-chloro-butanoic acid and (c)  $\alpha$ -di-chloro-butanoic acid, normalized to a unit peak height.

Measurements in the gas phase are also available in the form of concentration measurements for the species  $\text{Cl}_2$  and  $\text{HCl}$ . If the quantities  $u_{in,g}(t)$ ,  $u_{out,g}$ ,  $m_g(t)$  and  $V_g(t)$  are known, the  $p_m = 2$  extents of mass transfer can be computed using the MV-form of the measured numbers of moles  $n_{g,\text{Cl}_2}(t) = c_{g,\text{Cl}_2}(t)V_g(t)$  and  $n_{g,\text{HCl}}(t) = c_{g,\text{HCl}}(t)V_g(t)$ , as described in Eq. (6). The resulting extents of mass transfer  $\mathbf{x}_{m,g}(t)$  are used to compute the extents of mass transfer  $\mathbf{x}_{m,l}(t)$  for the liquid phase according to Eqs. (7) and (8). As the experimental conditions  $u_{in,l}$ ,  $u_{out,l}(t)$ ,  $m_l$  and  $V_l(t)$  are known, the extents of inlet flow  $x_{in,l}(t)$  and the discounting variable  $\lambda_l(t)$  can be computed from (3c) and (3d). Hence, since the inlet spectrum (pure component spectrum of butanoic acid) is also known, the spectral contributions of  $\mathbf{x}_{m,l}(t)$ ,  $x_{in,l}(t)$  and  $\lambda_l(t)$  can be subtracted from the absorbance spectrum  $\mathbf{a}(t)$  using Eq. (16), which allows computing the spectral contribution of the reactions  $\mathbf{a}^{RV}(t)$ . Note that, since the transferring species  $\text{Cl}_2$  and  $\text{HCl}$  are spectroscopically inactive ( $\mathbf{A}_m = \mathbf{0}_{2 \times 250}$ ), the extents of mass transfer are not needed to compute the spectral RV-form  $\mathbf{a}^{RV}(t)$ . The noise levels of 1%, 2% and 5% in  $\mathbf{c}(t)$  lead to noise levels of the same standard deviation in  $\mathbf{c}^{RV}(t)$ . The resulting noise levels in  $\mathbf{a}(t)$  and  $\mathbf{a}^{RV}(t)$  are 1%, 1.7% and 4%, respectively.

#### 4.2.3. Calibration experiment

The reactor is initially loaded with 5.63  $\text{m}^3$  of a solution of butanoic acid in ethanol that contains traces of MBA. This results in the initial conditions  $\mathbf{n}_{l0,cal} = [20 \ 0 \ 10^{-5} \ 0 \ 0 \ 61.5]^T$  kmol for the liquid phase. Only air is initially present in the gas phase so that the initial conditions are  $\mathbf{n}_{g0,cal} = [0 \ 0 \ 0.102]^T$  kmol and  $V_{g0,cal} = 3.37 \text{ m}^3$ . Butanoic acid is present initially, and there is no inlet flow in the liquid phase ( $p_l = 0$ ). The inlet mass flow rate to the gas phase is manipulated to keep the total pressure constant at 2 bar. Because of the limiting amount of butanoic acid present in the reactor, the experiment only lasts 108 min.

The calibration set comprises  $H = 12$  absorbance spectra recorded every 9 min. Each time a spectrum is measured, it is assumed that a sample is taken from the reaction mixture and analyzed offline to quantify the concentrations of BA, MBA and DBA. This measurement campaign results in 12 pairs of calibration vectors ( $\mathbf{c}_{l,cal}(t)$ ,  $\mathbf{a}_{cal}(t)$ ) for the  $S_c = 3$  species {BA, MBA, DBA}. Since there is no liquid inlet ( $x_{in,l,cal}(t) = 0$ ) and no calibrated species involved in the mass transfers ( $\mathbf{W}_{m,l,c} = \mathbf{0}_{3 \times 2}$ ), this calibration experiment does not require gas-phase measurements, and only the concentration and spectral contributions of  $\lambda_{l,cal}(t)$  have to be removed from  $\mathbf{c}_{l,cal}(t)$  and  $\mathbf{a}_{cal}(t)$  to generate  $\mathbf{c}_{l,cal}^{RV}(t)$  and  $\mathbf{a}_{cal}^{RV}(t)$ . This pretreatment with Eqs. (15) and (16) leads to the reduction of the matrix/vector dimension from  $S_l = 6$  species to  $S_c = 3$  calibrated species by removing the columns or rows corresponding to the species  $\text{Cl}_2$ ,  $\text{HCl}$  and  $\text{EtOH}$ .

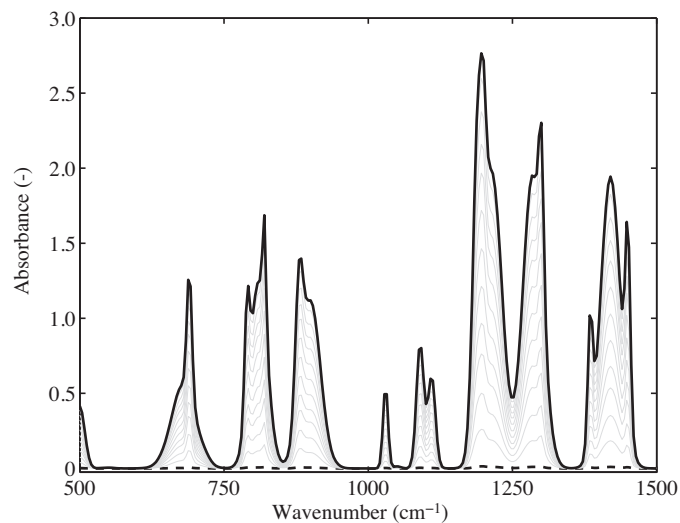
The  $(3 \times 250)$ -dimensional calibration matrices  $\mathbf{F}$  and  $\mathbf{F}^{RV}$  are constructed by PLS using  $\varepsilon = 3$  latent variables. One extra latent

variable ( $\varepsilon = 3$  instead of  $R = 2$ ) is necessary for the calibration model based on  $\mathbf{F}^{RV}$  to account for the effect of noise.

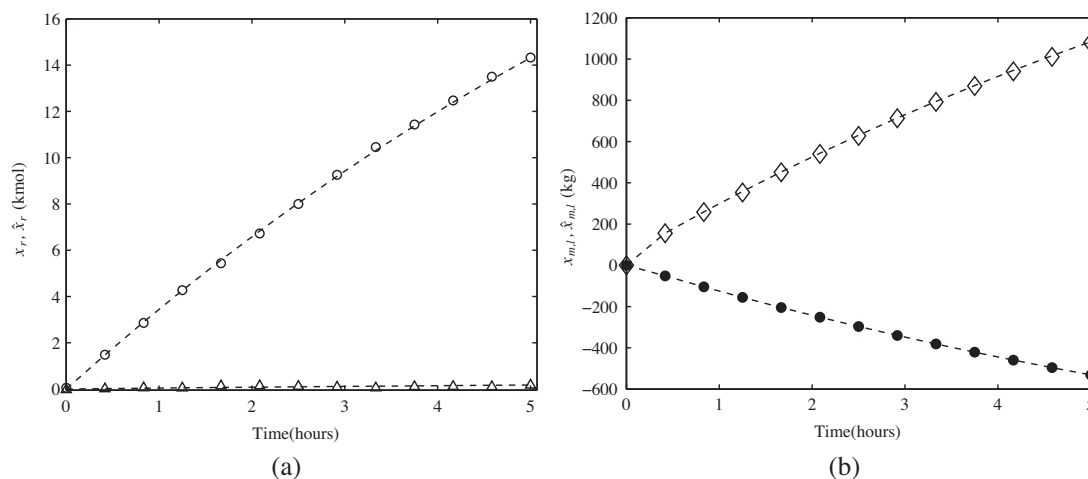
#### 4.2.4. Prediction experiment

The prediction experiment is performed under different experimental conditions. The reactor is loaded with  $V_{l0} = 5.82 \text{ m}^3$  of a liquid containing 100 kmol of ethanol and a small amount of MBA, so that the initial conditions are  $\mathbf{n}_{l0} = [0 \ 0 \ 10^{-5} \ 0 \ 0 \ 100]^T$  kmol,  $\mathbf{n}_{g0} = [0 \ 0 \ 0.0948]^T$  kmol and  $V_{g0} = 3.18 \text{ m}^3$ . The liquid phase is fed with pure butanoic acid ( $p_l = 1$ ) at the constant mass flow rate  $u_{in,l} = 324 \text{ kg h}^{-1}$ . The inlet mass flow rate  $u_{in,g}(t)$  is manipulated to keep the total pressure constant at 10 bar. The absorbance spectra of the prediction experiment are assumed to be measured every 5 min for 5 h, as shown in Fig. 7 for the case of 1% measurement noise.

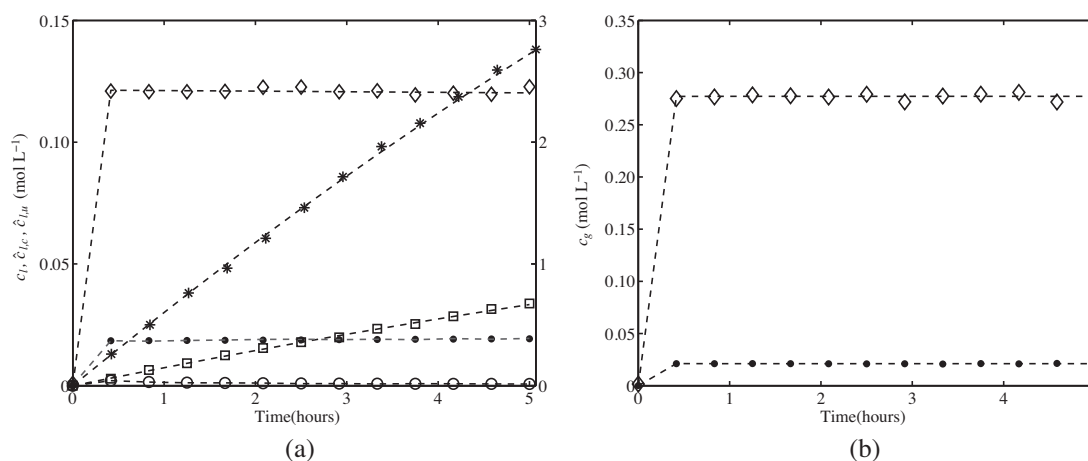
Since  $\text{rank}([\mathbf{N}_c^T \ \mathbf{w}_{in,l,c} \ \mathbf{n}_{l0,c}]) = R + p_l + 1 < 4$ , the linear transformation of Path 4 cannot be used, despite the fact that no calibrated species are involved in the mass transfers ( $\mathbf{W}_{m,l,c} = \mathbf{0}$ ). However, owing to the choice of the calibrated species,  $\hat{\mathbf{c}}_{l,c}(t)$  and  $\mathbf{a}(t)$  can be rearranged in  $\hat{\mathbf{c}}_{l,c}^{RV}(t)$  and  $\mathbf{a}^{RV}(t)$ , and thus either Path 3 – rearrangement of  $\mathbf{a}(t)$  in  $\mathbf{a}^{RV}(t)$  and then prediction of  $\hat{\mathbf{c}}_{l,c}^{RV}(t)$  via  $\mathbf{F}^{RV}$  – or Paths 1 and 6 – prediction of  $\hat{\mathbf{c}}_{l,c}(t)$  via  $\mathbf{F}$  and then rearrangement in  $\hat{\mathbf{c}}_{l,c}^{RV}(t)$  – can be followed. Both these routes involve the subsequent use of Eq. (23) to predict the extents of reaction. On the other hand, the extents of mass transfer will be obtained from the concentration measurements for the species  $\text{Cl}_2$  and  $\text{HCl}$  in the gas phase.



**Fig. 7.** Chlorination of butanoic acid: Mid-IR absorbance spectra  $\mathbf{a}(t)$  with 1% noise, measured every 5 min but plotted every 25 min. The black continuous and dashed lines indicate the first and the last spectra, respectively.



**Fig. 8.** Prediction experiment for chlorination of butanoic acid: (a) Noise-free simulated (dashed line) and noisy predicted (markers) extents of reaction for R1 ( $\circ$ ) and R2 ( $\Delta$ ), the prediction using Path 1 (PLS model) and Path 6 in Fig. 2 in the presence of 1% measurement noise; (b) noise-free simulated (dashed line) and noisy predicted (markers) extents of mass transfer for species  $\text{Cl}_2$  ( $\diamond$ ) and HCl ( $\bullet$ ), plotted every 25 min.



**Fig. 9.** Prediction experiment for chlorination of butanoic acid: (a) Noise-free simulated concentrations (dashed line) and noisy concentrations predicted via Path 1 (PLS model) and Path 6 for species BA ( $\circ$ , left axis),  $\text{Cl}_2$  ( $\diamond$ , left axis), MBA ( $*$ , right axis), HCl ( $\bullet$ , left axis) and DBA ( $\square$ , left axis) in the presence of 1% measurement noise; (b) noise-free simulated (dashed line) and noisy measured (markers) concentrations of species  $\text{Cl}_2$  ( $\diamond$ ) and HCl ( $\bullet$ ), plotted every 25 min.

As described in Section 4.2.2, these measurements allow computing the  $p_m = 2$  extents of mass transfer using the MV-form of the measured numbers of moles in the gas phase. The predicted extents of reaction and mass transfer obtained from absorbance spectra corrupted with 1% noise agree very well with their simulated values, as seen in Fig. 8.

The liquid concentrations  $\hat{c}_{i,u}(t)$  of the remaining  $S_u = 3$  uncalibrated species  $\{\text{Cl}_2, \text{HCl}, \text{EtOH}\}$  can be reconstructed from the predicted extents using Eq. (22). Fig. 9 compares simulated and predicted concentrations.

#### 4.2.5. Kinetic identification

As the objective of this study is the investigation of the two reaction rates, the prediction of the mass-transfer extents will be omitted, which illustrates the remark in Section 2.3 stating that, even in the absence of appropriate models for some of the extents, the remaining extents can be modeled correctly. Furthermore, reconstruction of the EtOH and HCl concentrations is not needed since these species are not involved in the rate expressions of R1 and R2.

The least-squares problem consists in adjusting the rate constant of each reaction by comparing the extents predicted by calibration and those modeled by a postulated rate law. The identification is performed using the extents of reaction predicted via both  $\mathbf{F}$  (Paths 1 and 6) and  $\mathbf{F}^{RV}$  (Path 3). Table 3 shows the estimated rate parameters with their 99% confidence interval resulting from absorbance spectra with 1% noise. The estimation is excellent, regardless of the calibration procedure that is used. Very similar results have been obtained using PCR (not shown in Table 3).

**Table 3**

Rate constants of the chlorination of butanoic acid: Simulated  $k_i$  and estimated  $\hat{k}_i$  rate constants obtained by fitting the predicted extents of reaction  $\hat{x}_r$  obtained by PLS via  $\mathbf{F}$  and  $\mathbf{F}^{RV}$ . The rate constant for reaction R1 is given in  $(\text{m}^3)^{3/2} \text{kmol}^{-3/2} \text{s}^{-1}$  and in  $(\text{m}^3)^{5/2} \text{kmol}^{-5/2} \text{s}^{-1}$  for reaction R2. The corresponding 99% confidence intervals are also given.

Reaction	Simulation $k_i$	Estimation via $\mathbf{F}$ $\hat{k}_i$ [99% C.I.]	Estimation via both $\mathbf{F}^{RV}$ $\hat{k}_i$ [99% C.I.]
R1	1.358	1.347 [1.333, 1.361]	1.361 [1.349, 1.373]
R2	0.136	0.136 [0.112, 0.161]	0.133 [0.108, 0.159]

The sensitivity of the estimated rate constants with respect to the noise level was investigated by using sets of calibration and prediction data corrupted with higher noise levels. It was observed that the range of the confidence intervals obtained via both  $\mathbf{F}$  and  $\mathbf{F}^{RV}$  are multiplied by 2 when the noise level is increased to 1.7%. For an increase of the noise level from 1% to 4.2%, the confidence intervals increase by a factor 3 to 5, depending on which extent of reaction is fitted and which calibration procedure is used.

## 5. Conclusion

PCR and PLS calibration techniques are widely used to predict concentrations from spectroscopic measurements. This article has discussed ways of predicting both extents and concentrations from absorbance measurements, which are then used to model the rate laws by the extent-based method of identification.

Efficient design of the calibration set consists in using reaction data measured in a preliminary experiment. Two sets of calibration pairs can be used, namely, the standard pairs of concentrations and absorbances or the pairs of concentrations and absorbances in RMV-form. In addition, when gas-phase measurements can be used to compute the concentration and spectral contributions of the mass transfers, the calibration pairs can be pretreated in RV-form. Calibration models based on calibration pairs pretreated in RMV- (or RV-) form use absorbance data in that form to predict concentrations in RMV- (or RV-) form. The use of both sets of calibration pairs has been illustrated.

The choice of the most appropriate calibration pairs depends on the fulfillment of the space-inclusion condition. If the space-inclusion condition is fulfilled, it can be preferable to use a calibration set based on the pairs  $(\mathbf{c}_{l,cal}(t), \mathbf{a}_{cal}(t))$ , even if the corresponding calibration models require more latent variables. This is due to the fact that this route (Paths 1 and 4 in Fig. 2) requires the least amount of measured information (only  $V_l(t)$ ) and thus is less prone to uncertainties in experimental conditions. This recommendation is particularly true when the extents can be predicted via the linear transformation in Path 4. When the space-inclusion condition is not fulfilled for the pairs  $(\mathbf{c}_{l,cal}(t), \mathbf{a}_{cal}(t))$ , pretreatment in RMV- or RV-form becomes a requirement for both building the calibration set and using the calibration model.

This paper has illustrated the potential of extent-based kinetic identification using spectroscopic measurements and multivariate calibration. The approach works well when a sufficient number of concentrations can be predicted from absorbance measurements and when the noise level is small. Further work is necessary to assess the effect of measurement noise on both the bias and the variance of the estimated kinetic parameters, but also to validate experimentally the potential of the method.

## References

- [1] D. Bonvin, Optimal operation of batch reactors – a personal view, *J. Proc. Control* 8 (1998) 355–368.
- [2] W. Marquardt, Model-based experimental analysis of kinetic phenomena in multi-phase reactive systems, *Chem. Eng. Res. Des.* 83 (2005) 561–573.
- [3] A. Bardow, W. Marquardt, Incremental and simultaneous identification of reaction kinetics: methods and comparison, *Chem. Eng. Sci.* 59 (2004) 2673–2684.
- [4] G. Puxty, M. Maeder, K. Hungerbühler, Tutorial on the fitting of kinetics models to multivariate spectroscopic measurements with non-linear least squares regression, *Chemometr. Intell. Lab. Syst.* 81 (2006) 149–164.
- [5] J. Billeter, Y.-M. Neuhold, L. Simon, G. Puxty, K. Hungerbühler, Uncertainties and error propagation in kinetic hard-modelling of spectroscopic data, *Chemometr. Intell. Lab. Syst.* 93 (2008) 120–131.
- [6] N. Bhatt, N. Kerimoglu, M. Amrhein, W. Marquardt, D. Bonvin, Incremental identification of reaction systems – a comparison between rate-based and extent-based approaches, *Chem. Eng. Sci.* 83 (2012) 24–38.
- [7] G. Jia, G. Stephanopoulos, R. Gunawan, Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles, *Metabolites* 2 (2012) 891–912.
- [8] M.L. Brendel, D. Bonvin, W. Marquardt, Incremental identification of kinetic models for homogeneous reaction systems, *Chem. Eng. Sci.* 61 (2006) 5404–5420.
- [9] N. Bhatt, M. Amrhein, D. Bonvin, Incremental identification of reaction and mass-transfer kinetics using the concept of extents, *Ind. Eng. Chem. Res.* 50 (2011) 12960–12974.
- [10] N. Bhatt, Extents of Reaction and Mass Transfer in the Analysis of Chemical Reaction Systems, Ph.D. thesis, No. 5028, EPFL, Lausanne, Switzerland, 2011.
- [11] S. Srinivasan, J. Billeter, D. Bonvin, Extent-based incremental identification of reaction systems using concentration and calorimetric measurements, *Chem. Eng. J.* 207–208 (2012) 785–793.
- [12] J. Workman, B. Lavine, R. Chrisman, M. Koch, Process analytical chemistry, *Anal. Chem.* 83 (2011) 4557–4578.
- [13] Y. Roggo, P. Chaluz, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, *J. Pharm. Biomed. Anal.* 44 (2007) 683–700.
- [14] W. Chew, P. Sharratt, Trends in process analytical technology, *Anal. Methods* 2 (2010) 1412–1438.
- [15] E. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons, New York, USA, 2002.
- [16] P. Geladi, B. Kowalski, Partial least-squares regression – a tutorial, *Anal. Chim. Acta* 185 (1986) 1–17.
- [17] G. Escandar, N. Faber, H. Goicoechea, A. de la Pena, A. Olivieri, R. Poppi, Second- and third-order multivariate calibration: data, algorithms and applications, *TRAC – Trends Anal. Chem.* 26 (2007) 752–765.
- [18] M. Maeder, Y.-M. Neuhold, *Practical Data Analysis in Chemistry*, Elsevier, Amsterdam, NL, 2007.
- [19] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Calibration of spectral reaction data, *Chemometr. Intell. Lab. Syst.* 46 (2) (1999) 249–264.
- [20] P.J. Gemperline, *Practical Guide to Chemometrics*, Taylor and Francis, Boca Raton, USA, 2006.
- [21] H. Martens, T. Naes, *Multivariate Calibration*, John Wiley & Sons, Chichester, UK, 2002.
- [22] Y. Wang, D. Veltkamp, B. Kowalski, Multivariate instrument standardization, *Anal. Chem.* 63 (1991) 2750–2756.
- [23] N. Bhatt, M. Amrhein, D. Bonvin, Extents of reaction, mass transfer and flow for gas–liquid reaction systems, *Ind. Eng. Chem. Res.* 49 (2010) 7704–7717.
- [24] M. Amrhein, N. Bhatt, B. Srinivasan, D. Bonvin, Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams, *AIChE J.* 56 (2010) 2873–2886.
- [25] M. Amrhein, Reaction and Flow Variants/Invariants for the Analysis of Chemical Reaction Data, Ph.D. thesis, No. 1861, EPFL, Lausanne, Switzerland, 1998.
- [26] G. Maria, A. Dan, D.N. Stefan, Model-based derivation of the safety operating limits of a semi-batch reactor for the catalytic acetoacetylation of pyrrole using a generalized sensitivity criterion, *Chem. Biochem. Eng. Q.* 24 (2010) 265–281.