Incremental Identification of Reaction and Mass–Transfer Kinetics Using the Concept of Extents

Nirav Bhatt, Michael Amrhein, and Dominique Bonvin[*]

Laboratoire d'Automatique

École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

E-mail: dominique.bonvin@epfl.ch

Abstract

This paper proposes a variation of the incremental approach to identify reaction and mass-transfer kinetics (rate expressions and the corresponding rate parameters) from concentration measurements for both homogeneous and gas-liquid reaction systems. This incremental approach proceeds in two steps: (i) computation of the extents of reaction and mass transfer from concentration measurements without explicit knowledge of the reaction and mass-transfer rate expressions, and (ii) estimation of the rate parameters for each rate expression individually from the computed extents using the integral method. The novelty consists in using extents that are computed from measured concentrations. For the computation of the individual extents, two cases are considered: if the concentrations of all the liquid-phase species can be measured, a linear transformation is used; otherwise, if the concentrations of only a subset of the liquid-phase species are available, an approach that uses flowrate and possibly gas-phase concentration measurements is proposed. The incremental identification approach is illustrated

[*]To whom correspondence should be addressed

1

in simulation via two reaction systems, namely the homogeneous acetoacetylation of pyrrole and the gas-liquid chlorination of butanoic acid.

Introduction

Dynamic models are used to analyze, monitor, control and optimize reaction systems. These models are often based on first principles and describe the evolution of the states (number of moles, temperature and volume) by means of conservation equations of differential nature and constitutive equations of algebraic nature. The models include information regarding the underlying reactions (stoichiometry and kinetics), the transfer of species between phases (mass-transfer kinetics), and the operation of the reactor (initial conditions, inlet and outlet flows, operational constraints). The identification of reaction and mass-transfer kinetics (rate expressions and the corresponding rate parameters) represents the main challenge in building first-principles models for reaction systems. The rate expressions, which are typically chosen from a set of candidates, need to be confronted to measured data. The identification task can be performed globally in one step via a simultaneous approach, or successively over several steps via an incremental approach, as discussed next.

Simultaneous identification proceeds as follows. From a library of reaction pathways and rate expressions, one chooses a rate expression candidate for the reaction system and estimates the rate parameters by comparing model predictions and measured data. The approach is termed 'simultaneous identification' since all reaction and mass transfer kinetics are identified simultaneously. The procedure needs to be repeated for all rate expression candidates. The candidate with the best fit is usually selected. Issues like parameter and structural identifiability [1,2] and experimental planning [3,4] are important to guarantee parameter estimates with little correlation and narrow confidence intervals. The main advantage of simultaneous identification is that it can deal with complex reaction and mass-transfer kinetics and lead to optimal parameters in the maximum-likelihood sense. [5] However, simultaneous identification can be computationally costly when several candidates are available for each rate expression. Furthermore, since the global model is fitted so as to reduce the

prediction error, structural mismatch in one part of the model will typically result in errors in all the rate parameters. Finally, it is often difficult to choose suitable initial parameter values, which may lead to convergence problems.[6]

On the other hand, *incremental identification* decomposes the identification task into a set of subproblems.[7–10] First, the reaction stoichiometry is identified from measured concentrations. For this, each reaction can be determined individually without explicit knowledge of reaction kinetics using target factor analysis (TFA).[11,12] Incremental TFA has been proposed to remove the variability associated with a reaction, once it has been accepted, before continuing target testing for the other reactions.[13] The next step computes the rate profiles of reaction and mass transfer from measured data and the known stoichiometry. Finally, the rate parameters are estimated from the computed rate profiles. For each subproblem, the number of model candidates can be kept low. This approach is also termed 'individual identification' since each reaction and mass transfer can be dealt with individually.

Regarding the identification of reaction and mass-transfer kinetics from measured data, two methods can be distinguished depending on the way data are handled, namely the differential and the integral methods.[14,15] These two methods are detailed next.

In the *differential method*, reaction rate profiles are computed through differentiation of concentration data. Furthermore, individual rate profiles can be computed upon knowledge of the stoichiometry. Then, for a given reaction, a rate expression is proposed and its parameters are estimated by fitting the simulated rate profile to the computed values. Note that the differentiation of measured concentrations is a difficult task due to noise and the sparsity of measurements.[16]

In the *integral method*, the rate expressions are integrated analytically or numerically to predict concentrations, and the unknown rate parameters are estimated by fitting these predictions to measured concentrations. The integral method is computationally intensive because of the need to integrate the rate expressions for each set of parameter values proposed by the optimization algorithm. However, in the absence of structural uncertainty and for Gaussian mea-

3

surement noise, the integral method leads to optimal estimates in the maximum-likelihood sense.[6,17]

The simultaneous and incremental approaches, which use the integral and differential methods, respectively, are illustrated in Figure 1 for homogeneous reaction systems:

Path "1" indicates the simultaneous identification approach that uses the integral method, whereby the rate expressions for all reactions are integrated to predict concentrations that are fitted to measured values via a least-squares problem.

Path "2" represents the incremental identification approach that uses the differential method, whereby the rate profile of the ith reaction is computed by differentiation of concentration measurements and use of information regarding the stoichiometry, the inlet composition, the volume, and the inlet and outlet flowrates. The ith rate model, which is chosen from a library of rate expressions, is fitted to the computed rate profile via a least-squares problem. Unfortunately, numerical differentiation introduces a bias in the computed rate profiles, thus leading to suboptimal parameters.[6,7] Hence, as part of a final adjustment step, simultaneous identification using the model structure identified by incremental approach is often performed to obtain unbiased parameter estimates.

For the sake of completeness, two special cases of generalized simultaneous and incremental approaches available in the literature are briefly mentioned next.

A framework for automatic modeling of chemical/biochemical reaction systems (TAM-C/B) based on concentrations and calorimetric data has been proposed.[18] TAM uses an automatic iterative procedure that imitates the human expert in modeling reaction systems. From measured data, TAM first generates a qualitative description of the dynamic behavior of the reaction system using a fuzzy interval identification method.[18] Then, based on the resulting qualitative description, prior knowledge regarding the reaction stoichiometry and a rule-based library, TAM postulates possible rate expressions and fits the global model to the concentrations and calorimetric data.

4

Experiments, measurements
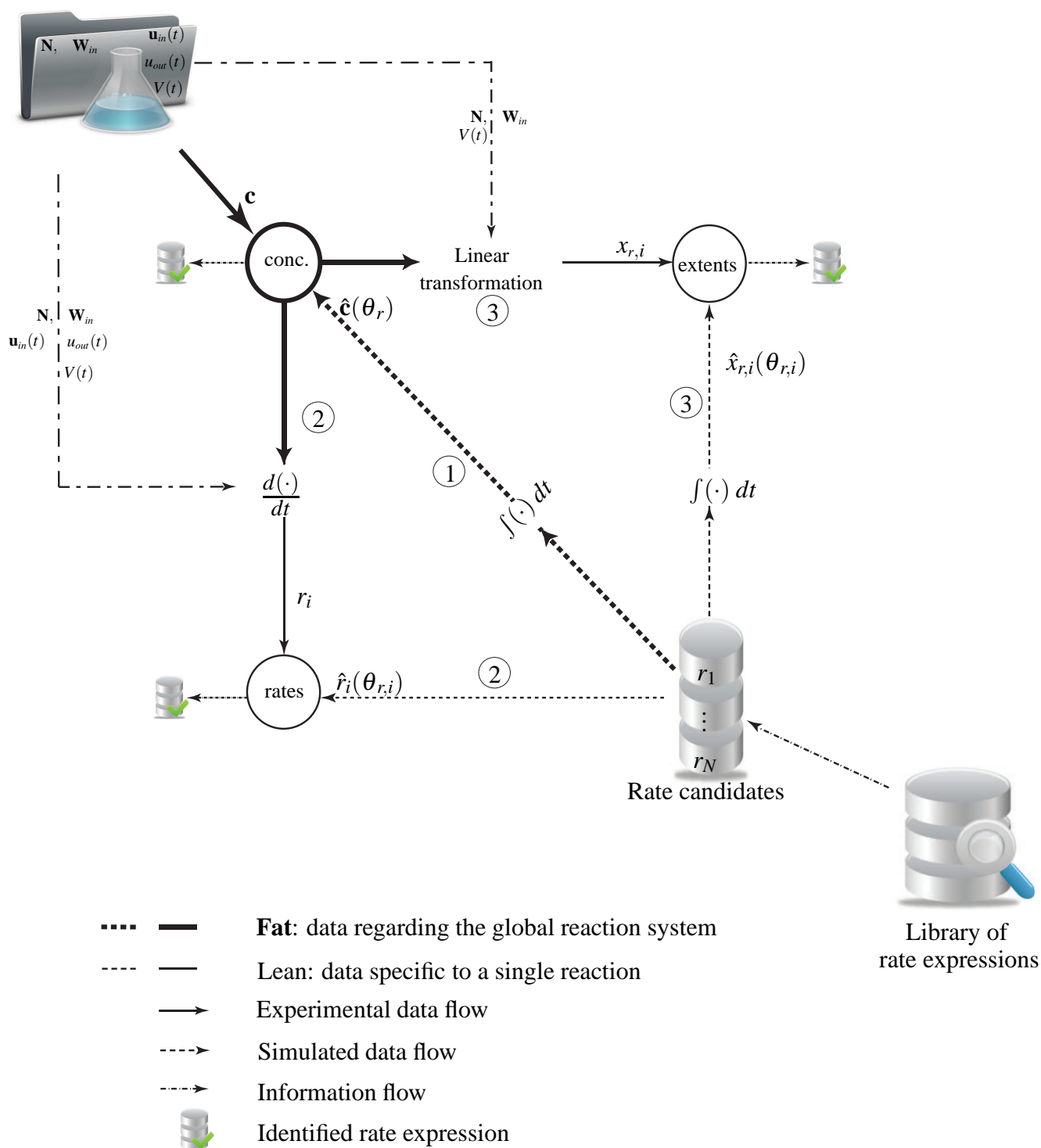and required information

Figure 1: Schematic comparison of the simultaneous and incremental identification approaches for homogeneous reaction systems. Path "1": simultaneous approach that uses the integral method to integrate all candidate rate expressions; Path "2": incremental approach that uses the differential method to differentiate concentrations; Path "3": incremental approach that uses the integral method to integrate a single rate expression at the time.

An iterative model identification framework that investigates model deficiencies and estimates nonparametric functional relationships from concentration data has also been proposed.[19,20] The authors suggest adding a stochastic process to selected mole balances exhibiting possible uncertainty. The mole balances with large fitted stochastic parameters are pinpointed as having model deficiencies. The modeler can then refine the pinpointed model equations in the next iteration.

This paper will develop an *incremental identification* approach that relies on the *integral method*, thereby combining the strengths of the incremental approach (can handle each rate individually) and the integral method (optimal handling of measurement noise). This endeavor will use the concept of *extents of reaction and mass transfer*.

In a batch reactor, the change in the extent of a reaction is given by the change in the number of moles of any species due to that reaction divided by the corresponding stoichiometric coefficient. Recently, Amrhein et al.[21] proposed a linear transformation that computes the extents of reaction from the numbers of moles in homogeneous reaction systems with inlet and outlet streams. The transformation uses only information regarding the stoichiometry, the inlet composition and the initial conditions, that is, it does not require rate expressions nor inlet and outlet flowrates. The approach has been extended to compute the extents of reaction and mass transfer in gas-liquid (G–L) reaction systems by Bhatt et al.[22] using only information regarding the stoichiometry, the inlet composition, the initial conditions, and the knowledge of the species transferring between phases.

As already mentioned, this paper proposes a novel variation of the incremental approach for the identification of reaction systems, which is based on the integral method. This novel incremental identification approach uses the concept of extents and corresponds to Path "3" in Figure 1. It proceeds in two steps as follows:

1. Data transformation. Measured concentrations are transformed, without explicit knowledge of rate expressions, to individual extents of reaction and mass transfer.

2. Identification of reaction and mass-transfer kinetics. The rate parameters of a candidate

reaction or mass-transfer rate expression are estimated by comparing predicted (using the integral method) and computed (using the transformation) extents.

In practice, concentrations of all the species are difficult to measure on-line. However, a subset of the concentrations and the flowrates can be measured on-line. For such cases, it is shown that the proposed incremental approach allows computing the individual extents of reaction and mass transfer using information on the inlet and outlet flowrates.

The paper is organized as follows. First, models of homogeneous and G–L reaction systems are briefly reviewed. Then, various procedures to compute the extents of reaction and mass transfer from measured data are developed; two cases are considered depending on whether or not all liquid-phase concentrations are measured. The next section describes the estimation of rate parameters from computed extents using the integral method. The proposed incremental identification approach is illustrated in simulation through two reaction systems, namely the acetoacetylation of pyrrole in a homogeneous reactor and the chlorination of butanoic acid in a G–L reactor. The last section concludes the paper.

Preliminaries

This section develops the models of homogeneous and G–L reaction systems with inlet and outlet streams.

Model of open homogeneous reaction systems

The mole balance equations for a homogeneous reaction system involving S species, R reactions, p inlet streams and one outlet stream can be written generically as follows:

$$\dot{\mathbf{n}}(t) = \mathbf{N}^{\mathrm{T}} V(t)\, \mathbf{r}(t) + \mathbf{W}_{in}\, \mathbf{u}_{in}(t) - \frac{u_{out}(t)}{m(t)}\mathbf{n}(t), \qquad \mathbf{n}(0) = \mathbf{n}_0, \tag{1}$$

where $\mathbf{n}$ is the S-dimensional vector of numbers of moles, $\mathbf{r}$ the R-dimensional reaction rate vector, $\mathbf{u}_{in}$ the p-dimensional inlet mass flowrate vector, u_{out} the outlet mass flowrate, V and m the volume and mass of the reaction mixture, $\mathbf{N}$ the $R \times S$ stoichiometric matrix, $\mathbf{W}_{in} = \mathbf{M}_w^{-1} \check{\mathbf{W}}_{in}$ the $S \times p$ inlet-composition matrix with $\mathbf{M}_w$ the S-dimensional diagonal matrix of molecular weights and $\check{\mathbf{W}}_{in} = [\check{\mathbf{w}}_{in}^1, \cdots, \check{\mathbf{w}}_{in}^p]$ with $\check{\mathbf{w}}_{in}^k$ being the S-dimensional vector of weight fractions of the kth inlet stream, and $\mathbf{n}_0$ the S-dimensional vector of initial numbers of moles. The flowrates $\mathbf{u}_{in}(t)$ and $u_{out}(t)$ are considered as independent (input) variables in Eq. (1). The way these variables are adjusted depends on the particular experimental situation; for example, some elements of $\mathbf{u}_{in}$ can be adjusted to control the temperature in a semi-batch reactor, or u_{out} can be a function of the inlet flows in a constant-volume reactor. The continuity (or total mass balance) equation reads:

$$\dot{m}(t) = \mathbf{1}_p^{\mathrm{T}} \mathbf{u}_{in} - u_{out}, \qquad m(0) = m_0, \tag{2}$$

where $\mathbf{1}_p$ is the p-dimensional vector filled with ones and m_0 the initial mass. Note that the mass $m(t)$ can also be computed from the numbers of moles as:

$$m(t) = \mathbf{1}_S^{\mathrm{T}} \mathbf{M}_w \mathbf{n}(t). \tag{3}$$

Indeed, from the relationships $\mathbf{1}_S^{\mathrm{T}} \mathbf{M}_w \mathbf{N}^{\mathrm{T}} = \mathbf{0}_R$ (each reaction conserves mass) and $\mathbf{1}_S^{\mathrm{T}} \mathbf{M}_w \mathbf{W}_{in} = \mathbf{1}_p^{\mathrm{T}}$ (the weight fractions of each inlet add up to unity), Eq. (2) can be obtained by differentiation of Eq. (3). Hence, the continuity equation (2) becomes redundant. The concentration of the sth species can be expressed in terms of the number of moles and the volume as follows:

$$c_s(t) = \frac{n_s(t)}{V(t)}, \qquad \forall s = 1, \ldots, S. \tag{4}$$

Model of open G–L reaction systems

The mole balance equations for a G–L reaction system are presented in this section. The gas and liquid phases will be modeled separately, with the mass-transfer rates $\boldsymbol{\zeta}$ connecting the two

phases. The gas phase contains S_g species, p_g inlets and one outlet, while the liquid phase contains S_l species, p_l inlets and one outlet. There are p_m mass transfer steps taking place between two phases. Let us consider the following assumptions:

(A1) The gas and liquid phases are homogeneous.

(A2) The reactor has a constant total volume.

(A3) The reactions take place in the liquid bulk only.

(A4) No accumulation in the boundary layer.

(A5) Mass transfer rates are considered positive from the gas to the liquid. Hence, negative mass-transfer rates are computed for species transferring from the liquid to the gas.

With these assumptions, the mole balances for the gas and liquid phases read:

Gas phase

$$\dot{\mathbf{n}}_g(t) = \mathbf{W}_{in,g}\,\mathbf{u}_{in,g}(t) - \mathbf{W}_{m,g}\,\boldsymbol{\zeta}(t) - \frac{u_{out,g}(t)}{m_g(t)}\mathbf{n}_g(t), \qquad \mathbf{n}_g(0) = \mathbf{n}_{g0}, \tag{5}$$

Liquid phase

$$\dot{\mathbf{n}}_l(t) = \mathbf{N}^{\mathrm{T}}V_l(t)\,\mathbf{r}(t) + \mathbf{W}_{in,l}\,\mathbf{u}_{in,l}(t) + \mathbf{W}_{m,l}\,\boldsymbol{\zeta}(t) - \frac{u_{out,l}(t)}{m_l(t)}\mathbf{n}_l(t), \qquad \mathbf{n}_l(0) = \mathbf{n}_{l0}, \tag{6}$$

where $\mathbf{n}_f$ is the S_f-dimensional vector of numbers of moles in the f phase, $f \in \{g,l\}$, $\mathbf{N}$ the $R \times S_l$ stoichiometric matrix, R the number of reactions, $\mathbf{W}_{in,f} = \mathbf{M}_{w,f}^{-1}\check{\mathbf{W}}_{in,f}$ 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 $\check{\mathbf{W}}_{in,f} = \begin{bmatrix} \check{\mathbf{w}}_{in,f}^1 & \cdots & \check{\mathbf{w}}_{in,f}^{p_k} \end{bmatrix}$ with $\check{\mathbf{w}}_{in,f}^k$ being the S_f-dimensional vector of weight fractions of the kth inlet to the f phase, $\mathbf{u}_{in,f}$ the p_f-dimensional inlet mass flowrate to the f phase, $\boldsymbol{\zeta}$ the p_m-dimensional mass-transfer rate vector, and $\mathbf{n}_{f0}$ the vector of initial moles in the f phase. $\mathbf{W}_{m,f} = \mathbf{M}_{w,f}^{-1}\check{\mathbf{E}}_{m,f}$ is the $S_f \times p_m$ mass-transfer matrix to the f phase, $\check{\mathbf{E}}_{m,f} = \begin{bmatrix} \check{\mathbf{e}}_{m,f}^1 & \cdots & \check{\mathbf{e}}_{m,f}^{p_m} \end{bmatrix}$

with $\check{\mathbf{e}}_{m,f}^{k}$ being the S_f-dimensional vector with the element corresponding to the kth transferring species equal to unity and the other elements equal to zero.

The relationships between mole fractions, concentrations and numbers of moles are:

$$y_s(t) = \frac{n_{g,s}(t)}{\mathbf{1}_{S_g}^{\mathrm{T}} \mathbf{n}_g(t)}, \qquad \forall s = 1, ..., S_g, \tag{7}$$

$$c_s(t) = \frac{n_{l,s}(t)}{V_l(t)}, \qquad \forall s = 1, ..., S_l. \tag{8}$$

Note that the reactor masses $m_f(t)$ can be inferred from the numbers of moles as:

$$m_f(t) = \mathbf{1}_{S_f}^{\mathrm{T}} \mathbf{M}_{w,f} \mathbf{n}_f(t), \qquad f \in \{g, l\}. \tag{9}$$

Upon grouping the inlet flowrates $\mathbf{u}_{in,f}$ and the mass-transfer rates $\boldsymbol{\zeta}$, Eqs. (5) and (6) become:

$$\dot{\mathbf{n}}_g(t) = \bar{\mathbf{W}}_{in,g} \, \bar{\mathbf{u}}_{in,g}(t) - \frac{u_{out,g}(t)}{m_g(t)} \mathbf{n}_g(t), \qquad \mathbf{n}_g(0) = \mathbf{n}_{g0}, \tag{10}$$

$$\dot{\mathbf{n}}_l(t) = \mathbf{N}^{\mathrm{T}} V_l(t) \, \mathbf{r}(t) + \bar{\mathbf{W}}_{in,l} \, \bar{\mathbf{u}}_{in,l}(t) - \frac{u_{out,l}(t)}{m_l(t)} \mathbf{n}_l(t), \qquad \mathbf{n}_l(0) = \mathbf{n}_{l0}, \tag{11}$$

where $\bar{\mathbf{W}}_{in,g} = [\mathbf{W}_{in,g}, -\mathbf{W}_{m,g}]$ is a matrix of dimension $S_g \times \bar{p}_g$, $\bar{\mathbf{W}}_{in,l} = [\mathbf{W}_{in,l}, \mathbf{W}_{m,l}]$ a matrix of dimension $S_l \times \bar{p}_l$, and $\bar{\mathbf{u}}_{in,f} = \begin{bmatrix} \mathbf{u}_{in,f} \\ \boldsymbol{\zeta} \end{bmatrix}$ a vector of dimension $\bar{p}_f$, with $\bar{p}_f = p_f + p_m$, $f \in \{g, l\}$.

Throughout this paper, the R reactions, and the $\bar{p}_f$ extended inlets are assumed to be independent, according to the definitions given in Amrhein *et al.*[21] and Bhatt *et al.*[22]

Computation of extents of reaction and mass transfer

The transformation of concentration data to the extents of reaction and mass transfer is presented next for both homogeneous and G–L reaction systems. Note that the transformation of concentration data is performed without knowledge of the reaction and mass-transfer kinetics.

Homogeneous reaction systems

For homogeneous reaction systems, two cases will be distinguished depending upon the available measurements: (a) all concentrations are measured, and (b) only a subset of the concentrations are measured.

All concentrations measured: Linear transformation

Let $\mathbf{c}(t_h)$ be the concentrations measured at time instant t_h, with $h = 0, 1, \ldots, H$. Let $V(t)$ denote the volume measured at time t.[1] The numbers of moles $\mathbf{n}(t_h)$ can be computed as $V(t_h)\mathbf{c}(t_h)$. The following proposition states the conditions to be able to compute the extents of reaction $\mathbf{x}_r(t_h)$ from $\mathbf{n}(t_h)$ using the linear transformation proposed in Amrhein *et al.*[21]

Proposition 1 (Linear transformation of $\mathbf{n}(t_h)$)

Consider the following assumptions:

(i) $\mathbf{N}$, $\mathbf{W}_{in}$ *and* $\mathbf{n}_0$ *are known,*

(ii) $\mathrm{rank}\left([\mathbf{N}^{\mathrm{T}} \ \mathbf{W}_{in} \ \mathbf{n}_0]\right) = R + p + 1$, *and*

(iii) $\mathbf{n}(t_h)$ *is available from measurements.*

Then, the extent of the ith reaction $x_{r,i}(t_h), \forall i = 1, \ldots, R$, can be computed as:

$$x_{r,i}(t_h) = (\mathbf{S}_0^{\mathrm{T}})_i \, \mathbf{n}(t_h), \tag{12}$$

where $(\mathbf{S}_0^{\mathrm{T}})_i$ denotes the ith row of the $(R \times S)$-dimensional $\mathbf{S}_0^{\mathrm{T}}$ matrix defined in Appendix A.

Furthermore, the extents of the kth inlet $x_{in,k}(t_h)$, $k = 1, \ldots, p$, and the outlet extent $x_{out}(t_h)$ can

---

[1]In practice, concentrations are typically measured infrequently, whereas volumes and flowrates are available nearly continuously. This leads to infrequent concentration data, denoted $\mathbf{c}(t_h)$, and frequent volume, inlet flowrates and outlet flowrate, denoted $V(t)$, $\mathbf{u}_{in}(t)$ and $u_{out}(t)$, respectively.

be computed from $\mathbf{n}(t_h)$ *as follows:*

$$x_{in,k}(t_h) = (\mathbf{M}_0^T)_k \, \mathbf{n}(t_h),$$

$$x_{out}(t_h) = 1 - \mathbf{q}_0^T \mathbf{n}(t_h),$$

(13)

where $(\mathbf{M}_0^T)_k$ *denotes the kth row of the* $(p_l \times S)$-*dimensional* $\mathbf{M}_0^T$ *matrix. The matrices* $\mathbf{M}_0$ *and* $\mathbf{q}_0$ *are defined in Appendix A.*

(Proof follows from Theorem 2 in Amrhein et al.[21])

Note that the transformation in Proposition 1 computes the extents of reaction and flow from the measured numbers of moles without the knowledge of inlet and outlet flowrates. However, the inlet and outlet flowrates are required in the identification of reaction and mass-transfer kinetics in the next step.

A subset of concentrations measured: Flow-dependent approach

Let $\mathbf{n}_a(t_h)$ be the numbers of moles of the S_a available species at time instant t_h. In practice, the inlet and outlet flowrates can be measured frequently. Let $\mathbf{u}_{in}(t)$ and $u_{out}(t)$ denote the inlet and outlet flowrates measured at time t. The following proposition states the conditions to be able to compute the extents of reaction $\mathbf{x}_r(t_h)$ from $\mathbf{n}_a(t_h)$ using the measured inlet and outlet flowrates.

Proposition 2 (Flow-dependent approach: Use of $\mathbf{n}_a(t_h)$ and integration of $\mathbf{u}_{in}(t)$ and $u_{out}(t)$)
Consider the following assumptions:

(i) $\mathbf{N}_a$, $\mathbf{W}_{in,a}$ *and* $\mathbf{n}_{0,a}$ *are known,*

(ii) $\mathrm{rank}(\mathbf{N}_a) = R$, *and*

(iii) $\mathbf{n}_a(t_h)$, $\mathbf{u}_{in}(t)$ *and* $u_{out}(t)$ *are available from measurements.*

Then, the extents of reaction $x_{r,i}(t_h)$, $\forall i = 1, \ldots, R$, *can be computed in two steps as follows:*

1. Compute the extents $\mathbf{x}_{in}(t)$ and $\lambda(t)$:

$$\dot{\mathbf{x}}_{in}(t) = \mathbf{u}_{in}(t) - \frac{u_{out}(t)}{m(t)}\mathbf{x}_{in}(t), \quad \mathbf{x}_{in}(0) = \mathbf{0}_p,$$

$$\dot{\lambda}(t) = -\frac{u_{out}(t)}{m(t)}\lambda(t), \qquad\qquad \lambda(0) = 1,$$

$$(14)$$

with

$$m(t) = \mathbf{1}_p^{\mathrm{T}}\mathbf{x}_{in}(t) + m_0\,\lambda(t), \tag{15}$$

where λ is a scalar dimensionless variable used to account for the effect of the outlet on the initial conditions.

2. Compute the extents of reaction $x_{r,i}(t_h)$, $i = 1,\ldots,R$:

$$x_{r,i}(t_h) = (\mathbf{N}_a^{\mathrm{T}+})_i\left(\mathbf{n}_a(t_h) - \mathbf{W}_{in,a}\mathbf{x}_{in}(t_h) - \mathbf{n}_{0,a}\,\lambda(t_h)\right), \tag{16}$$

where $(\mathbf{N}_a^{\mathrm{T}+})_i$ denotes the ith row of the $(R \times S_a)$-dimensional $\mathbf{N}_a^{\mathrm{T}+}$ matrix. (Proof see Appendix B)

Note that $\lambda(t) = 1$ for reactors without outlet stream. The numbers of moles of the $S_u = S - S_a$ unmeasured species $\mathbf{n}_u(t_h)$ can be reconstructed using the computed extents of reaction and flow. If the initial conditions of the species that are not measured on-line, $\mathbf{n}_{0,u}$, are known, the corresponding numbers of moles can be reconstructed using the computed $\mathbf{x}_r(t_h)$ in Eq. (16) and $\mathbf{x}_{in}(t)$ and $\lambda(t)$ in Eq. (14) as follows:

$$\mathbf{n}_u(t_h) = \mathbf{N}_u^{\mathrm{T}}\mathbf{x}_r(t_h) + \mathbf{W}_{in,u}\mathbf{x}_{in}(t_h) + \mathbf{n}_{0,u}\,\lambda(t_h), \tag{17}$$

where $\mathbf{N}_u$ is the known $(R \times S_u)$-dimensional stoichiometric matrix and $\mathbf{W}_{in,u}$ the known $(S_u \times p)$-dimensional inlet-composition matrix.

Remarks

1. If Assumptions (i)–(iii) in Proposition 1 or 2 are satisfied, then the extents of reaction can be computed from the measured experimental data. In the other words, the measured data is sufficiently informative in the sense that one can compute the reaction extents. However, fulfillment of Assumptions (i)–(iii) does not provide any indication about parameter identifiability and the quality of the parameters computed from measured data.

2. Note that, for situations where Assumption (ii) in Proposition 1 does not hold, e.g. when $S < R + p + 1$, the extents of reactions can be computed as per Proposition 2.

Error in extents of reaction computed from noisy composition measurements

In practice, the numbers of moles are obtained from noisy concentration and volume measurements. Noise propagates to the computed extents as illustrated next. Let $\check{\mathbf{n}} = \mathbf{n} + \mathbf{e}_n$ denote the noisy value of $\mathbf{n}$, where $\mathbf{e}_n$ is a vector of zero-mean Gaussian noise with variance $\Sigma_{S \times S}$, that is, $\mathbf{e}_n \sim \mathcal{N}(\mathbf{0}_S, \Sigma_{S \times S})$. Since the errors in flowrate measurements are usually negligible with respect to errors in concentration measurements, it is assumed that the flowrates are noise free. Furthermore, for the sake of simplicity, we also assume that the initial conditions $\mathbf{n}_0$ are noise free. Since the extents are linear functions of the numbers of moles, unbiased extent estimates can be computed from the numbers of moles as given next.

The mean and the variance of the extent of the ith reaction computed via linear transformation are:

$$
\begin{aligned}
\mathrm{E}[x_{r,i}(t_h)] &= (\mathbf{S}_0^{\mathrm{T}})_i \, \mathbf{n}(t_h), \\
\mathrm{var}[x_{r,i}(t_h)] &= (\mathbf{S}_0^{\mathrm{T}})_i \, \Sigma_{S \times S} (\mathbf{S}_0^{\mathrm{T}})_i^{\mathrm{T}},
\end{aligned}
\tag{18}
$$

where $\mathrm{E}[\cdot]$ is the expectation operator and $\mathrm{var}[\cdot]$ is the variance operator.

Similarly, the mean and the variance of the extent of the ith reaction computed via the flow-

14

dependent approach read:

$$E[x_{r,i}(t_h)] = (\mathbf{N}_a^{T+})_i \left(\mathbf{n}_a(t_h) - \mathbf{W}_{in,a}\mathbf{x}_{in}(t_h) - \mathbf{n}_{0,a}\,\lambda(t_h) \right),$$

$$\text{var}[x_{r,i}(t_h)] = (\mathbf{N}_a^{T+})_i\, \Sigma_{S_a \times S_a} (\mathbf{N}_a^{T+})_i^{T}. \tag{19}$$

Eqs. (18) and (19) show that the computed extent $x_{r,i}(t_h)$ is unbiased and its variance depends on the variances of the measured concentrations.

Gas-liquid reaction systems

The objective of this section is to compute the extents of reaction and mass transfer from concentration measurements. Note that the extents of mass transfer can be computed from either the gas or the liquid concentrations. On the other hand, the extents of reaction can only be computed from the liquid concentrations. Again, two cases are distinguished depending upon the number of species for which concentration measurements are available.

All liquid-phase concentrations measured: Linear transformation

Let $\mathbf{n}_l(t_h)$ be the numbers of moles of the S_l species in the liquid phase measured at time instant t_h. The following proposition states the conditions to be able to compute the extents of reaction $\mathbf{x}_r(t_h)$ and mass transfer $\mathbf{x}_{m,l}(t_h)$ from $\mathbf{n}_l(t_h)$ using the linear transformation proposed in Bhatt *et al.*[22]

Proposition 3 (Linear transformation of $\mathbf{n}_l(t_h)$)

Consider the following assumptions:

  (i) $\mathbf{N}$, $\mathbf{W}_{m,l}$, $\mathbf{W}_{in,l}$ *and* $\mathbf{n}_{l0}$ *are known,*

  (ii) $rank([\mathbf{N}^{T}\ \ \mathbf{W}_{m,l}\ \ \mathbf{W}_{in,l}\ \ \mathbf{n}_{l0}]) = R + p_m + p_l + 1$, *and*

  (iii) $\mathbf{n}_l(t_h)$ *is available from measurements.*

Then, the extent of the ith reaction $x_{r,i}(t_h)$, $\forall i = 1,\ldots,R$, and the extent of the jth mass transfer in the liquid phase $x_{m,l,j}(t_h)$, $\forall j = 1,\ldots,p_m$, can be computed as:

$$x_{r,i}(t_h) = (\mathbf{S}_{l0}^{\mathrm{T}})_i \, \mathbf{n}_l(t_h), \tag{20}$$

$$x_{m,l,j}(t_h) = (\mathbf{M}_{l0}^{\mathrm{T}})_j \, \mathbf{c}_l(t_h), \tag{21}$$

where $(\mathbf{S}_{l0}^{\mathrm{T}})_i$ denotes the ith row of the $(R \times S_l)$-dimensional $\mathbf{S}_{l0}^{\mathrm{T}}$ matrix and $(\mathbf{M}_{l0}^{\mathrm{T}})_j$ the jth row of the $(p_l \times S_l)$-dimensional $\mathbf{M}_{l0}^{\mathrm{T}}$ matrix, which can be computed using the algorithm given in Appendix A. (Proof follows from Theorem 1 in Bhatt et al.[22])

Similarly, if the measurements of all the species in the gas phase $\mathbf{n}_g(t_h)$ are available, the linear transformation can be applied to compute the extent of the jth mass transfer $x_{m,g,j}(t_h)$ in the gas phase.[22] Note that the extents $x_{m,l,j}(t_h)$ and $x_{m,g,j}(t_h)$ are typically different as they represent the mass transferred between phases, which is discounted for the outlet flow of liquid and gas, respectively.

A subset of concentrations measured in the gas and liquid phases: Flow-dependent approach

Let $\mathbf{n}_{g,a}(t_h)$ and $\mathbf{n}_{l,a}(t_h)$ be the $S_{g,a}$- and $S_{l,a}$-dimensional vectors of available numbers of moles in the gas and liquid phases at time instant t_h, respectively. In addition, the inlet flowrates $\mathbf{u}_{in,l}(t)$ and $\mathbf{u}_{in,g}(t)$, the outlet flowrates $u_{out,l}(t)$ and $u_{out,g}(t)$, and the masses $m_l(t)$ and $m_g(t)$ are also measured.

The extents of mass transfer will be computed from information stemming from both phases, namely p_{m_g} extents will be computed from the gas phase and p_{m_l} extents from the liquid phase, with $p_{m_l} + p_{m_g} = p_m$. Without loss of generality, it is assumed that the $S_{g,a}$ species measured in the gas phase are involved in mass transfer.[2] In other words, the p_m transferring species are measured such that $S_{g,a} = p_{m_g}$ species are measured in the gas phase and p_{m_l} are measured in the liquid phase.

---

[2]If some of the $S_{g,a}$ species measured in the gas phase are not involved in mass transfer, these species will simply be discarded since they are not useful in the computation of the extents of mass transfer.

The extents of mass transfer and the mass-transfer matrices are written accordingly. For instance, $\mathbf{x}_{m_g,g}$ and $\mathbf{x}_{m_g,l}$ represent p_{m_g}-dimensional vectors of extents of mass transfer computed from gas-phase measurements for the gas phase and the liquid phase, respectively; $\mathbf{x}_{m_l,l}$ is the p_{m_l}-dimensional vector of extents of mass transfer computed from and for the liquid phase; $\mathbf{W}_{m_g,g,a}$ is the $(S_{g,a} \times p_{m_g})$-dimensional mass-transfer matrix associated with the p_{m_g} mass transfers and the $S_{g,a}$ species; $\mathbf{W}_{m_l,l,a}$ is the $(S_{l,a} \times p_{m_l})$-dimensional mass-transfer matrix associated with the p_{m_l} mass transfers and the $S_{l,a}$ species; $\mathbf{W}_{m_g,l,a}$ is the $(S_{l,a} \times p_{m_g})$-dimensional mass-transfer matrix associated with the p_{m_g} mass transfers and the $S_{l,a}$ species. Furthermore, let $\bar{\mathbf{N}}_a^{\mathrm{T}} := [\mathbf{N}_a^{\mathrm{T}} \ \mathbf{W}_{m_l,l,a}]$ be the extended $[S_{l,a} \times (R + p_{m_l})]$-dimensional stoichiometric matrix.

Proposition 4 (Flow-dependent approach: Use of $\mathbf{n}_{f,a}(t_h)$, $\mathbf{u}_{in,f}(t)$ and $u_{out,f}(t)$, $f \in \{g,l\}$)

Consider the following assumptions:

(i) $\mathbf{N}_a$, $\mathbf{W}_{m_g,g,a}$, $\mathbf{W}_{m_l,l,a}$, $\mathbf{W}_{m_g,l,a}$, $\mathbf{W}_{in,l,a}$, $\mathbf{W}_{in,g,a}$ *and the initial conditions* $\mathbf{n}_{l0,a}$ *and* $\mathbf{n}_{g0,a}$ *are known,*

(ii) $S_{l,a} + S_{g,a} \geq R + p_m,$

(iii) $\mathrm{rank}\,(\mathbf{N}_a) = R,$ *and*

(iv) $\mathrm{rank}\,(\bar{\mathbf{N}}_a) = R + p_{m_l},$

(v) $\mathbf{n}_{g,a}(t_h)$, $\mathbf{n}_{l,a}(t_h)$, $\mathbf{u}_{in,l}(t)$, $\mathbf{u}_{in,g}(t)$, $u_{out,l}(t)$, $u_{out,g}(t)$, $V_l(t)$, $m_l(t)$ *and* $m_g(t)$ *are available from measurements.*

Then, the extents of reaction and mass transfer can be computed in three steps as follows:

1. *Compute the extents of mass transfer* $\mathbf{x}_{m_g,g}(t_h)$ *in the gas phase from:*

$$
\begin{aligned}
\dot{\mathbf{x}}_{in,g}(t) &= \mathbf{u}_{in,g}(t) - \frac{u_{out,g}(t)}{m_g(t)}\mathbf{x}_{in,g}(t), & \mathbf{x}_{in,g}(0) &= \mathbf{0}_{p_g}, \\
\dot{\lambda}_g(t) &= -\frac{u_{out,g}(t)}{m_g(t)}\lambda_g(t), & \lambda_g(0) &= 1, \\
\mathbf{W}_{m_g,g,a}\,\mathbf{x}_{m_g,g}(t_h) &= \mathbf{n}_{g,a}(t_h) - \mathbf{W}_{in,g,a}\mathbf{x}_{in,g}(t_h) - \mathbf{n}_{g0,a}\,\lambda_g(t_h).
\end{aligned}
\tag{22}
$$

2. Compute the extents of mass transfer $\mathbf{x}_{m_g,l}(t_h)$ in the liquid phase:

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

$$\mathbf{x}_{m_g,l}(t_h) = \mathbf{x}_{m_g,g}(t_h) - \delta_m(t_h), \tag{23}$$

where $\delta_m(t)$ is a p_{m_g}-dimensional vector expressing the difference in the extents of mass transfer computed for the gas and liquid phases.

3. Compute the extents of reaction $\mathbf{x}_r(t_h)$ and mass transfer $\mathbf{x}_{m_l,l}(t_h)$ in the liquid phase:

$$\dot{\mathbf{x}}_{in,l}(t) = \mathbf{u}_{in,l}(t) - \frac{u_{out,l}(t)}{m_l(t)}\mathbf{x}_{in,l}(t), \quad \mathbf{x}_{in,l}(0) = \mathbf{0}_{p_l},$$

$$\dot{\lambda}_l(t) = -\frac{u_{out,l}(t)}{m_l(t)}\lambda_l(t), \qquad \lambda_l(0) = 1, \tag{24}$$

$$\bar{\mathbf{N}}_a^{\mathrm{T}}\begin{bmatrix}\mathbf{x}_r(t_h)\\[1mm]\mathbf{x}_{m_l,l}(t_h)\end{bmatrix} = \mathbf{n}_{l,a}(t_h) - \mathbf{W}_{in,l,a}\mathbf{x}_{in,l}(t_h) - \mathbf{W}_{m_g,l,a}\mathbf{x}_{m_g,l}(t_h) - \mathbf{n}_{l0,a}\lambda_l(t_h).$$

(See Proof in Appendix C)

Remarks

1. The differential-algebraic equation system (22)-(24) can be solved as follows:

   - *Step 1:* Since the matrix $\mathbf{W}_{m_g,g,a}$ is full rank by construction, the differential-algebraic system (22) can be solved to compute the extents of mass transfer $\mathbf{x}_{m_g,g}(t_h)$ involving the $S_{g,a}$ species measured in the gas phase.

   - *Step 2:* The corresponding extents of mass transfer for the liquid phase, $\mathbf{x}_{m_g,l}(t_h)$, can be estimated from Eq. (23).

   - *Step 3:* Since the matrix $\bar{\mathbf{N}}_a$ is full rank by Assumption (iv), the R extents of reaction $\mathbf{x}_r(t_h)$ and the remaining p_{m_l} extents of mass transfer $\mathbf{x}_{m_l,l}(t_h)$ can be estimated by solving Eq. (24).

2. Assumptions (i)–(v) specify the minimal number of concentration measurements and the conditions that are needed to compute the extents of reaction and mass transfer. In other words, the extents of reaction and mass transfer can be computed from subsets of the species measurements, if Assumptions (i)–(v) are fulfilled.

3. As with homogeneous reaction systems, since the computed extents are linear function of the numbers of moles in the liquid and gas phases, they are unbiased in G–L reaction systems when the measured numbers of moles are corrupted with zero-mean Gaussian noise.

4. If the volumetric flowrates $\mathbf{q}_{in,f}$ and $q_{out,f}$, $f \in \{g, l\}$, and the liquid volume V_l are measured, then $\frac{u_{out,f}}{m_f}$ can be replaced by $\frac{q_{out,f}}{V_f}$ in Eqs. (22)-(24). Moreover, the masses of the two phases need not be measured.

5. If the initial conditions of the species that are not measured on-line in the liquid, $\mathbf{n}_{l0,u}$, and in the gas, $\mathbf{n}_{g0,u}$, are known, the corresponding numbers of moles $\mathbf{n}_{l,u}(t_h)$ and $\mathbf{n}_{g,u}(t_h)$ can be computed from $\mathbf{x}_r(t_h), \mathbf{x}_{in,l}(t_h), \mathbf{x}_{in,g}(t_h), \mathbf{x}_{m,l}(t_h), \mathbf{x}_{m,g}(t_h), \mathbf{x}_{m,l}(t_h), \lambda_l(t_h)$ and $\lambda_g(t_h)$ as follows:

$$\mathbf{n}_{l,u}(t_h) = \mathbf{N}_u^{\mathrm{T}} \mathbf{x}_r(t_h) + \mathbf{W}_{in,l,u} \mathbf{x}_{in,l}(t_h) + \mathbf{W}_{m,l,u} \mathbf{x}_{m,l}(t_h) + \mathbf{n}_{l0,u} \lambda_l(t_h), \tag{25}$$

$$\mathbf{n}_{g,u}(t_h) = \mathbf{W}_{in,g,u} \mathbf{x}_{in,g}(t_h) - \mathbf{W}_{m,g,u} \mathbf{x}_{m,g}(t_h) + \mathbf{n}_{g0,u} \lambda_g(t_h). \tag{26}$$

Special case: G–L reactors without outlet

The computation of extents of reaction and mass transfer for G–L reactors without outlet, such as batch and semi-batch reactors, is discussed in this section. In such a case, since $x_{m,l,j}(t) = x_{m,g,j}(t)$ and thus $\delta_m(t) = \mathbf{0}$, the gas and liquid phases can be treated simultaneously. When the concentrations of all species in the liquid phase are measured, Proposition 3 can be used to compute the extents of reaction and mass transfer from measurements.

Next, the case when a subset of the concentrations are measured in the gas and liquid phases is considered. Let $\mathbf{n}_a(t_h) = \begin{bmatrix} \mathbf{n}_{g,a}(t_h) \\ \mathbf{n}_{l,a}(t_h) \end{bmatrix}$ be a S_a-dimensional vector of measured numbers of moles

in the gas and liquid phases at time instant t_h, with $S_a = S_{g,a} + S_{l,a}$. The $[S_a \times (p_g + p_l + p_m)]$-dimensional extended inlet matrix of available species is defined as $\bar{\mathbf{W}}_{in,a} := \begin{bmatrix} \mathbf{W}_{in,g,a} & \mathbf{W}_{in,l,a} & \mathbf{W}_{m,a} \end{bmatrix}$, where $\mathbf{W}_{in,g,a}$ is the $(S_a \times p_g)$-dimensional matrix associated with the p_g gas inlets, $\mathbf{W}_{in,l,a}$ is the $(S_a \times p_l)$-dimensional matrix associated with the p_l liquid inlets, and $\mathbf{W}_{m,a}$ is the $(S_a \times p_m)$-dimensional matrix associated with the p_m mass transfers. Also, let $\bar{\mathbf{N}}_a^{\mathrm{T}} := [\mathbf{N}_a^{\mathrm{T}} \ \mathbf{W}_{m,a}]$ be the $[S_a \times (R + p_m)]$-dimensional extended stoichiometric matrix in this case.

Proposition 5 (Flow-dependent approach for G–L reactors without outlet)

Consider the following assumptions:

(i) $\bar{\mathbf{N}}_a$ and $\bar{\mathbf{W}}_{in,a}$ are known,

(ii) rank $(\bar{\mathbf{N}}_a) = R + p_m$, and

(iii) $\mathbf{n}_a(t_h)$, $\mathbf{u}_{in,g}(t)$, and $\mathbf{u}_{in,l}(t)$ are available from measurements.

Then, the extents of reaction $x_{r,i}(t_h)$, $\forall i = 1, \ldots, R$, and of mass transfer $x_{m,j}$, $\forall j = 1, \ldots, p_m$, can be computed in two steps as follows:

1. Compute the extents of inlet flow $\mathbf{x}_{in,l}(t)$ and $\mathbf{x}_{in,g}(t)$:

$$\dot{\mathbf{x}}_{in,l}(t) = \mathbf{u}_{in,l}(t), \qquad \mathbf{x}_{in,l}(0) = \mathbf{0}_{p_l}, \tag{27}$$

$$\dot{\mathbf{x}}_{in,g}(t) = \mathbf{u}_{in,g}(t), \qquad \mathbf{x}_{in,g}(0) = \mathbf{0}_{p_g}. \tag{28}$$

2. Compute the extents of the ith reaction and jth mass transfer:

$$
\begin{aligned}
x_{r,i}(t_h) &= (\bar{\mathbf{N}}_a^{\mathrm{T}+})_i \left(\mathbf{n}_a(t_h) - \mathbf{W}_{in,g,a}\mathbf{x}_{in,g}(t_h) - \mathbf{W}_{in,l,a}\mathbf{x}_{in,l}(t_h) - \mathbf{n}_{0,a} \right), \\
x_{m,j}(t_h) &= (\bar{\mathbf{N}}_a^{\mathrm{T}+})_{R+j} \left(\mathbf{n}_a(t_h) - \mathbf{W}_{in,g,a}\mathbf{x}_{in,g}(t_h) - \mathbf{W}_{in,l,a}\mathbf{x}_{in,l}(t_h) - \mathbf{n}_{0,a} \right),
\end{aligned}
\tag{29}
$$

where $(\bar{\mathbf{N}}_a^{\mathrm{T}+})_i$ and $(\bar{\mathbf{N}}_a^{\mathrm{T}+})_{R+j}$ are the ith and $(R+j)$th rows of the $[(R+p_m) \times S_a]$-dimensional matrix $(\bar{\mathbf{N}}_a^{\mathrm{T}+})$.

(See Proof in Appendix D)

20

The $S_{l,u}$ unavailable numbers of moles in the liquid phase $\mathbf{n}_{l,u}(t_h)$ and the $S_{g,u}$ unavailable numbers of moles in the gas phase $\mathbf{n}_{g,u}(t_h)$ can be reconstructed from the computed extents as:

$$\mathbf{n}_{l,u}(t_h) = \mathbf{N}_u^T \mathbf{x}_r(t_h) + \mathbf{W}_{in,l,u}\mathbf{x}_{in,l}(t_h) + \mathbf{W}_{m,u}\mathbf{x}_{m,l}(t_h) + \mathbf{n}_{l0,u}\lambda_l(t_h), \tag{30}$$

$$\mathbf{n}_{g,u}(t_h) = \mathbf{W}_{in,g,u}\mathbf{x}_{in,g}(t_h) - \mathbf{W}_{m,u}\mathbf{x}_m(t_h) + \mathbf{n}_{g0,u}\lambda_g(t_h). \tag{31}$$

Estimation of rate parameters from computed extents

With the integral method, the rate parameters are estimated by comparing measured concentrations and concentrations that are obtained by integration of the reactor model using postulated rate expressions. In this section, the integral method is used to compare extents that are computed from measured concentrations and extents that are obtained by integration of postulated rate expressions. This can be done *individually* for each extent of reaction and mass transfer.

For the ith reaction, let $\mathbf{x}_{r,i}$ and $\hat{\mathbf{x}}_{r,i}$ denote the H-dimensional vectors of computed (according to the procedures described in the previous section) and simulated (according to a rate expression involving the parameters $\theta_{r,i}$) extents of reaction at H time instants, respectively. The following weighted least-squares problem can be formulated to estimate the parameters $\theta_{r,i}$:

$$\min_{\theta_{r,i}} (\mathbf{x}_{r,i} - \hat{\mathbf{x}}_{r,i}(\theta_{r,i}))^T \mathbf{W}_r (\mathbf{x}_{r,i} - \hat{\mathbf{x}}_{r,i}(\theta_{r,i}))$$

$$\text{s.t.} \quad \dot{\hat{x}}_{r,i}(t) = V_l(t)\, r_i(\mathbf{c}_l(t), \theta_{r,i}) - \frac{u_{out,l}(t)}{m_l(t)}\hat{x}_{r,i}(t), \quad \hat{x}_{r,i}(0) = 0, \tag{32}$$

$$\theta_{r,i}^L \leq \theta_{r,i} \leq \theta_{r,i}^U,$$

where $\mathbf{W}_r$ is an $(H \times H)$-dimensional weighting matrix, r_i is the postulated rate expression for the ith reaction, which is a known function of the molar concentrations $\mathbf{c}_l(t)$ and the l-dimensional unknown parameter vector $\theta_{r,i}$ that can vary between the lower bound $\theta_{r,i}^L$ and the upper bound $\theta_{r,i}^U$. The dynamic equation for $\hat{x}_{r,i}$ in Eq. (32) results from the linear transformation in Bhatt et al.[22] Note that $\mathbf{c}_l(t)$ needs to be reconstructed from the H measured values $\mathbf{c}_l(t_h)$.

Similarly, let $\mathbf{x}_{m,l,j}$ and $\hat{\mathbf{x}}_{m,l,j}$ denote the H-dimensional vectors of computed and simulated extents of the jth mass transfer in the liquid phase, respectively. The following weighted least-squares problem can be formulated to estimate the parameters $\theta_{m,j}$:

$$\min_{\theta_{m,j}} (\mathbf{x}_{m,l,j} - \hat{\mathbf{x}}_{m,l,j}(\theta_{m,j}))^{\mathrm{T}} \mathbf{W}_m (\mathbf{x}_{m,l,j} - \hat{\mathbf{x}}_{m,l,j}(\theta_{m,j}))$$

$$\text{s.t.} \quad \dot{\hat{x}}_{m,l,j}(t) = \zeta_j(\mathbf{c}_l(t), \mathbf{c}_g(t), \theta_{m,j}) - \frac{u_{out,l}(t)}{m_l(t)} \hat{x}_{m,l,j}, \quad \hat{x}_{m,l,j}(0) = 0, \tag{33}$$

$$\theta_{m,j}^L \leq \theta_{m,j} \leq \theta_{m,j}^U,$$

where $\mathbf{W}_m$ is an $(H \times H)$-dimensional weighting matrix, and ζ_j is the postulated rate expression for the jth mass transfer, which is a known function of the measured concentrations $\mathbf{c}_l(t)$ and $\mathbf{c}_g(t)$ and the q-dimensional unknown parameter vector $\theta_{m,j}$ that can vary between the lower bound bound $\theta_{m,j}^L$ and the upper bound $\theta_{m,j}^U$.

Eq. (32) can be used to estimate the reaction rate parameters in homogeneous and G–L reaction systems for $i = 1, \ldots, R$. Eq. (33) can be used to estimate the mass-transfer rate parameters in G–L reaction systems for $j = 1, \ldots, p_m$. When only a subset of the concentrations is measured, the unavailable numbers of moles can be reconstructed from the available measurements using Eq. (17) for homogeneous reactors and Eqs. (25) and (26) for G–L reaction systems.

Integration of the differential equations (32) and (33) requires the knowledge of the concentrations $\mathbf{c}_l(t)$ and $\mathbf{c}_g(t)$, which can be achieved by polynomial interpolation of the measured values $\mathbf{c}_l(t_h)$ and $\mathbf{c}_g(t_h)$, $h = 0, 1, \ldots, H$. For example, $\mathbf{c}(t) = \sum_{h=0}^{H} \mathbf{c}(t_h) \phi_h(t)$, where $\mathbf{c}(t)$ represents a continuous function approximation of the concentrations based on the low-resolution data $\mathbf{c}(t_h)$ and the basis functions $\phi_h(t)$. In this paper, "ode45" of MATLAB$^{\circledR}$, which is based on an explicit Runge-Kutta (4,5) formula, is used to integrate the differential equations.[23]

Remarks

1. **Parameter identifiability:** Identifiability of parameters implies uniqueness of the rate parameters determined from experimental data.[1] Several methods can be used to check pa-

rameter identifiability a priori.[24–27] In contrast, a parametric sensitivity analysis can be performed after parameter estimation to identify the parameters that affect the fit the most.[28] Moreover, the quality of parameter estimates can be evaluated using statistical methods. For example, the confidence intervals of the estimated parameters can be calculated using significance testing or bootstrapping.[29] In this paper, the function "nlparci" in the statistical toolbox of MATLAB®, which is based on "asymptotic normal distribution of the parameter estimates", is used to compute the confidence intervals.[30]

2. **Error propagation from the computed extents to the estimated rate parameters:** Since noisy measurements are used to compute $\mathbf{c}_l(t)$ and $\mathbf{c}_g(t)$ and simulate the profiles $\hat{x}_{r,i}(t)$ and $\hat{x}_{m,l,j}(t)$, error propagation can affect the accuracy of the estimated parameters. A numerical investigation of the error propagation from the computed extents to the estimated parameters has been carried out.[31] The main result is that the extent-based (integral) incremental approach gives parameter estimates with tighter confidence intervals in comparison to the rate-based (differential) incremental approach. The numerical investigation also indicates that both incremental approaches introduce bias in parameter estimates, particularly in the presence of large measurement noise or sparse data.
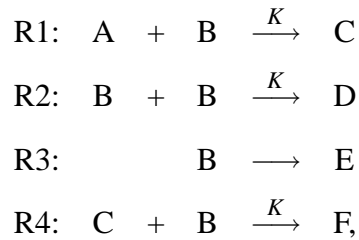
Illustrative simulated examples

This section illustrates the computation of extents of reaction and mass transfer from measured concentrations and the estimation of reaction and mass-transfer rate parameters. The acetoacetylation of pyrrole illustrates a homogeneous reaction system, while the chlorination of butanoic acid is representative of a G–L reaction system.

Homogeneous reaction system: Acetoacetylation of pyrrole

Generation of simulated data: The acetoacetylation of pyrrole (A) with diketene (B) involves one main reaction and three side reactions.[32] The main reaction (R1) between pyrrole and diketene

produces 2-acetoacetyl pyrrole (C). The side reactions include (R2), the dimerization of diketene to dehydroacetic acid (D); (R3), the oligomerization of diketene to oligomers (E); and (R4), a consecutive reaction between diketene and 2-acetoacetyl pyrrole to the by-product F. The reactions R1, R2 and R4 are catalyzed by pyridine (K). The reaction stoichiometry reads:

$$
\begin{array}{lllclcl}
\text{R1:} & A & + & B & \xrightarrow{K} & C \\
\text{R2:} & B & + & B & \xrightarrow{K} & D \\
\text{R3:} & & & B & \longrightarrow & E \\
\text{R4:} & C & + & B & \xrightarrow{K} & F,
\end{array}
$$

from which one can write the stoichiometric matrix $\mathbf{N}$:

$$
\mathbf{N} = \begin{bmatrix}
-1 & -1 & 1 & 0 & 0 & 0 \\
0 & -2 & 0 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 \\
0 & -1 & -1 & 0 & 0 & 1
\end{bmatrix}. \tag{34}
$$

The kinetic expressions are as follows:

$$
\begin{aligned}
r_1 &= k_1 c_A c_B c_K \tag{35} \\
r_2 &= k_2 c_B^2 c_K \\
r_3 &= k_3 c_B \\
r_4 &= k_4 c_C c_B c_K,
\end{aligned}
$$

with the "true" parameter values used for data generation given in Table 2.[32]

The startup of an isothermal CSTR is considered. The initial concentrations of the 6 species $A - F$ are $\mathbf{c}_0 = \begin{bmatrix} 0.3 & 1 & 0.1 & 0.01 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$ mol L^{-1}. The species A and B are added continuously through one inlet with composition $\mathbf{c}_{in} = \begin{bmatrix} 2 & 3.5 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$ mol L^{-1}. The volumetric inlet flowrate is $q_{in} = 0.1$ L min^{-1}. Under the assumption of constant density, the volumetric outlet flowrate is $q_{out} = 0.1$ L min^{-1}. The volume of the reactor is 1 L. There is initially 0.5 mol of catalyst in the reactor. The concentrations of all species are measured every 30 sec for 1 h. The

measurements are corrupted with additive zero-mean Gaussian noise. The standard deviation for each species is taken as 5% of the maximal concentration of that species, i.e. $\sigma_s = 0.05 \, c_s^{max}$ for $s = \{A, B, C, D, E, F\}$. The noisy concentration measurements are shown in Figure 2.
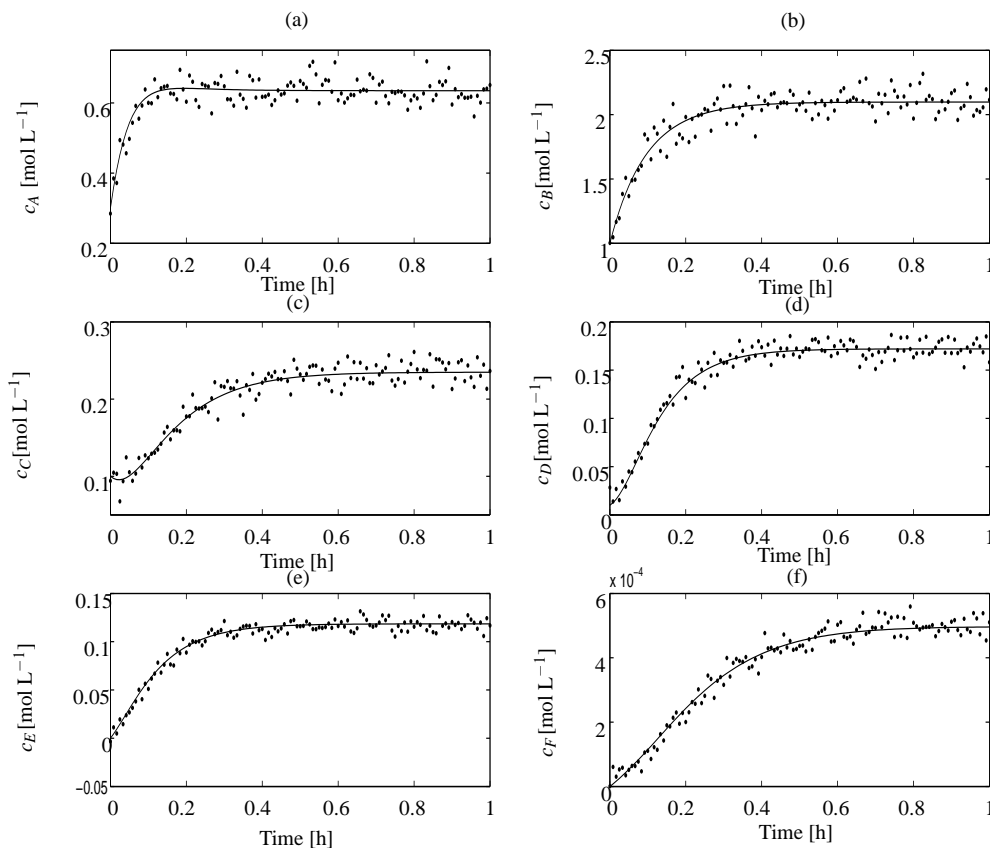

Figure 2: : Concentrations of the 6 species in the acetoacetylation of pyrrole. The solid lines indicate the generated (true) concentrations, while the markers indicate the noisy measurements.

Identification of reaction kinetics: Since the concentrations of all species are measured and rank $([\mathbf{N}^T \ \mathbf{c}_{in} \ \mathbf{n}_0]) = R + p + 1 = 4 + 1 + 1 = 6$, Proposition 1 can be used to compute the extents of reaction from concentration data. The computed extents, shown in Figure 3, are within the 95% confidence intervals that are calculated using the mean (the simulated "true" extents) and the variance computed from Eq. (18) using the noise distribution used in the data generation.

The next step is to identify the rate expressions from among the set of rate expression candidates
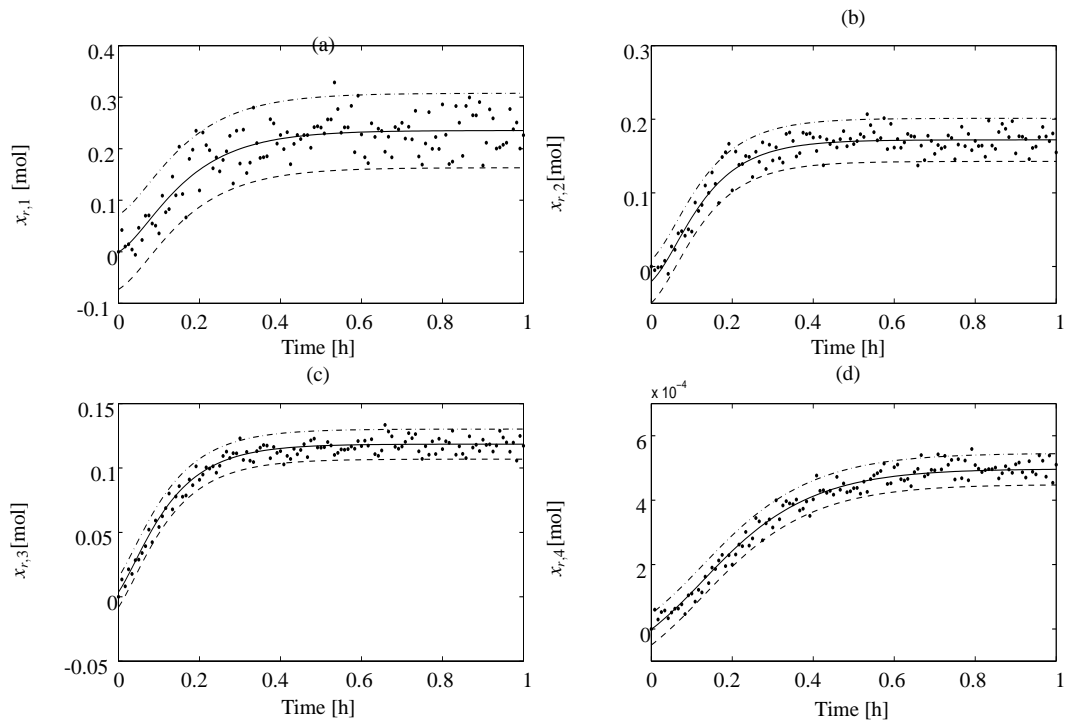
Figure 3: Extents of (a) reaction R1, (b) reaction R2, (c) reaction R3, and (d) reaction R4 computed from measured concentrations. The solid lines indicate the true extents, while the markers indicate the extents computed from noisy concentration measurements. The dash lines indicate the 95% confidence intervals.

26

Table 1: Rate expression candidates for the acetoacetylation of pyrrole. Candidates $r_1^{(8)}$, $r_2^{(5)}$, $r_3^{(2)}$ and $r_4^{(6)}$ correspond to the simulated (true) rate expressions

Reaction R1:	Reaction R2:	Reaction R3:	Reaction R4:
$r_1^{(1)} = k_1$	$r_2^{(1)} = k_2$	$r_3^{(1)} = k_3$	$r_4^{(1)} = k_4$
$r_1^{(2)} = k_1 c_B$	$r_2^{(2)} = k_2 c_B$	$r_3^{(2)} = k_3 c_B$	$r_4^{(2)} = k_4 c_B$
$r_1^{(3)} = k_1 c_A$	$r_2^{(3)} = k_2 c_B^2$	$r_3^{(3)} = k_3 c_B^2$	$r_4^{(3)} = k_4 c_C$
$r_1^{(4)} = k_1 c_K$	$r_2^{(4)} = k_2 c_B c_K$	$r_3^{(4)} = k_3 c_B c_K$	$r_4^{(4)} = k_4 c_K$
$r_1^{(5)} = k_1 c_A c_B$	$r_2^{(5)} = k_2 c_B^2 c_K$	$r_3^{(5)} = k_3 c_B^2 c_K$	$r_4^{(5)} = k_4 c_B c_C$
$r_1^{(6)} = k_1 c_A c_K$	$r_2^{(6)} = k_2 c_K$	$r_3^{(6)} = k_3 c_K$	$r_4^{(6)} = k_4 c_K c_B c_C$
$r_1^{(7)} = k_1 c_B c_K$			$r_4^{(7)} = k_4 c_K c_B$
$r_1^{(8)} = k_1 c_A c_B c_K$			$r_4^{(8)} = k_4 c_K c_C$
$r_1^{(9)} = k_1 c_A^2 c_B$			$r_4^{(9)} = k_4 c_K c_B c_C^2$
$r_1^{(10)} = k_1 c_A c_B^2$			$r_4^{(10)} = k_4 c_K c_B^2 c_C$

given for each reaction in Table 1. Note that each rate expression involves a single unknown parameter and thus $\theta_{r,1} = k_1$, $\theta_{r,2} = k_2$, $\theta_{r,3} = k_3$, and $\theta_{r,4} = k_4$. For a given reaction, each rate expression candidate is fitted to the corresponding computed extent using the least-squares problem (32). For example, for the main reaction R1, the quality of fit of rate expression candidates $r_1^{(4)}$, $r_1^{(6)}$ and $r_1^{(8)}$ is shown in Figure 4. Candidate $r_1^{(8)}$ leads to the best fit for $\mathbf{W}_r = \mathbf{I}_H$. Similarly, the rate expressions $r_2^{(5)}$, $r_3^{(2)}$ and $r_4^{(6)}$ are identified as the suitable rate expressions for reactions R2, R3, and R4, which are indeed the rate expressions used for generating the data. The true values, the initial guesses, the estimated values and their 95% confidence intervals are given in Table 2. The estimated values are close to the true values.

G–L reaction system: Chlorination of butanoic acid

The chlorination of butanoic acid (BA), which takes place in the organic liquid phase with ethanol as solvent, involves two parallel auto-catalytic reactions that consume dissolved Cl_2. The main
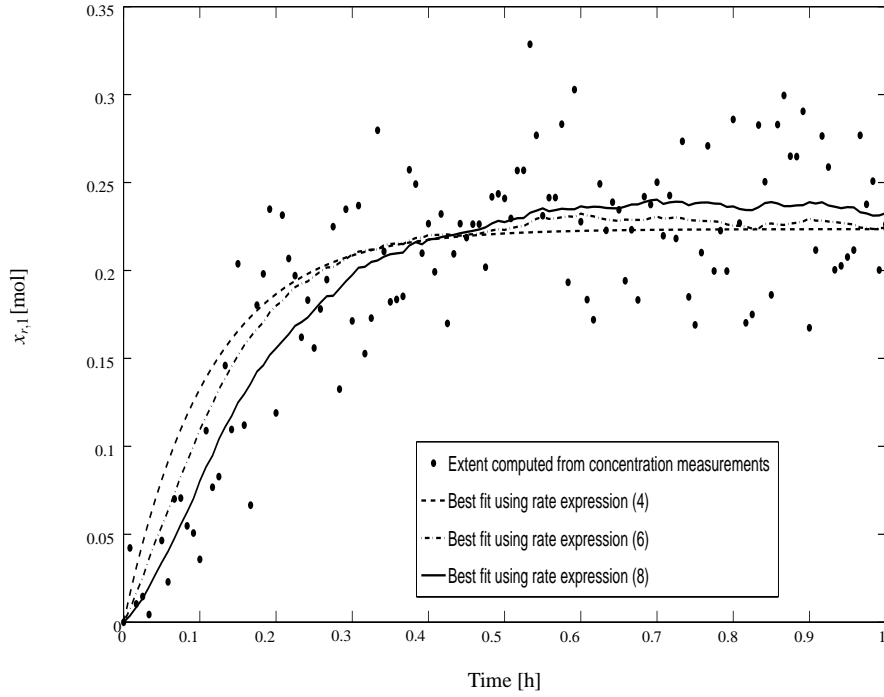
Figure 4: Fit of three rate expressions to the computed extent $x_{r,1}$.

Table 2: Parameter estimation using the extent-based incremental identification approach: True, initial and estimated values of the four rate parameters for the selected rate expressions $r_1^{(8)}$, $r_2^{(5)}$, $r_3^{(2)}$, and $r_4^{(6)}$. The last column indicates the 95% confidence interval for the estimated rate parameters.

Parameter	True value	Initial value	Estimated value	Confidence interval
k_1 [L^2 mol^{-2} min^{-1}]	0.0530	0.8000	0.0531	[0.0509, 0.0553]
k_2 [L^2 mol^{-2} min^{-1}]	0.1280	0.8000	0.1281	[0.1250, 0.1312]
k_3 [L min^{-1}]	0.0280	0.8000	0.0279	[0.0276, 0.0283]
k_4 [L^2 mol^{-2} min^{-1}]	0.0010	0.8000	0.0010	[0.0010, 0.0010]

reaction produces the desired product α-monochlorobutanoic acid (MBA) and hydrochloric acid (HCl). The second reaction produces the side product α-dichlorobutanoic acid (DBA) and HCl. HCl is a highly volatile product that is found in both phases.[33] The reaction stoichiometry is:

R1: BA $+$ Cl$_2$ $\rightarrow$ MBA $+$ HCl

R2: BA $+$ 2Cl$_2$ $\rightarrow$ DBA $+$ 2HCl.
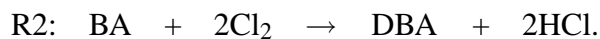
Generation of simulated data: The rate expressions of the reactions R1 and R2 are:

$$r_1 = k_1 \, c_{l,BA} \, c_{l,Cl_2} \, \sqrt{c_{l,MBA}}, \tag{36}$$

$$r_2 = k_2 \, r_1 \, c_{l,Cl_2}.$$

The mass-transfer rate expressions (in kg s^{-1}) are given by:

$$\zeta_{Cl_2} = k_{Cl_2} A_s V_l M_{w,Cl_2} \left(c^\star_{Cl_2} - c_{l,Cl_2}\right), \quad c^\star_{Cl_2} = \mathscr{P}_{Cl_2}/H_{Cl_2},$$

$$\zeta_{HCl} = -k_{HCl} A_s V_l M_{w,HCl} \left(c_{l,HCl} - c^\star_{HCl}\right), \quad c^\star_{HCl} = \mathscr{P}_{HCl}/H_{HCl}, \tag{37}$$

where A_s is the specific interfacial area, $c^\star_{Cl_2}$ and $c^\star_{HCl}$ are the equilibrium molar concentrations at the interface, c_{l,Cl_2} and $c_{l,HCl}$ are the molar concentrations in the liquid bulk, H_{Cl_2} and H_{HCl} are Henry's law constants, $\mathscr{P}_{Cl_2}$ and $\mathscr{P}_{HCl}$ are the partial pressures in the gas phase calculated using the ideal gas law from the numbers of moles in the gas phase. It is assumed that the gas side resistance is negligible. Since HCl transfers from the liquid to the gas, its mass-transfer rate is negative. The values of the rate parameters are given in Table 4. The thermodynamic and hydrodynamic parameter values are given in Appendix E.

The measurements are generated by simulating the transient operation of a reactor with inlet and outlet flows. The initial pressure in the reactor is 1 atm. The total pressure is regulated at 10 atm by manipulating the inlet flowrate of Cl$_2$ as shown in Figure 5a and Figure 5d. The gaseous outlet flowrate $u_{out,g}$ is constant at 3600 kg h^{-1}. The liquid inlet flowrate of BA $u_{in,BA}$ is 324 kg h^{-1}, while the liquid outlet flow $u_{out,l}$ is manipulated to regulate the liquid mass at 483 kg as shown

29

in Figure 5b. It is assumed that concentration measurements of Cl_2, BA and MBA in the liquid phase ($S_{l,a} = 3$) and HCl ($S_{g,a} = 1$) in the gas phase are available. They are corrupted with additive zero-mean Gaussian noise with standard deviation equal to 2% of the maximal concentration of the corresponding species.
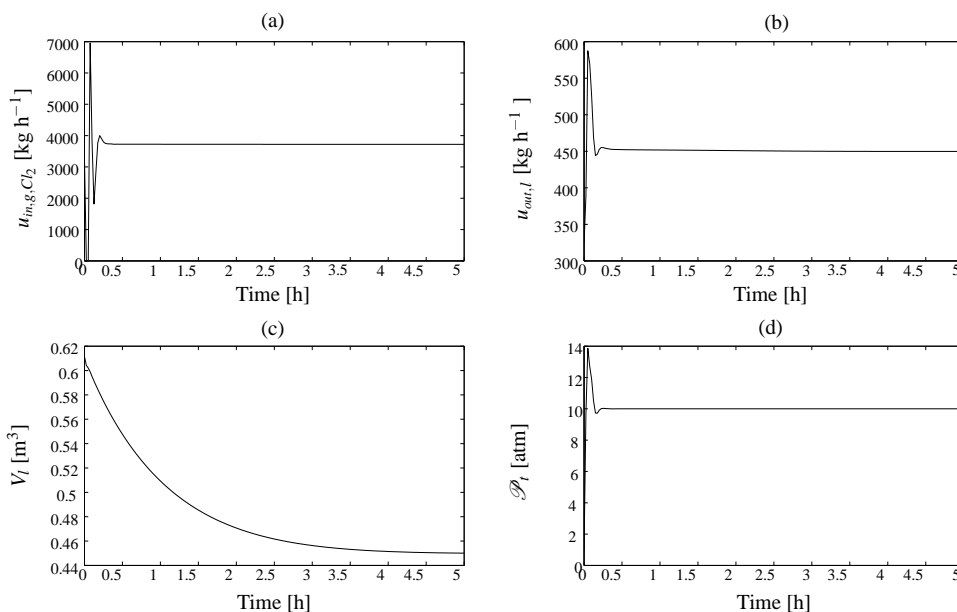

Figure 5: Noise-free measurements in the chlorination of butanoic acid: (a) Inlet flowrate of Cl_2 in the gas phase (u_{in,g,Cl_2}) manipulated to control the total pressure, (b) liquid outlet flowrate ($u_{out,l}$) manipulated to control the liquid mass, (c) liquid volume (V_l), and (d) total pressure ($\mathcal{P}_t$).

Identification of reaction and mass-transfer kinetics: The measurements include the liquid volume, the inlet flowrate of Cl_2, the liquid outlet flowrate, and the total pressure are measured (Figure 5) as well as four concentrations or numbers of moles (Figure 6). Since the reactions R1 and R2 are autocatalytic, they start slowly because the amount of MBA present in the reactor initially is small. This leads to accumulation of Cl_2 in the liquid phase (Figure 6a). After about 0.15 h, both reaction rates increase, which leads to a sharp decrease in the amount of Cl_2 in the liquid phase. The amount of BA in the liquid phase decreases momentarily to increase later since the amount of BA added by the liquid inlet surpasses its removal via the reactions and the liquid

outlet. Moreover, the behavior of $n_{g,HCl}$ in Figure 6d indicates that the gas phase has reached steady state after 0.2 h.
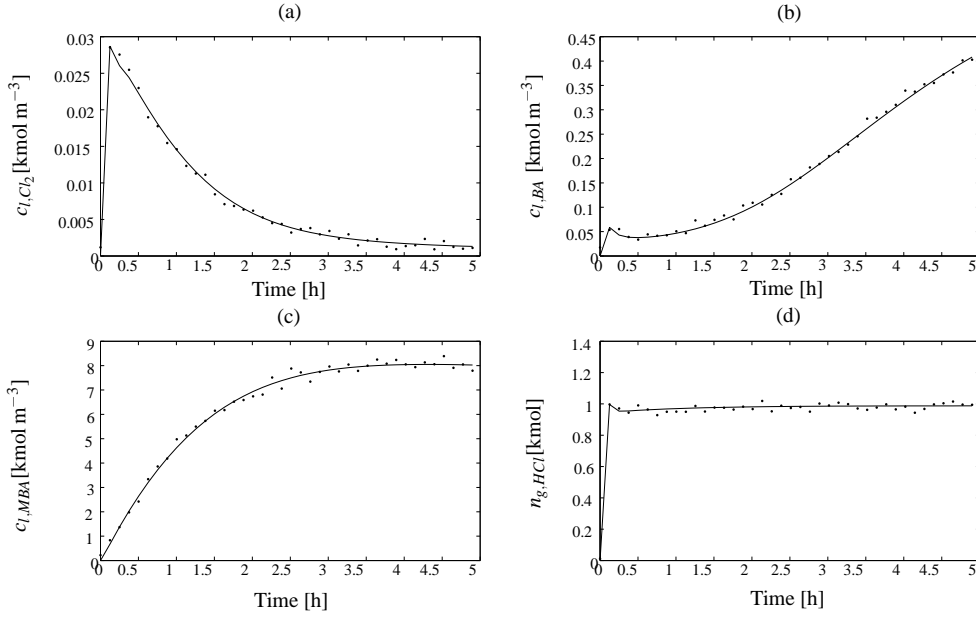

Figure 6: Composition measurements in the chlorination of butanoic acid: (a) Concentration of Cl_2 in the liquid phase, (b) concentration of MBA in the liquid phase, (c) concentration of BA in the liquid phase, and (d) number of moles HCl in the gas phase. The solid lines indicate the generated (true) compositions, while the markers indicate the noisy measurements.

The stoichiometric and inlet-composition matrices corresponding to the measured (available) quantities are:

$$\mathbf{N}_a = \begin{bmatrix} -1 & -1 & 1 \\ -2 & -1 & 0 \end{bmatrix}; \ \mathbf{W}_{m_l,l,a} = \begin{bmatrix} 0 & 0.0141 & 0 \end{bmatrix};$$

$$\mathbf{W}_{m_g,g,a} = \begin{bmatrix} -0.0274 \end{bmatrix}; \ \mathbf{W}_{in,g,a} = 0; \ \mathbf{W}_{m_g,l,a} = \begin{bmatrix} 0.0274 \end{bmatrix}; \ \mathbf{W}_{in,l,a} = \begin{bmatrix} 0.0113 \end{bmatrix}.$$

The measurements satisfy Assumptions (i)-(v) in Proposition 4. Hence, the extents of reaction $x_{r,1}$ and $x_{r,2}$ and the extents of mass transfer x_{m,l,Cl_2} and $x_{m,g,HCl}$ can be computed as mentioned in Proposition 4 and the results are shown in Figure 7. One sees that $x_{r,1}$ is significantly smaller than $x_{r,2}$ (Figure 7a-b). Since HCl transfers from the liquid to the gas, $x_{m,g,HCl}$ is negative and, furthermore, it is nearly constant after 0.2 h due to the steady-state condition in the gas phase

31

(Figure 7c). The mass-transfer extent x_{m,l,Cl_2} increases with time because of the large driving force maintained by the consumption of Cl_2 in the reactions. The number of moles of Cl_2 in the gas phase and the concentration of HCl in the liquid phase are reconstructed from the available measurements using Eqs. (30) and (31) (Figure 8). Note that these reconstructed quantities are required for parameter estimation.
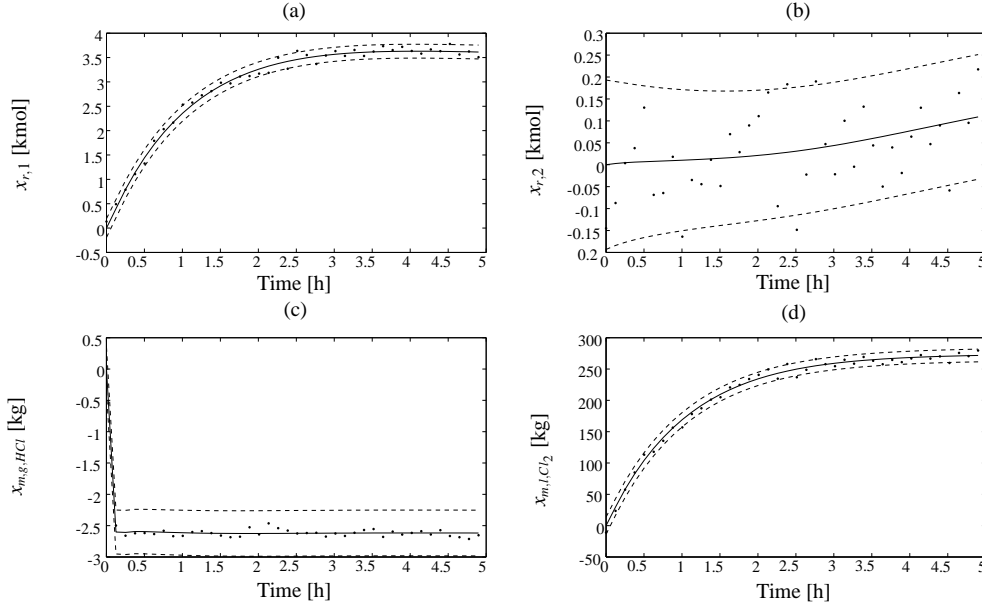

Figure 7: Extents of (a) reaction R1, (b) reaction R2, (c) HCl mass transfer to the gas phase, and (d) Cl_2 mass transfer to the liquid phase. The solid lines indicate the true extents of reaction and mass transfer, while the markers indicate the extents that are computed from the noisy measurements. The dash lines indicate the 95% confidence intervals.

The next step is to identify the rate expressions and the corresponding rate parameters. In this example, we assume that the the mass-transfer rate expressions are known.[3] The reaction rate expressions need to be identified from the set in Table 3. Note that the candidates $r_1^{(4)}$ and $r_2^{(4)}$ correspond to the true rate expressions. For a given reaction, each rate expression candidate is fitted to the computed extent using the least-squares problem (32). For Reaction R1, the quality of fit of the four rate expression candidates is shown in Figure 9. Candidate $r_1^{(4)}$ leads to the best fit for $\mathbf{W}_r = \mathbf{I}_H$. Similarly, for Reaction R2, the rate expression $r_2^{(4)}$ is found suitable. The identified

---

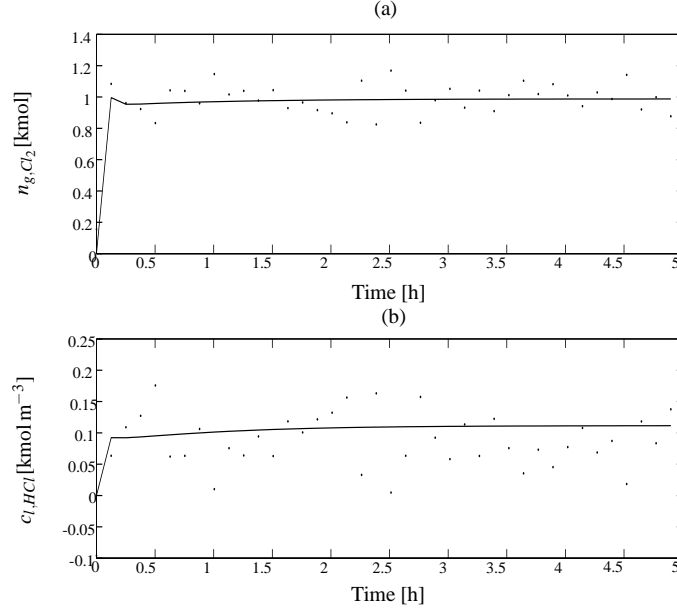[3]Additional mass-transfer rate expressions could be found elsewhere.[34]

Figure 8: Reconstructed compositions: (a) Number of moles of Cl_2 in the gas phase and (b) concentration of HCl in the liquid phase. The solid lines indicate the true compositions, while the markers indicate the values reconstructed from noisy measurements.

rate expressions correspond to the "simulated true" rate expressions used for generating the data.

The true values, the initial guesses, and the estimated values of the rate parameters with their 95% confidence intervals are given in Table 4. True and estimated values of the reaction rate parameters are very close to each other. However, there is a bias in the estimated mass-transfer rate parameters, for which the true values lie outside the computed confidence intervals.

Table 3: Reaction rate candidates for the chlorination of butanoic acid. Candidates $r_1^{(4)}$ and $r_2^{(4)}$ correspond to the simulated (true) reaction rate expressions.

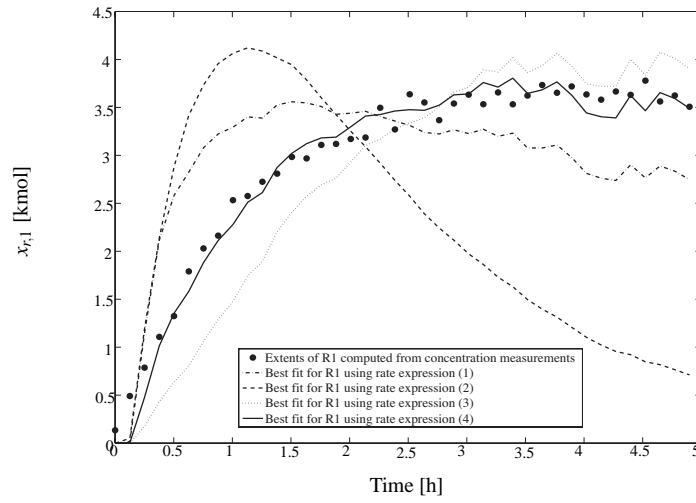| Reaction R1:	Reaction R2:
$r_1^{(1)} = k_1 \, c_{l,BA} \, c_{l,Cl_2}$	$r_1^{(1)} = k_2 \, c_{l,BA} \, c_{l,Cl_2}^2$
$r_1^{(2)} = k_1 \, c_{l,Cl_2}$	$r_2^{(2)} = k_2 \, c_{l,BA} \, c_{l,Cl_2}$
$r_1^{(3)} = k_1 \, c_{l,BA} \, c_{l,Cl_2} \, c_{l,MBA}$	$r_2^{(3)} = k_1 \, k_2 \, c_{l,BA} \, c_{l,Cl_2}^2 \, c_{l,MBA}$
$r_1^{(4)} = k_1 \, c_{l,BA} \, c_{l,Cl_2} \, \sqrt{c_{l,MBA}}$	$r_2^{(4)} = k_2 k_1 \, c_{l,BA} \, c_{l,Cl_2} \, \sqrt{c_{l,MBA}}$

Figure 9: Fit of four rate expressions to the computed extent $x_{r,1}$.

Table 4: Parameter estimation using the extent-based incremental identification approach: True, initial and estimated values of the reaction and mass-transfer rate parameters. The last column indicates the 95% confidence interval for the estimated parameters.

Parameter	True value	Initial value	Estimated value	Confidence interval
k_1 [m^3 kmol^{-1}]	1.3577	1.8000	1.3488	[1.3250, 1.3726]
k_2 [-]	0.1	0.0100	0.1002	[0.0877, 0.1126]
k_{Cl_2} [m s^{-1}]	0.666×10^{-4}	2×10^{-4}	0.591×10^{-4}	[0.580×10^{-4}, 0.602×10^{-4}]
k_{HCl} [m s^{-1}]	0.845×10^{-4}	2×10^{-4}	0.810×10^{-4}	[0.804×10^{-4}, 0.815×10^{-4}]

Conclusions

Being able to compute the extent of each reaction and each mass transfer from measured concentrations without knowledge of the reaction and mass-transfer kinetics is of great help in the investigation of reaction systems. This paper has proposed an incremental approach for the identification of reaction systems that uses the concept of extents and the integral method of parameter estimation. The proposed approach proceeds in two steps: (i) computation of the extents of reaction and mass transfer from concentration measurements, and (ii) identification of the individual reaction and mass-transfer kinetics from the computed extents using the integral method.

For homogeneous reaction systems, when the concentrations of all species are measured, the extents of reaction are computed in a straightforward manner via linear transformation of the numbers of moles. When the concentrations of only a subset of species are measured, additional information is necessary, in particular regarding the inlet and outlet flowrates, thus leading to a flow-dependent approach.

For G–L reaction systems, a similar linear transformation computes the extents of reaction and mass transfer from the concentrations of all species in the liquid phase. When the concentrations are available for only subsets of the species in the gas and liquid phases, a flow-dependent approach computes the extents of reaction and mass transfer by solving a set of differential-algebraic equations that involves the inlet and outlet flowrates and the liquid and gas masses. Proposition 4 specifies the minimal number of concentration measurements and the conditions needed to compute the extents of reaction and mass transfer.

Future work will compare the incremental identification approach proposed in this paper (Path "3" in Figure 1) with the one in the literature (Path "2" in Figure 1). The differential method in Path "2" requires time differentiation of noisy and sparse data, which is a difficult task that calls for appropriate regularization, while the integral method in Path "3" requires integration of both low- and high-resolution data. A detailed investigation of the relative merits and drawbacks of numerical differentiation and numerical integration will help compare the accuracy and the computational efficiency of the approaches. Moreover, it is proposed to develop an appropriate

metric to investigate the discrimination power of the two incremental identification approaches with respect to competing rate expressions.

Appendix A: Algorithm to compute S_0, M_0 and q_0

The algorithm is written generically for S species, R independent reactions, p inlet streams, the stoichiometric matrix $\mathbf{N}$ and the inlet matrix $\mathbf{W}_{in}$. It assumes $\text{rank}([\mathbf{N}^{\mathrm{T}}\ \mathbf{W}_{in}]) = R + p$, although it can easily be extended to the case of $\text{rank}([\mathbf{N}^{\mathrm{T}}\ \mathbf{W}_{in}]) < R + p$.

Let's first compute the matrices $\mathbf{Q}$, $\mathbf{L}$, $\mathbf{M}$ and $\mathbf{S}$, where $\mathbf{L}$ is a $(S \times p)$-dimensional auxiliary matrix, which fulfill the conditions:

C1: The $S \times S$ matrix $[\mathbf{N}^{\mathrm{T}}\ \mathbf{L}\ \mathbf{Q}]$ is of rank S,

C2: The columns of $\mathbf{Q}$ are orthonormal and span the null space of $[\mathbf{N}^{\mathrm{T}}\ \mathbf{W}_{in}]^{\mathrm{T}}$,

C3: The columns of $\mathbf{L}$ are orthonormal and span the null space of $[\mathbf{N}^{\mathrm{T}}\ \mathbf{Q}]^{\mathrm{T}}$,

C4: $\mathbf{M}^{\mathrm{T}}\mathbf{W}_{in} = \mathbf{I}_p$, which can be achieved by choosing $\mathbf{M} = \mathbf{L}(\mathbf{W}_{in}^{\mathrm{T}}\mathbf{L})^{+}$.

The following algorithm computes successively the matrices $\mathbf{Q}$, $\mathbf{L}$, $\mathbf{M}$ and $\mathbf{S}$:

1. Apply the singular value decomposition (SVD) to the matrix $[\mathbf{N}^{\mathrm{T}}\ \mathbf{W}_{in}]$:

$$[\mathbf{N}^{\mathrm{T}}\ \mathbf{W}_{in}] = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^{\mathrm{T}}.$$

   Let $\mathbf{U}_1 = [\mathbf{U}_{1,1}\ \mathbf{U}_{1,2}]$, where $\mathbf{U}_{1,1}$ and $\mathbf{U}_{1,2}$ are of dimension $S \times (R+p)$ and $S \times (S-R-p)$, respectively. Then, $\mathbf{Q} = \mathbf{U}_{1,2}$.

2. Note that $\text{rank}([\mathbf{N}^{\mathrm{T}}\ \mathbf{Q}]) = S - p$. Apply SVD to the matrix $[\mathbf{N}^{\mathrm{T}}\ \mathbf{Q}]$:

$$[\mathbf{N}^{\mathrm{T}}\ \mathbf{Q}] = \mathbf{U}_2\mathbf{S}_2\mathbf{V}_2^{\mathrm{T}}.$$

   Let $\mathbf{U}_2 = [\mathbf{U}_{2,1}\ \mathbf{U}_{2,2}]$, where $\mathbf{U}_{2,1}$ and $\mathbf{U}_{2,2}$ are of dimension $S \times (S-p)$ and $S \times (p)$, respectively. Then, $\mathbf{L} = \mathbf{U}_{2,2}$.

3. Compute $\mathbf{M} = \mathbf{L}(\mathbf{W}_{in}^{\mathrm{T}}\mathbf{L})^{+}$.

4. Compute $\mathbf{S}^{\mathrm{T}} = \mathbf{N}^{\mathrm{T}+}(\mathbf{I}_S - \mathbf{W}_{in}\mathbf{M}^{\mathrm{T}})$.

5. Using the initial conditions $\mathbf{n}_0$, the matrices $\mathbf{q}_0$, $\mathbf{S}_0$ and $\mathbf{M}_0$ are defined as:

$$\mathbf{q}_0^{\mathrm{T}} = \frac{\mathbf{1}_{S-R-p}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}}{\mathbf{1}_{S-R-p}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{n}_0}, \qquad \mathbf{S}_0^{\mathrm{T}} = \mathbf{S}^{\mathrm{T}}(\mathbf{I}_S - \mathbf{n}_0\,\mathbf{q}_0^{\mathrm{T}}), \qquad \mathbf{M}_0^{\mathrm{T}} = \mathbf{M}^{\mathrm{T}}(\mathbf{I}_S - \mathbf{n}_0\,\mathbf{q}_0^{\mathrm{T}}). \tag{38}$$

Appendix B: Proof of Proposition 2

The numbers of moles of the available species $\mathbf{n}_a(t_h)$ can be computed from the volume $V(t_h)$ and the available concentrations $\mathbf{c}_a(t_h)$. On the other hand, from Theorem 2 in Amrhein et al.[21], $\mathbf{n}_a(t_h)$ can be written as:

$$\mathbf{n}_a(t_h) = \mathbf{N}_a^{\mathrm{T}}\,\mathbf{x}_r(t_h) + \mathbf{W}_{in,a}\,\mathbf{x}_{in}(t_h) + \mathbf{n}_{0,a}\,\lambda(t_h). \tag{39}$$

The states $\mathbf{x}_{in}(t)$ and $\lambda(t)$ can be computed from $\mathbf{u}_{in}(t)$ and $u_{out}(t)$ using Eqs. (14) and (15). Furthermore, $(\mathbf{N}_a^{\mathrm{T}})^+$ exists from the assumption rank $(\mathbf{N}_a) = R$. Hence, the extents of reaction $\mathbf{x}_r(t_h)$ can be computed from $\mathbf{n}_a(t_h)$ and the flow extents $\mathbf{x}_{in}(t_h)$ and $\lambda(t_h)$ according to Eq. (16).

Appendix C: Proof of Proposition 4

The numbers of moles of the available species in the gas phase $\mathbf{n}_{g,a}(t_h)$ can be computed from $\mathbf{c}_{g,a}(t_h)$ as follows: $\mathbf{n}_{g,a}(t_h) = (V_r - V_l(t_h))\mathbf{c}_{g,a}(t_h)$. On the other hand, from Corollary 1 in Bhatt et al.[22], $\mathbf{n}_{g,a}(t_h)$ can be written in terms of various extents as:

$$\mathbf{n}_{g,a}(t_h) = -\mathbf{W}_{m,g,a}\mathbf{x}_{m,g}(t_h) + \mathbf{W}_{in,g,a}\mathbf{x}_{in,g}(t_h) + \mathbf{n}_{g0,a}\lambda_g(t_h). \tag{40}$$

Since $\mathbf{u}_{in,g}(t)$, $u_{out,g}(t)$ and $m_g(t)$ are available, $\mathbf{x}_{in,g}(t)$ and $\lambda_g(t)$ can be computed from Eqs. (22)-(24). Furthermore, since it is assumed that the $S_{g,a} = p_{m_g}$ species measured in the gas phase are involved in mass transfer, $\mathbf{W}_{m,g,a}$ can be partitioned as $\mathbf{W}_{m,g,a} = [\mathbf{W}_{mg,g,a}\ \mathbf{0}_{p_{m_g}\times(p_m-p_{m_g})}]$ and Eq. (40)

written as:

$$\mathbf{n}_{g,a}(t_h) = -\mathbf{W}_{m_g,g,a}\mathbf{x}_{m_g,g}(t_h) + \mathbf{W}_{in,g,a}\mathbf{x}_{in,g}(t_h) + \mathbf{n}_{g0,a}\lambda_g(t_h), \tag{41}$$

which corresponds to the last equation in Eq. (22). Since $\text{rank}\left(\mathbf{W}_{m_g,g,a}\right) = p_{m_g}$, its inverse exists, and the p_{m_g} extents $\mathbf{x}_{m_g,g}(t)$ can be computed.

Furthermore, since $u_{out,l}(t)$, $m_l(t)$, $u_{out,g}(t)$ and $m_g(t)$ are measured, the p_{m_g} extents of mass transfer $\mathbf{x}_{m_g,l}(t)$ can be computed from Eq. (23).

The numbers of moles $\mathbf{n}_{l,a}(t_h)$ can be computed from $\mathbf{c}_{l,a}(t_h)$ and $V_l(t_h)$ as $\mathbf{n}_{l,a}(t_h) = V_l(t_h)\mathbf{c}_{l,a}(t_h)$. On the other hand, from Theorem 1 in Bhatt $et\ al.$[22], $\mathbf{n}_{l,a}(t_h)$ can be written as:

$$\begin{aligned}
\mathbf{n}_{l,a}(t_h) &= \mathbf{N}_a^{\mathrm{T}}\mathbf{x}_r(t_h) + \mathbf{W}_{m,l,a}\mathbf{x}_{m,l}(t_h) + \mathbf{W}_{in,l,a}\mathbf{x}_{in,l}(t_h) + \mathbf{n}_{l0,a}\lambda_l(t_h) \tag{42}\\
&= \bar{\mathbf{N}}_a^{\mathrm{T}}\begin{bmatrix} \mathbf{x}_r(t_h) \\ \mathbf{x}_{m_l,l}(t_h) \end{bmatrix} + \mathbf{W}_{m_g,l,a}\mathbf{x}_{m_g,l}(t_h) + \mathbf{W}_{in,l,a}\mathbf{x}_{in,l}(t_h) + \mathbf{n}_{l0,a}\lambda_l(t_h), \tag{43}
\end{aligned}$$

which corresponds to the last equation in Eq. (24). Since $\mathbf{u}_{in,l}(t)$, $u_{out,l}(t)$ and $m_l(t)$ are available, $\mathbf{x}_{in,l}(t)$ and $\lambda_l(t)$ can be computed from Eq. (24). The pseudo inverse of matrix $\bar{\mathbf{N}}_a^{\mathrm{T}}$ exists since it is full rank by assumption. Hence, the extents of reaction $\mathbf{x}_r(t_h)$ and $\mathbf{x}_{m_l,l}(t_h)$ can be computed by inverting the matrix $\bar{\mathbf{N}}_a^{\mathrm{T}}$, which proves the proposition.

Appendix D: Proof of Proposition 5

The numbers of moles of the available species $\bar{\mathbf{n}}_a(t_h)$ can be computed from the volume $V_l(t_h)$ and the concentrations $\bar{\mathbf{c}}_a(t_h)$. On the other hand, $\bar{\mathbf{n}}_a(t_h)$ can be written as: [22]

$$
\begin{aligned}
\bar{\mathbf{n}}_a(t_h) &= \mathbf{N}_a^{\mathrm{T}} \mathbf{x}_r(t_h) + \bar{\mathbf{W}}_{in,a} \mathbf{x}_{in}(t_h) + \bar{\mathbf{n}}_{0,a} \\
&= \mathbf{N}_a^{\mathrm{T}} \mathbf{x}_r(t_h) + \begin{bmatrix} \mathbf{W}_{in,g,a} & \mathbf{W}_{in,l,a} & \mathbf{W}_{m,a} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{in,g}(t_h) \\ \mathbf{x}_{in,l}(t_h) \\ \mathbf{x}_m(t_h) \end{bmatrix} + \bar{\mathbf{n}}_{0,a} \\
&= \mathbf{N}_a^{\mathrm{T}} \mathbf{x}_r(t_h) + \mathbf{W}_{in,g,a} \mathbf{x}_{in,g}(t_h) + \mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t_h) + \mathbf{W}_{m,a} \mathbf{x}_m(t_h) + \bar{\mathbf{n}}_{0,a} \\
&= \bar{\mathbf{N}}_a^{\mathrm{T}} \begin{bmatrix} \mathbf{x}_r(t_h) \\ \mathbf{x}_m(t_h) \end{bmatrix} \mathbf{x}_r(t_h) + \mathbf{W}_{in,g,a} \mathbf{x}_{in,g}(t_h) + \mathbf{W}_{in,l,a} \mathbf{x}_{in,l}(t_h) + \bar{\mathbf{n}}_{0,a}.
\end{aligned}
$$

$$(44)$$
$$(45)$$
$$(46)$$
$$(47)$$

Since $\mathbf{u}_{in,g}(t)$ and $\mathbf{u}_{in,l}(t)$ are measured, $\mathbf{x}_{in,g}(t_h)$ and $\mathbf{x}_{in,l}(t_h)$ can be computed from Eqs. (27) and (28). Furthermore, $(\bar{\mathbf{N}}_a^{\mathrm{T}})^+$ exists from the assumption rank $(\bar{\mathbf{N}}_a) = R + p_m$. Hence, the vector of extents of reaction $\mathbf{x}_r(t_h)$ and mass transfer $\mathbf{x}_m(t_h)$ can be computed from $\bar{\mathbf{n}}_a(t_h)$ and the flow extents $\mathbf{x}_{in,g}(t_h)$ and $\mathbf{x}_{in,l}(t_h)$ according to Eq. (29). The first R elements of the computed vector correspond to the extents of reaction, while the remaining p_m elements correspond to the extents of mass transfer.

Appendix E: Data for the Chlorination of Butanoic Acid

Thermodynamic data	Henry constants, $H_{c,Cl_2} = H_{c,HCl} = 70.33$ [bar m^3 kmol^{-1}]
	Molecular weights [kg kmol^{-1}]
	$\quad M_{w,BA} = 88.12$
	$\quad M_{w,MBA} = 122.52$
	$\quad M_{w,DBA} = 156.97$
	$\quad M_{w,Cl_2} = 71$
	$\quad M_{w,HCl} = 36.45$
	$\quad M_{w,EtOH} = 46$
	Liquid densities of pure species [kg m^{-3}]
	$\quad \rho_{l,BA} = 859.17$
	$\quad \rho_{l,MBA} = 1085.53$
	$\quad \rho_{l,DBA} = 1070$
	$\quad \rho_{l,Cl_2} = 1093$
	$\quad \rho_{l,HCl} = 1486.26$
	$\quad \rho_{l,EtOH} = 790$
Hydrodynamic data	Specific interfacial area, $A_s = 254.9$ [m^{-1}]

References

1. Walter, E.; Pronzato, L. *Identification of Parametric Models from Experimental Data*; Springer-Verlag, Berlin, 1997.

2. Asprey, S.; Macchietto, S. Statistical tools for optimal dynamic model building. *Comp. Chem. Eng.* **2000**, *24*, 2161.

3. Montgomery, D. C. *Design and Analysis of Experiments*, 6th ed.; John Wiley & Sons, New York, 2005.

4. Franceschini, G.; Macchietto, S. Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.* **2008**, *63*, 4846.

5. Bard, Y. *Nonlinear Parameter Estimation*; Academic Press, New York, 1974.

6. Brendel, M. L. *Incremental Identification of Complex Reaction Systems*; Doctoral thesis, RWTH Aachen University, Germany, 2005.

7. Bardow, A.; Marquardt, W. Incremental and simultaneous identification of reaction kinetics: Methods and comparison. *Chem. Eng. Sci.* **2004**, *59*, 2673.

8. Brendel, M.; Bonvin, D.; Marquardt, W. Incremental identification of kinetic models for homogeneous reaction systems. *Chem. Eng. Sci.* **2006**, *61*, 5404.

9. Marquardt, W. Model-based experimental analysis of kinetic phenomena in multi-phase reactive systems. *Chem. Eng. Res. Des.* **2005**, *83(A6)*, 561–573.

10. Michalik, C.; Brendel, M.; Marquardt, W. Incremental identification of fluid multi-phase reaction systems. *AIChE J.* **2009**, *55*, 1009.

11. Bonvin, D.; Rippin, D. Target factor analysis for the identification of stoichiometric models. *Chem. Eng. Sci.* **1990**, *45*, 3417.

12. Amrhein, M.; Srinivasan, B.; Bonvin, D. Target factor analysis of reaction data: Use of data pre-treatment and reaction-invariant relationships. *Chem. Eng. Sci.* **1999**, *54*, 579.

13. Prinz, O.; Bonvin, D. Monitoring chemical reaction systems using incremental target factor analysis. *IFAC Symposium ADCHEM*, Kyoto, Japan, 1994.

14. Levenspiel, O. *Chemical Reaction Engineering*, 3rd ed.; John Wiley & Sons, New York, 1998.

15. Froment, G. F.; Bischoff, K. B. *Chemical Reactor Analysis and Design*; John Wiley & Sons, New York, 1979.

16. Yeow, Y. L.; Wickramasinghe, S. R.; Han, B.; Leong, Y. K. A new method of processing the time-concentration data of reaction kinetics. *Chem. Eng. Sci.* **2003**, *58*, 3601–3610.

17. Hosten, L. H. A comparative study of short cut procedures for parameter estimation in differential equations. *Comp. Chem. Eng.* **1979**, *3*, 117–126.

18. Schaich, D.; Becker, R.; King, R. Qualitative modelling for automatic identification of mathematical models of chemical reaction systems. *Control Engineering Practice* **2001**, *9*, 1373–1381.

19. Kristensen, N. R.; Madsen, H.; Jørgensen, S. B. Developing phenomena models from experimental data. *336th European Symposium of the Working Party on Computer Aided Process Engineering*, Lappeenranta, Finland, 2003; pp 1091–1096.

20. Kristensen, N. R.; Madsen, H.; Jørgensen, S. B. A method for systematic improvement of stochastic grey-box models. *Comp. Chem. Eng.* **2004**, *28*, 1431–1449.

21. Amrhein, M.; Bhatt, N.; Srinivasan, B.; Bonvin, D. Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams. *AIChE J.* **2010**, *56*, 2873.

22. Bhatt, N.; Amrhein, M.; Bonvin, D. Extents of reaction, mass transfer and flow for Gas–Liquid reaction systems. *Ind. Eng. Chem. Res.* **2010**, *49*, 7704.

23. Butcher, J. *Numerical Methods for Ordinary Differential Equations*; John Wiley & Sons, England, 2003.

24. Ljung, L.; Glad, T. On global identifiability for arbitrary model parametrizations. *Automatica* **1994**, *30*, 265.

25. Walter, E.; Pronzato, L. On the identifiability and distinguishability of nonlinear parametric models. *Math. and Comp. in Sim.* **1996**, *42*, 125.

26. Davidescu, F. P.; Jørgensen, S. B. Structural parameter identifiability analysis for dynamic reaction networks. *Chem. Eng. Sci.* **2008**, *63*, 4754.

27. Srinatha, S.; Gunawan, R. Parameter identifiability of power-law biochemical system models. *J. of Biotech.* **2010**, *149*, 132.

28. Turányi, T. Sensitivity analysis of complex kinetic systems. Tools and applications. *J. of Math. Chem.* **1990**, *5*, 203.

29. Seber, G.; Wild, C. *Nonlinear Regression*; John Wiley & Sons, New York, 1989.

30. van der Heijden, F.; Duin, R.; de Ridder, D.; Tax, D. *Classification, Parameter Estimation, and State Estimation: An Engineering Approach using MATLAB®*; John Wiley & Sons, Chichester, 2004.

31. Bhatt, N.; Karimoglu, N.; Amrhein, M.; Marquardt, W.; Bonvin, D. Incremental model identification for reaction systems - A comparison of rate-based and extent-based approaches. *Chem. Eng. Sci.* **2011**, Submitted.

32. Ruppen, D.; Bonvin, D.; Rippin, D. W. T. Implementation of adaptive optimal operation for a semi-batch reaction system. *Comp. Chem. Eng.* **1998**, *22*, 185–199.

33. Zhang, J.; Smith, R. Design and optimization of batch and semi-batch reactors. *Chem. Eng. Sci.* **2004**, *59*, 459.

34. Taylor, R.; Krishna, R. *Multicomponent Mass Transfer*, 1st ed.; John Wiley & Sons, New York, 1993.