

Data Reconciliation in Reaction Systems using the Concept of Extents

Sriniketh Srinivasan^a, Julien Billeter^a, Shankar Narasimhan^b and Dominique Bonvin^a

^a*Laboratoire d'Automatique, EPFL, Lausanne, Switzerland*

^b*Department of Chemical Engineering, IIT Madras, Chennai, India*
sriniketh.srinivasan@epfl.ch

Abstract

Concentrations measured during the course of a chemical reaction are corrupted with noise, which reduces the quality of information. When these measurements are used for identifying kinetic models, the noise impairs the ability to identify accurate models. The noise in concentration measurements can be reduced using data reconciliation, exploiting for example the material balances as constraints. However, additional constraints can be obtained via the transformation of concentrations into extents and invariants. This paper uses the transformation to extents and invariants and formulates the data reconciliation problem accordingly. This formulation has the advantage that non-negativity and monotonicity constraints can be imposed on selected extents. A simulated example is used to demonstrate that reconciled measurements lead to the identification of more accurate kinetic models.

Keywords: Kinetic identification, Data reconciliation, Model discrimination, Vessel extents

1. Introduction

Chemical reactions are used in the chemical, biotechnological and pharmaceutical industries to convert feed materials into manufactured products. Measurements of concentrations made during the course of the reaction are vital for efficient reactor operation. Process monitoring, control and optimization can be carried out using measurements by either building first-principles (kinetic) models or using data-driven models, or by combining these two approaches.

Kinetic modeling of chemical reaction systems is generally performed via simultaneous identification, Bardow and Marquardt (2004). This identification path suffers from combinatorial complexity and is therefore computationally intensive. As an alternative, the extent-based incremental identification introduced by Amrhein et al. (2010) can be used to build first-principles kinetic models incrementally. The procedure involves a transformation of the measured numbers of moles to extents. This transformation decouples the modeling task into a set of sub-problems, thereby reducing the combinatorial complexity when there are several candidate models for each reaction. Since measurements are corrupted by noise, the performance of the modeling/identification task, and thus also of the subsequent monitoring, control and optimization steps, depends highly on the accuracy of the measurements.

To reduce noise and improve the accuracy of the measured information, data reconciliation (DR) techniques are often used as a pre-processing step, see Narasimhan and Jordache (1999). Data reconciliation exploits process constraints derived from conservation equations to reconcile measurements, that is, to correct the measured data so as to satisfy constraints that are valid at all times. This paper describes a reconciliation approach that is based on extents instead of concentrations. This novel problem formulation allows using additional constraints such as the non-negativity and

monotonicity of extents, which improves significantly the accuracy of the reconciled concentrations and of the identified kinetic models, as illustrated through a simulated example.

2. Preliminaries

2.1. Mole Balance Equations

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

$$\dot{\mathbf{n}}(t) = \mathbf{N}^T \mathbf{r}_v(t) + \mathbf{W}_{in} \mathbf{u}_{in}(t) - \omega(t) \mathbf{n}(t), \quad \mathbf{n}(0) = \mathbf{n}_0, \quad (1)$$

where \mathbf{n} is the S -dimensional vector of the numbers of moles, $\mathbf{r}_v := V \mathbf{r}$ with V the volume and \mathbf{r} the R -dimensional vector of reaction rates, \mathbf{u}_{in} the p -dimensional vector of inlet mass flowrates, $\omega := \frac{u_{out}}{m}$ the inverse of the residence time with m the mass of the reaction mixture and u_{out} the outlet mass flowrate, \mathbf{N} the $R \times S$ stoichiometric matrix, $\mathbf{W}_{in} = \mathbf{M}_w^{-1} \check{\mathbf{W}}_{in}$ the $S \times p$ matrix of inlet compositions, \mathbf{M}_w the S -dimensional diagonal matrix of molecular weights, $\check{\mathbf{W}}_{in} = [\check{\mathbf{w}}_{in}^1 \cdots \check{\mathbf{w}}_{in}^p]$ with $\check{\mathbf{w}}_{in}^j$ being the S -dimensional vector of weight fractions of the j th inlet flow, and \mathbf{n}_0 the S -dimensional vector of initial numbers of moles. If needed, the concentration are computed from the numbers of moles as $\mathbf{c}(t) = \frac{\mathbf{n}(t)}{V(t)}$.

2.2. Conservation Equations

The S dynamic equations (1) are often redundant, as the variability in the system is determined by the number of independent reactions and inlet/outlet streams, and not by the number of chemical species. These redundancies can usually be expressed in terms of algebraic constraints. The number and nature of these constraints depend on the operating mode of the reactor. Table 1 lists the number of constraints q and the procedure for deriving them from structural elements of (1), in particular \mathbf{N} , \mathbf{W}_{in} and \mathbf{n}_0 , for three different operating modes. The matrix \mathbf{P} represents the null space of the corresponding structural matrix.

Table 1: Algebraic constraints under different operating modes.

Operation	# constraints	Constraint derivation	Constraints
Batch	$q = S - R$	$\mathbf{P}^T [\mathbf{N}^T] = \mathbf{0}_{q \times R}$	$\mathbf{P}^T \mathbf{n}(t) = \mathbf{P}^T \mathbf{n}_0$
Fed-batch	$q = S - R - p$	$\mathbf{P}^T [\mathbf{N}^T \mathbf{W}_{in}] = \mathbf{0}_{q \times (R+p)}$	$\mathbf{P}^T \mathbf{n}(t) = \mathbf{P}^T \mathbf{n}_0$
Open	$q = S - R - p - 1$	$\mathbf{P}^T [\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0] = \mathbf{0}_{q \times (R+p+1)}$	$\mathbf{P}^T \mathbf{n}(t) = \mathbf{0}_q$

2.3. From numbers of moles to vessel extents

Amrhein et al. (2010) and Rodrigues et al. (2015) have developed a linear transformation for open reaction systems that transforms the numbers of moles \mathbf{n} into four contributions, namely, the R extents of reaction \mathbf{x}_r , the p extents of inlet \mathbf{x}_{in} , a dimensionless extent of initial conditions x_{ic} , and the $q = S - R - p - 1$ invariants \mathbf{x}_{iv} that are identically zero. The linear transformation \mathcal{T} reads:

$$\begin{bmatrix} \mathbf{x}_r(t) \\ \mathbf{x}_{in}(t) \\ x_{ic}(t) \\ \mathbf{x}_{iv}(t) \end{bmatrix} = \mathcal{T} \mathbf{n}(t) := \begin{bmatrix} \mathbf{R} \\ \mathbf{F} \\ \mathbf{i}^T \\ \mathbf{Q} \end{bmatrix} \mathbf{n}(t), \quad (2)$$

The transformation matrix \mathcal{T} is given by $[\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0 \mathbf{P}]^{-1}$, where \mathbf{P} represents the null space of $[\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0]$, and brings the dynamic model (1) to the following decoupled form:

$$\dot{\mathbf{x}}_r(t) = \mathbf{r}_v(t) - \omega(t) \mathbf{x}_r(t) \quad \mathbf{x}_r(0) = \mathbf{0}_R \quad (3a)$$

$$\dot{\mathbf{x}}_{in}(t) = \mathbf{u}_{in}(t) - \omega(t) \mathbf{x}_{in}(t) \quad \mathbf{x}_{in}(0) = \mathbf{0}_p \quad (3b)$$

$$\dot{x}_{ic}(t) = -\omega(t) x_{ic}(t) \quad x_{ic}(0) = 1 \quad (3c)$$

$$\dot{\mathbf{x}}_{iv}(t) = \mathbf{0}_q, \quad (3d)$$

where \mathbf{R} , \mathbf{F} and $\mathbf{Q} = \mathbf{P}^+$ are matrices of dimensions $R \times S$, $p \times S$ and $q \times S$, respectively, and \mathbf{i} is a S -dimensional vector. Each individual extent $x_{r,i}$ or $x_{in,j}$ defined in (3) is in fact a vessel extent expressing the amount of material (due to a reaction or an inlet) that is still in the reactor, the negative terms on the right-hand side accounting for what has left the reactor. Similarly, the extent $x_{ic}(t)$ in (3c) indicates the fraction of the initial conditions that is still in the reactor at time t .

The vector of numbers of moles $\mathbf{n}(t)$ can be reconstructed from the various extents by pre-multiplying (2) by $\mathcal{F}^{-1} = [\mathbf{N}^T \mathbf{W}_{in} \mathbf{n}_0 \mathbf{P}]$ and considering that $\mathbf{x}_{iv} = \mathbf{0}_q$, which yields:

$$\mathbf{n}(t) = \mathbf{N}^T \mathbf{x}_r(t) + \mathbf{W}_{in} \mathbf{x}_{in}(t) + \mathbf{n}_0 x_{ic}(t). \quad (4)$$

Remark 1 (Fed-batch reactor)

For fed-batch reactors, the transformation (2) transforms the \mathbf{n} into the R extents of reactions \mathbf{x}_r , the p extents of inlets \mathbf{x}_{in} and the $q = S - R - p$ invariants \mathbf{x}_{iv} .

Remark 2 (Batch reactor)

For batch reactors, the transformation (2) transforms \mathbf{n} into the R extents of reactions \mathbf{x}_r and the $q = S - R$ invariants \mathbf{x}_{iv} .

Proposition 1 (Monotonicity of x_{ic})

The extent $x_{ic}(t)$ is strictly positive and monotonically decreasing over time in an open reactor.

Proof: The solution to (3c) is $x_{ic}(t) = e^{-\int_0^t \omega(\tau) d\tau}$ with $\omega(t) = \frac{u_{out}(t)}{m(t)}$. Since $\omega(t) \geq 0$, $x_{ic}(t)$ varies as a negative exponential function and hence is monotonically decreasing.

Proposition 2 (Monotonicity of \mathbf{x}_r and \mathbf{x}_{in} in batch and fed-batch modes)

The extents \mathbf{x}_r for reactions with net positive rates and the extents \mathbf{x}_{in} are strictly increasing in batch and semi-batch modes.

Proof: In absence of outlet, that is for $u_{out} = 0$, (3a) reduces to $\dot{\mathbf{x}}_r(t) = \mathbf{r}_v(t)$ and (3b) to $\dot{\mathbf{x}}_{in}(t) = \mathbf{u}_{in}(t)$. It follows from $\mathbf{r}_v(t) \geq \mathbf{0}_R$ and $\mathbf{u}_{in}(t) \geq \mathbf{0}_p$ that both extents are strictly positive and increasing.

3. Data Reconciliation

Data reconciliation applied to reaction systems uses redundancy expressed as algebraic constraints, such as those provided in Table 1, to improve the accuracy of measured concentrations. The following assumptions are made in this study:

- A1. the concentrations of all S species are measured and available,
- A2. the volume is measured accurately without noise, and
- A3. the initial conditions are known perfectly.

Let $\tilde{\mathbf{c}}(t_h) = \mathbf{c}(t_h) + \boldsymbol{\epsilon}_c$ denote the S -dimensional vector of noisy concentrations measured at H time instants, $t_h \in [t_1, t_H]$ with $t_1 = 0$, where $\boldsymbol{\epsilon}_c$ is an S -dimensional vector of zero-mean Gaussian noise with the constant variance-covariance matrix $\boldsymbol{\Sigma}_c$. The noise in the numbers of moles $\tilde{\mathbf{n}}(t_h) = V(t_h) \tilde{\mathbf{c}}(t_h)$ also follows a zero-mean normal distribution with the time-varying variance-covariance matrix $\boldsymbol{\Sigma}_n(t_h) = V(t_h)^2 \boldsymbol{\Sigma}_c$.

3.1. Data reconciliation in terms of numbers of moles

For all operating modes, the reconciliation of the measured numbers of moles $\tilde{\mathbf{n}}$ can be formulated as an optimization problem constrained by algebraic relationships:

$$\begin{aligned} \min_{\hat{\mathbf{n}}(t_i)} \quad & (\tilde{\mathbf{n}}(t_i) - \hat{\mathbf{n}}(t_i))^T \mathbf{W}(t_i) (\tilde{\mathbf{n}}(t_i) - \hat{\mathbf{n}}(t_i)) \\ \text{s.t.} \quad & \mathbf{P}^T \hat{\mathbf{n}}(t_i) = \mathbf{0}_q, \quad \hat{\mathbf{n}}(t_i) \geq \mathbf{0}_S \end{aligned} \quad (5)$$

where $\hat{\mathbf{n}}$ is the vector of reconciled numbers of moles and $\mathbf{W}(t_i) = \boldsymbol{\Sigma}_n^{-1}(t_i)$ is the weighting matrix. Note that, due to the presence of inequality constraints, Problem (5) does not have an analytical solution and hence must be solved numerically. In this formulation, the DR problem for different time instants are decoupled, that is, the reconciled number of moles can be estimated independently at each time instant.

3.2. Data reconciliation in terms of extents

An alternative consists in defining the DR problem directly in terms of extents. The formulation of this problem depends on whether the reactor has an outlet (open reactors) or not (batch and fed-batch reactors). In this formulation, the reconciliation in terms of extents at instant t_i involves reconciling all the previous measurements as well.

3.2.1. Open reactors

In the presence of an outlet, there is a monotonicity constraint only for x_{ic} . The reconciliation problem for reactions with net positive rates reads:

$$\begin{aligned} \min_{\substack{\hat{\mathbf{x}}_r(t_1), \dots, \hat{\mathbf{x}}_r(t_H) \\ \hat{\mathbf{x}}_{in}(t_1), \dots, \hat{\mathbf{x}}_{in}(t_H) \\ \hat{x}_{ic}(t_1), \dots, \hat{x}_{ic}(t_H)}}} \quad & \sum_{i=1}^{t_H} (\tilde{\mathbf{n}}(t_i) - \hat{\mathbf{n}}(t_i))^T \mathbf{W}(t_i) (\tilde{\mathbf{n}}(t_i) - \hat{\mathbf{n}}(t_i)) \\ \text{s.t.} \quad & \hat{x}_{ic}(t_i) - \hat{x}_{ic}(t_{i-1}) \leq 0 \quad \forall i > 1, \\ & \hat{\mathbf{n}}(t_i) \geq \mathbf{0}, \quad \hat{\mathbf{x}}_r(t_i) \geq \mathbf{0}, \quad \hat{\mathbf{x}}_{in}(t_i) \geq \mathbf{0}, \quad \hat{x}_{ic}(t_i) \geq 0 \end{aligned} \quad (6)$$

with $\hat{\mathbf{n}}(t_i) = \mathbf{N} \hat{\mathbf{x}}_r(t_i) + \mathbf{W}_{in} \hat{\mathbf{x}}_{in}(t_i) + \mathbf{n}_0 \hat{x}_{ic}(t_i)$ from (4).

3.2.2. Batch and fed-batch reactors

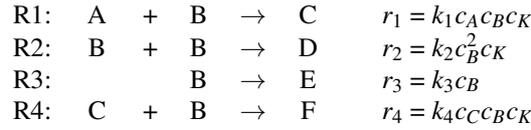
In the absence of outlet, monotonicity constraints can be imposed on \mathbf{x}_r in batch mode, and on \mathbf{x}_r and \mathbf{x}_{in} in fed-batch mode. For a fed-batch reactor, the reconciliation problem for reactions with net positive rates then reads:

$$\begin{aligned} \min_{\substack{\hat{\mathbf{x}}_r(t_1), \dots, \hat{\mathbf{x}}_r(t_H) \\ \hat{\mathbf{x}}_{in}(t_1), \dots, \hat{\mathbf{x}}_{in}(t_H)}}} \quad & \sum_{i=1}^{t_H} (\tilde{\mathbf{n}}(t_i) - \hat{\mathbf{n}}(t_i))^T \mathbf{W}(t_i) (\tilde{\mathbf{n}}(t_i) - \hat{\mathbf{n}}(t_i)) \\ \text{s.t.} \quad & \hat{\mathbf{x}}_r(t_i) - \hat{\mathbf{x}}_r(t_{i-1}) \geq \mathbf{0}_R \quad \forall i > 1, \quad \hat{\mathbf{x}}_{in}(t_i) - \hat{\mathbf{x}}_{in}(t_{i-1}) \geq \mathbf{0}_p \quad \forall i > 1, \\ & \hat{\mathbf{n}}(t_i) \geq \mathbf{0}, \quad \hat{\mathbf{x}}_r(t_i) \geq \mathbf{0}_R, \quad \hat{\mathbf{x}}_{in}(t_i) \geq \mathbf{0}_p \end{aligned} \quad (7)$$

with $\hat{\mathbf{n}}(t_i) = \mathbf{N} \hat{\mathbf{x}}_r(t_i) + \mathbf{W}_{in} \hat{\mathbf{x}}_{in}(t_i) + \mathbf{n}_0$ from (4) with $x_{ic}(t_i) = 1$. The constraint $\mathbf{P}^T \hat{\mathbf{n}}(t_i) = \mathbf{0}_q$ is implicitly satisfied in (6) and (7) since the invariants \mathbf{x}_{iv} are zero.

4. Application to Model Identification

The different formulations of the reconciliation problem are compared via a simulated example. The chosen reaction system is the catalyzed acetoacetylation of pyrrole, which consists of $R = 4$ reactions and $S = 6$ species (plus the catalyst), Ruppen et al. (1998). The reaction scheme and the kinetic models used to generate the data are:



This reaction system is simulated in an isothermal batch reactor. The values of the rate constants are $k_1 = 0.0530$, $k_2 = 0.1280$, $k_3 = 0.0280 \text{ s}^{-1}$ and $k_4 = 0.003 \text{ L}^2 \text{ mol}^{-2} \text{ s}^{-1}$. The initial volume is 1 L with the concentrations $\mathbf{c}_0 = [1.25 \ 1 \ 0 \ 0 \ 0 \ 0]^T \text{ mol L}^{-1}$ and $c_{K0} = 0.5 \text{ mol L}^{-1}$ for the catalyst K. Since the density is assumed constant, the volume is also constant.

According to Table 1, this batch reaction system has $q = S - R = 2$ invariant relationships:

$$\begin{aligned}
 -2c_A(t) + c_B(t) - c_C(t) + 2c_D(t) + c_E(t) &= -2c_{A,0} + c_{B,0} - c_{C,0} + 2c_{D,0} + c_{E,0} \\
 c_A(t) + c_C(t) + c_F(t) &= c_{A,0} + c_{C,0} + c_{F,0}.
 \end{aligned} \tag{8}$$

Each concentration is corrupted with additive independent zero-mean Gaussian noise of standard deviation corresponding to 10% of its maximum concentration. Catalyst K and species F are assumed to be noise free. The variance-covariance matrix Σ_n is constant (batch conditions) and is assumed to be known. Measurements are taken every 0.5 min for 30 min.

The measured numbers of moles $\hat{\mathbf{n}}$ are reconciled in two ways. First, in terms of numbers of moles, according to the problem formulation (5) using constraints given by the invariant relationships (8). Second, in terms of extents, according to the formulation (7) (with $\hat{\mathbf{x}}_m = \mathbf{0}$) with constraints on the positivity and monotonicity of the extents of reaction $\hat{\mathbf{x}}_r$. The performance of the different approaches is assessed using the residual sum of squares (RSS) calculated as the difference between the *true* simulated data and the measured/reconciled data. Table 2 shows that the formulation (7) improves significantly the accuracy of the reconciled data compared to the unreconciled (original) measurements and the data reconciled according to the formulation (5). Figure 1 shows the simulated, measured and extent-based reconciled concentrations of species A to E.

Table 2: Residual sum of squares between the *true* simulated numbers of moles and the numbers of moles obtained without reconciliation ($\hat{\mathbf{n}}$) and with reconciliation according to (5) and (7).

Species	Measurements via $\hat{\mathbf{n}}$	Data reconciliation	
		via $\hat{\mathbf{n}}$ (5)	via $\hat{\mathbf{x}}$ (7)
A	0.9164	0.0332	0.0108
B	0.4615	0.1317	0.0355
C	0.0316	0.0332	0.0109
D	0.0392	0.0268	0.0062
E	0.0162	0.0155	0.0115

Note that incremental model identification can be performed independently for each reaction, as discussed in Srinivasan et al. (2012) and Billeter et al. (2013). For each rate law, kinetic identification is performed using the measured data ($\hat{\mathbf{n}}$) and the data reconciled according to (5) and (7). For each set of data (measurements, reconciled estimates via $\hat{\mathbf{n}}$ and reconciled estimates via $\hat{\mathbf{x}}$), the capability of discriminating between correct and incorrect kinetic laws is assessed by comparing the RSS of each candidate rate law to that of the correct law. Table 3 shows drastic improvement when measurements are reconciled with (7), that is, using constraints on extents. In all cases, the least RSS is obtained for the correct kinetic model, which implies that the correct model is identified. However, the use of the raw measurements (without reconciliation) makes it difficult to discriminate between models 1 and 3 for R1 and models 1 and 4 for R2. In contrast, the use of reconciled estimates based on the formulation (7) enables unambiguous identification of the correct kinetic model.

Table 3: Ratios of RSS (with respect to the correct model, thus giving 1 for the correct model) for different kinetic models for reactions R1 and R2 using the measured numbers of moles $\hat{\mathbf{n}}$ and the numbers of moles reconciled according to 5 ($\hat{\mathbf{n}}$) and 7 ($\hat{\mathbf{x}}$).

R1	Measurements via $\hat{\mathbf{n}}$	Reconciled estimates		R2	Measurements via $\hat{\mathbf{n}}$	Reconciled estimates	
		via $\hat{\mathbf{n}}$ (5)	via $\hat{\mathbf{x}}$ (7)			via $\hat{\mathbf{n}}$ (5)	via $\hat{\mathbf{x}}$ (7)
$r_1 = k_1 c_A c_B c_K$	1	1	1	$r_2 = k_2 c_B^2 c_K$	1	1	1
$r_1 = k_1 c_A$	1.1761	3.3886	10.1717	$r_2 = k_2 c_B$	1.518	695	8.1858
$r_1 = k_1 c_A^2 c_B$	1.0017	1.0365	1.2355	$r_2 = k_2$	3.9516	12.6008	42.5469
$r_1 = k_1 c_A c_B^2$	1.1201	2.1703	6.7232	$r_2 = k_2 c_B^3$	1.0593	1.4487	2.4062

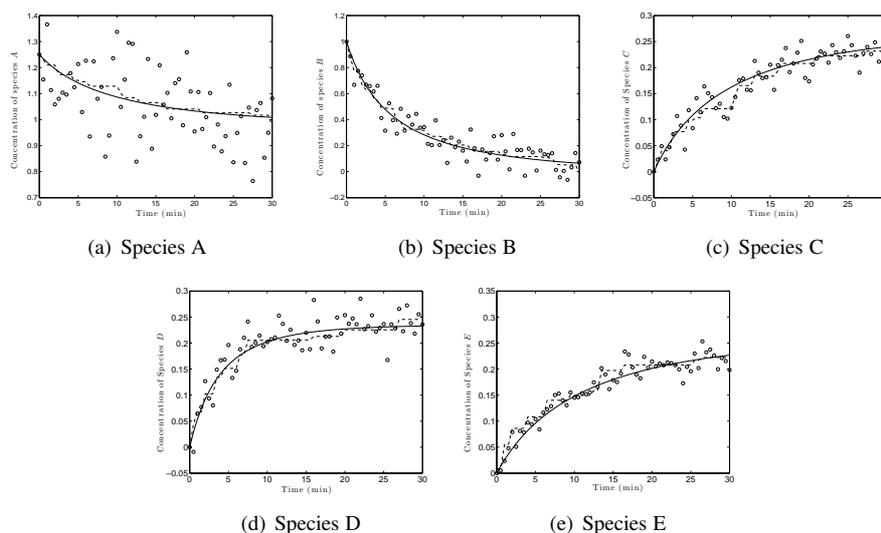


Figure 1: Simulated (continuous line), measured (dots) and extent-based reconciled (dashed line) concentrations of the species A to E.

5. Conclusion

This paper has shown that data reconciliation using invariant relationships helps reduce the effect of measurement noise. Furthermore, formulating the reconciliation problem in terms of extents allows exploiting additional monotonicity constraints. It has also been shown through a simulated example that the use of reconciled estimates together with the extent-based incremental approach leads to improved model discrimination. Future work will focus on developing monotonicity constraints for the extents of reaction and mass transfer in the context of open heterogeneous reaction systems.

References

- M. Amrhein, N. Bhatt, B. Srinivasan, D. Bonvin, 2010, Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams, *AIChE J.*, 56, 2873–2886.
- A. Bardow, W. Marquardt, 2004, Incremental and simultaneous identification of reaction kinetics: Methods and comparison, *Chem. Eng. Sci.*, 59, 13, 2673–2684.
- J. Billeter, S. Srinivasan, D. Bonvin, 2013, Extent-based kinetic identification using spectroscopic measurements and multivariate calibration, *Anal. Chim. Acta*, 767, 21–34.
- S. Narasimhan, C. Jordache, 1999, *Data Reconciliation and Gross Error Detection*. Elsevier.
- D. Rodrigues, S. Srinivasan, J. Billeter, D. Bonvin, 2015, Variant and invariant states for reaction systems, *Comp. Chem. Eng.*, 73, 1, 23–33.
- D. Ruppen, D. Bonvin, D. Rippin, 1998, Implementation of adaptive optimal operation for a semi-batch reaction system, *Comp. Chem. Eng.*, 22, 185 – 199.
- S. Srinivasan, J. Billeter, D. Bonvin, 2012, Extent-based incremental identification of reaction systems using concentration and calorimetric measurements, *Chem. Eng. J.*, 207–208, 785 – 793.